

論文 / 著書情報
Article / Book Information

題目(和文)	相関のあるパターンに基づくパーセプトロン学習とその応用に関する研究
Title(English)	A study on perceptron learning from correlated patterns and its applications
著者(和文)	新里隆
Author(English)	Takashi Shinzato
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第7530号, 授与年月日:2009年3月26日, 学位の種別:課程博士, 審査員:樺島 祥介
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第7530号, Conferred date:2009/3/26, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

相関のあるパターンに基づく
パーセプトロン学習とその応用に関する研究
A Study on Perceptron Learning from
Correlated Patterns and its Applications

新里 隆
東京工業大学大学院
総合理工学研究科
知能システム科学専攻

概要

人間は高度な学習機能を備えている。この知的能力を機械で実現させようという試みは計算機の出現以来、計算機研究者の中心的課題の1つである。機械学習 (machine learning) とは、パーセプトロン (perceptron) やサポートベクターマシン (support vector machine) などに代表される学習器 (learning scheme) の学習能力 (learning performance) を議論し、またそれを実行するアルゴリズムを開発することを目的とした人工知能研究の1分野である。高機能な計算機が昔に比べて安価になった今日、遺伝子解析や時系列解析などの大規模データからその現象に含まれる規則性を抽出し、利用することを目的とする問題の重要性が増しており、それらを行うことをも想定した「機械学習」は、理論・応用の両面で広く注目を集めている。本学位論文ではランダム行列理論と統計力学的手法を用いて、機械学習の一テーマであるパーセプトロン学習に関する学習能力の評価と、それを実現するアルゴリズムの開発を行い、パーセプトロン学習に対する新たな視点を提示する。

パーセプトロンとは、入力パターン（以下入力もしくはパターンとも記す）を高次元のベクトルで表したとき、それを含む高次元の入力空間を、結合加重を法線ベクトルとする「超平面」を用いて線形分離し、与えられた入力に対するクラス分類を行う学習器である。この識別器の入出力関係の表現能力は必ずしも高いわけではないが、構造が単純であるため扱いやすく、過学習の危険性が少ないという利点がある。また様々な用途に合わせたパーセプトロンやそれを拡張したニューラルネットワーク (neural networks) がいくつも提案され、記憶容量問題 (storage capacity problem) や規則性抽出問題 (regularity extraction problem) などの学習課題に対する学習効率に関して、これまでの研究で多くの知見が得られている。

さてパーセプトロン学習に対する既存の理論研究では、与えられた入力に対して「独立同分布の仮定 (independently and identically distributed)」を暗黙のうちにしていることが少なくない（以下 i.i.d. パターン）。その1つの理由として、入力が各々独立である課題 (examples) は一般的に言って、入力に相関が含まれている課題に比べて解析的に扱いやすく、また確率伝搬法 (belief propagation) などの実効的なアルゴリズムが構成しやすいという利点が挙げられる。しかしながら実用の面で考えると、実データに対して多くの場合、入力の「独立同分布の仮定」の妥当性が保証されているとは言い難い。そのため独立同分布の仮定のままで実データに適用した場合、パーセプトロンを用いて各種の学習課題を学習した後のクラス識別における、「系統的な誤り（相関が含まれるパターンに対して i.i.d. パターンだと仮定することによる、モデルミスマッチングに起因する予測誤差に含まれる誤りなど）」が発生する可能性がある。つまりこのような状況では、i.i.d. パターンに対する学習器のクラス識別に対する知見やアルゴリズムなど、クラス識別器の性能を十分に引き出せているとは言い難い。また他の例として、実データ解析において、入力が各々直交化する「前処理 (preprocessing)」をあらかじめデータに施している場合もある。問題にも依存するが、適切な「相関」を積極的に導入することで、クラス分類能力が上がるということが知られているため、「前処理」の有無で参照例題に対するクラス分類能力が向上するかどうか評価する必要がある。つまり入力間に相関が含まれるデータに対する、パーセプトロンのデータ解析能力の典型的な性能評価や実効的なアルゴリズムの開発など理論的整備が必要である。

我々はこれまで、既存研究で扱われてきた独立同分布の解析結果を含む形で「より広いクラス」のパターンに対して、統計力学に基づいて開発されてきた既存研究の解析手法の拡張を行ってきた。実際、与えられた参照例題の入力間に含まれる相関のクラスを、入力で構成される行列で定義される自己相関行列の固有値分布で特徴付けられる相関パターンに限定し（既存の i.i.d. パターンもこの特徴付けのクラスに含まれる）、ランダム行列理論で開発された行列積分公式を用いることで既存の解析手法の拡張・発展を行ってきた。特にそのクラスに含まれる「直交例題」と呼ばれるパターン例題と既存研究で扱われてきた独立同分布のパターン例題を用いて、機械学習で議論されている典型的なタスクに対する学習効率などの比較検証を行い、パターン間の相関が学習効率に与える影響に関して一定の知見

が得られた。またその結果を情報理論で現在も活発に研究が行われている通信方式（符号化方法）の一つである「誤り訂正符号」に対して適用し、その有効性を検証した。本学位論文ではそれらを詳説する。

謝辞

指導教員である樺島祥介先生に、懐の深い研究指導をしていただきました、感謝いたします。特に統計力学、情報理論、機械学習、組み合わせ最適化、数理統計学に関連する広範な知識に基づいた、たくさんの有益な助言を賜りました、さらに研究に取り組む姿勢や着眼点を教授していただきました。心から感謝いたします。青西亨先生、Isaac Pérez Castillo 先生、福島孝治先生、林正人先生、保坂忠明先生、井上純一先生、井上真郷先生、伊庭幸人先生、小林重信先生、村山立人先生、中村一尊先生、西森秀稔先生、岡田真人先生、大久保潤先生、大森敏明先生、Andrea Pagnani 先生、鈴木正先生、高安美佐子先生、竹田晃人先生、田中和之先生、田中利幸先生、安田宗樹先生、Martin Weigt 先生、和田山正先生、渡辺一帆先生、渡辺治先生、渡辺澄夫先生との創意に富む議論、助言、叱咤激励に感謝いたします。加えて樺島研究室の赤池洋俊さん、幡生敦史さん、林愛空さん、金子勇治さん、三屋諒太郎さん、白木善史さん、高橋久尚さん、高橋憲介さん、外崎幸徳さん、宇田新介さん、山中正雄さんには多くの協力をいただきました。また研究室合同のゼミを行った渡辺研究室の皆様、居室を共にした青西研究室、小林研究室の皆様に感謝いたします。学会や研究会などで有益な議論を行っていただいた荒井ひろみさん、比護貴之さん、金城敬太さん、北島顕正さん、三浦佳二さん、永田賢二さん、西山悠さん、岡野原大輔さん、太田桂輔さん、斎藤稔さん、杉浦正康さん、竹田隆治さん、竹内啓悟さん、佐藤一誠さん、白石友一さんに感謝いたします。在学中、様々な面で支援してくれた家族に感謝いたします。

最後に本学位論文におさめられている研究の一部は、科学研究費特定領域プロジェクト「情報統計力学の深化と展開 (DEX-SMI)」とグローバル COE プロジェクト「計算可能性の深化と展開 (COMPVIEW)」の補助を受けて遂行しました。深く感謝いたします。

目次

第 1 章 序論	9
1.1 背景	9
1.1.1 統計力学とスピングラス理論	9
1.1.2 機械学習	12
1.1.3 3つの論文	19
1.2 研究目的	21
1.3 アウトライン	21
第 2 章 予備知識	23
2.1 パーセプトロン学習	23
2.1.1 Bayes 推論	23
2.1.2 パーセプトロン	24
2.1.3 記憶容量問題	24
2.1.4 教師推定問題	24
2.1.5 コスト関数の導入	25
2.1.6 事後確率と Boltzmann 分布	26
2.1.7 まとめ	26
2.2 スピングラス理論	27
2.2.1 レプリカ法	27
2.2.2 Thouless-Anderson-Palmer 法	30
2.2.3 まとめ	31
2.3 ランダム行列理論	32
2.3.1 ランダム行列の固有値分布	32
2.3.2 ランダム正方行列の積分公式	39
2.3.3 ランダム長方形行列の積分公式	40
2.3.4 まとめ	43
2.4 まとめ	43
第 3 章 先行研究	45
3.1 パーセプトロン学習の統計力学的アプローチ	45
3.2 レプリカ解析と既存結果	47
3.2.1 レプリカ解析	47
3.2.2 レプリカ対称解	48
3.2.3 結合加重が離散値の場合	48

3.2.4	結合加重が連続値の場合	49
3.2.5	記憶容量問題	51
3.2.6	教師推定問題	52
3.3	平均場近似法	54
3.3.1	Bethe 自由エネルギー	55
3.3.2	離散値の場合	56
3.3.3	連続値の場合	59
3.3.4	アルゴリズム	61
3.4	まとめ	61
第 4 章	相関の特徴付けと定式化	63
4.1	相関の特徴付け	63
4.2	レプリカ解析	64
4.2.1	$\mathcal{F}(Q_w, Q_u)$ の導出	65
4.2.2	その他の項	66
4.2.3	レプリカ対称解	67
4.2.4	結合加重が離散の場合	69
4.2.5	結合加重が連続の場合	69
4.2.6	既存研究との整合性	69
4.3	Thouless-Anderson-Palmer 法	70
4.3.1	$N\tilde{g}^\beta(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u)$ の計算	71
4.3.2	$Ng^\beta(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u)$ の計算	71
4.3.3	アルゴリズム	72
4.4	まとめ	72
第 5 章	結果	75
5.1	直交例題	75
5.2	結果 1: 記憶容量問題	75
5.2.1	レプリカ解析	76
5.2.2	離散値の場合	77
5.2.3	連続値の場合	78
5.2.4	まとめ	79
5.3	結果 2: 教師推定問題	80
5.3.1	レプリカ解析	81
5.3.2	離散値の場合	82
5.3.3	連続値の場合	82
5.3.4	まとめ	86
5.4	結果 3: 通信方式への応用	86
5.4.1	誤り訂正符号の一般論	87
5.4.2	雑音あり通信路のモデル化	88

5.4.3	通信容量	89
5.4.4	通信路符号化定理	90
5.4.5	誤り訂正符号の提案	90
5.4.6	数値結果	91
5.5	まとめ	92
第 6 章	結論と今後の展望	95
6.1	本学位論文の結論	95
6.2	応用；機械学習	95
6.2.1	アンサンブル学習	96
6.2.2	バイオインフォマティクス	96
6.3	応用；情報技術	96
6.3.1	歪みありデータ圧縮	96
6.3.2	符号分割多元接続方式	97
付 録 A	$H(u)$ について	99
付 録 B	式 (4.9) の鞍点条件	101
付 録 C	線形回帰問題の統計力学的解析	105
C.1	線形回帰問題のレプリカ解析	105
C.2	線形回帰問題の TAP 法	107
C.3	拘束条件付き線形回帰問題	107
付 録 D	求解アルゴリズム	109
D.1	逐次過緩和法	109
D.1.1	SOR 法	109
D.1.2	応用例	110
D.2	Newton 法	110
D.2.1	1 変数の場合の Newton 法	110
D.2.2	多変数の場合の Newton 法	111
付 録 E	大偏差原理と Legendre 変換	113
E.1	大偏差統計もしくは鞍点法	113
E.2	Legendre 変換	116

第1章 序論

この章では博士論文全体を通して議論する，研究目的とそれに関連する研究背景について概説する．本章の構成は以下の通りである．次の節では物理学の1分野である「統計力学」と人工知能の1分野である「機械学習」の関係について触れる．さらに本研究の目的に関係する，3つの論文，E. Gardner (1988) [71], E. Levin, N. Tishby and S. A. Solla (1990) [129], M. Oppen and O. Winther (2001) [167] を後半に取り上げる．それを踏まえて続く節で本学位論文の研究目的を述べる．最終節にこの博士論文のアウトラインを示す．

1.1 背景

本節では本論文の背骨となる統計力学の説明からはじめる．

1.1.1 統計力学とスピングラス理論

(平衡) 統計力学とはミクロな運動法則とマクロな振る舞いをつなげる架け橋として導入された物理学の手法のことである．詳しくは [1, 14, 26, 27, 28, 73, 74, 100, 115, 116, 141, 159, 164, 197] などを参照されたい．平衡統計力学にはいくつかの形式が知られているが，ここでは本学位論文で用いる「正準形式 (canonical formalism)」を紹介する． N 個の構成要素からなる閉じた系に対して，その構成要素の状態を N 次元のベクトル $\vec{S} := \{S_1, S_2, \dots, S_N\}^T$ で表し，各状態 $\vec{S}$ に対して実数値関数 $\mathcal{H}(\vec{S}) \in \mathbf{R}$ が割り当てられた絶対温度 T の平衡状態に達した集団 (ensemble) を考える． $\mathcal{H}(\vec{S})$ はハミルトニアン (Hamiltonian) と呼ばれ，この系の全エネルギーを表す¹．このとき考えている集団の中で $\vec{S}$ の出現する確率 $P(\vec{S})$ は次の Boltzmann 分布に従うことが知られている ($\mathcal{H}(\vec{S})$ が小さいほど確率が高い)．

$$P(\vec{S}) := \frac{e^{-\beta \mathcal{H}(\vec{S})}}{Z} \quad (1.2)$$

ただし分母の規格化項 $Z := \sum_{\vec{S}} e^{-\beta \mathcal{H}(\vec{S})}$ は分配関数 (partition function) を表し，統計力学の中で重要な役割を果たす．また記号 $\sum_{\vec{S}}$ は $\vec{S}$ の取り得るすべての状態で和もしくは積分を実行することを表す (またこれ以降混乱のない限り断りなく記号 $\sum_{\vec{S}}$ を変数 $\vec{S}$ の「和」の便法として用いることにする)．これより内部エネルギー E とエントロピー S をそれぞれ Boltzmann 分布を用いて

$$E := \sum_{\vec{S}} \mathcal{H}(\vec{S}) P(\vec{S}), \quad S := -k_B \left(\sum_{\vec{S}} P(\vec{S}) \log P(\vec{S}) \right) = k_B \beta E + k_B \log Z \quad (1.3)$$

で表すことができる．定義よりあきらかに $\min_{\vec{S}} \mathcal{H}(\vec{S}) \leq E \leq \max_{\vec{S}} \mathcal{H}(\vec{S})$ である．ここで新たな関数 F を次のように定義する．

$$F := -\frac{1}{\beta} \log Z \quad (1.4)$$

これより式 (1.3) は $F = E - \frac{1}{k_B \beta} S$ となる．ここで熱力学で知られた経験則である熱力学的関係式 $F = E - TS$ と比較することで，式 (1.4) で定義された F は Helmholtz の自由エネルギーに対応し，さらに $\beta = \frac{1}{k_B T}$ となり，変

¹たとえば N 次正方行列 $J := \{J_{ij}\} \in \mathcal{M}_{N \times N}$ と N 次ベクトル $\vec{h} := \{h_1, h_2, \dots, h_N\}^T \in \mathbf{R}^N$ を用いて，

$$\mathcal{H}(\vec{S}) := -\frac{1}{2} \vec{S}^T J \vec{S} - \vec{h}^T \vec{S} \quad (1.1)$$

などと表される．ただし記号 T は行列やベクトルの転置を表す．

統計力学

$N < \infty$ では正準形式において以下の熱力学的関係式が成り立つ。

$$\text{Boltzmann 分布} \dots P(\vec{S}) = \frac{e^{-\beta \mathcal{H}(\vec{S})}}{Z}$$

$$\text{自由エネルギー} \dots F = E - TS = -\frac{1}{\beta} \log Z$$

$$\text{内部エネルギー} \dots E = \sum_{\vec{S}} \mathcal{H}(\vec{S}) P(\vec{S}) = -\frac{\partial}{\partial \beta} \log Z$$

$$\text{エントロピー} \dots S = -\sum_{\vec{S}} P(\vec{S}) \log P(\vec{S}) = -\frac{\partial F}{\partial T}$$

$$\text{分配関数} \dots Z = \sum_{\vec{S}} e^{-\beta \mathcal{H}(\vec{S})} = e^{-\beta F}$$

また重要なこととして、 $N \rightarrow \infty$ で平均場近似など（物理学の意味で）厳密な解析結果を導出する近似手法が存在する。

情報理論・機械学習

データ ξ^p を観測したときの $\vec{w}$ の事後確率 $P(\vec{w}|\xi^p)$ と相互情報量 $I(\vec{w}, \xi^p)$ は同時確率 $P(\vec{w}, \xi^p)$ を用いて

$$P(\vec{w}|\xi^p) = \frac{P(\vec{w}, \xi^p)}{\sum_{\vec{w}} P(\vec{w}, \xi^p)} = \frac{P(\xi^p|\vec{w})P(\vec{w})}{\sum_{\vec{w}} P(\xi^p|\vec{w})P(\vec{w})}$$

$$I(\vec{w}, \xi^p) = \sum_{\vec{w}, \xi^p} P(\vec{w}, \xi^p) \log \frac{P(\vec{w}, \xi^p)}{P(\vec{w})P(\xi^p)} \\ = H(\vec{w}) + H(\xi^p) - H(\vec{w}, \xi^p)$$

$$H(*) = -\sum_{*} P(*) \log P(*)$$

で定義される。ただし条件付き確率 $P(\xi^p|\vec{w})$ は与えられており、操作の面で見れば $P(\vec{w}|\xi^p)$ を用いて推論を行ったり、平均値を計算することが目的となる。

表 1.1: 統計力学と情報理論・機械学習の関係。左は条件付き確率を使わずに構成されており、右は条件付き確率を基礎においている。

数 β は絶対温度 T を用いて定義されることがわかる。また式 (1.4) は熱力学的関係式として知られている

$$S = -\frac{\partial F}{\partial T} \quad (1.5)$$

を満たしていることも確認できる。さらに

$$E = -\frac{\partial}{\partial \beta} \log Z \quad (1.6)$$

が成り立つ。これより先、エネルギーや温度などの物理的な次元がない問題に対して統計力学の方法を用いるため、物理学のシナリオの中でメトリックとして導入された Boltzmann 定数 $k_B \simeq 1.381 \times 10^{-23} \text{ J/K}$ は 1 と見なす [28]。ここで重要な点を 4 つ指摘しておく。

1. 自由エネルギー F や分配関数 Z はシステムを解析する上で重要な量（もしくは関数）である。なぜなら式 (1.5) や式 (1.6) を見ても分かるように、自由エネルギー F や分配関数 Z が幸運にも求めることができれば、その微分量によってシステムの物理量（たとえば E や S など）を算出することができるためである。同様に以下で説明するスピングラスでも重要な役割を果たす量であることに注意。
2. 式 (1.3) で定義されるエントロピーは C. E. Shannon (1948) が導入した情報エントロピーと対応する [182]。すなわち数理的な構造の類似性により、情報理論や統計的学習理論の問題に統計力学的手法が援用できることが（完全な証明が与えられているわけではないが）示唆される。
3. N が大きい極限でもいくつかの特殊な場合を除いて、分配関数 Z を厳密に評価することは困難である。なぜならシステムサイズ N に対して指数関数的に和、もしくは積分の計算量が増えるためである。しかしながら多くの模型に関して、対称性を用いた各種の変換（ゲージ変換やくりこみ変換など）や平均場近似などを用いることで、その計算困難を回避することができ、またその操作の正当性は、数値実験の外挿値や実際の物理実験の観測値の振る舞いを定性的・定量的に説明できることから、対称性を用いた各種の変換や平均場近似など、理論的な操作としてサポートされている。
4. 表 1.1 にあるように、統計力学と情報理論・機械学習の基本的な数理構造は似ているが、数理的な類似の中で決定的に異なる箇所が 1 つある。スピングラス理論（後述する）以前の統計力学では条件付き確率を念頭に構築されているわけではない。一方情報理論や機械学習では与えられた参照例題から、事後確率を用いてパラメータを推定するという「条件付き確率の枠組み」であり、両者には隔たりがある。

最後のコメントに関連して、物理学、特に S. F. Edwards らにより導入されたスピングラス理論の構築の過程で、条件付き確率の必要性が認識されるようになり、条件付き確率を記述する統計力学の理論整備がなされてきた。しかしながらそれに伴い、次に説明する新たな問題が発生してくる。

スピングラス理論

ここでは上記の問題、条件付き確率で記述される描像の解析に取り掛かる足がかりとして、あるランダムネス J に固定されたときのランダム磁性体を考える (図 1.1)。このときの Boltzmann 分布は

$$P(\vec{S}|J) = \frac{e^{-\beta\mathcal{H}(\vec{S}|J)}}{\sum_{\vec{S}} e^{-\beta\mathcal{H}(\vec{S}|J)}} \quad (1.7)$$

のようにランダムネス J に依存したハミルトニアン $\mathcal{H}(\vec{S}|J)$ を用いて定義される。この分布は対数尤度を $-\beta\mathcal{H}(\vec{S}|J)$ で定義する「条件付き確率」と対応していることが直感的に分かる。Boltzmann 分布と条件付き確率の関係については 20 ページで詳しく議論する。このとき均一系 (homogeneous system) と同様にランダム磁性体においても分配関数 $Z(J) := \sum_{\vec{S}} e^{-\beta\mathcal{H}(\vec{S}|J)}$ と自由エネルギー $F(J) = -\frac{1}{\beta} \log Z(J)$ を定義することができ、上記で紹介した均一系に対する「統計力学」のシナリオをランダムネス J で特徴付けられる系を扱えるように拡張できていることを意味しているが、このときシステムの典型的なふるまいを取り扱う状況下において、均一系では議論されない「ランダムネス平均」の問題が生じる。このランダムネスでの平均を次のように 2 種類考えることができる。

$$F_{\text{annealed}} = -\frac{1}{\beta} \log [Z(J)]_J \quad (1.8)$$

$$F_{\text{quenched}} = -\frac{1}{\beta} [\log Z(J)]_J \quad (1.9)$$

ただし記号 $[A(J)]_J$ はランダムネス J の関数 $A(J)$ の、 J が従う確率測度 $P(J)$ での平均を表す。式 (1.8) と式 (1.9) はそれぞれアニール平均 (annealed averaging)、クエンチ平均 (quenched averaging) と呼ばれ、一般に凸関数 (convex function) の性質から $F_{\text{annealed}} \leq F_{\text{quenched}}$ が成り立つ。ここでいくつかの特殊な場合を除いて一般的に式 (1.9) の計算は式 (1.8) に比べて難しいことが知られている。そこで S. F. Edwards と P. W. Anderson はレプリカ法 (replica method) と呼ばれる解析手法を用いてその困難を回避する方法を提案した。レプリカ法とは、以下で定義するレプリカトリック (replica trick) を用いてクエンチ平均をアニール平均と同程度の計算量で抑える方法であり、歴史は古く M. Kac (1968)[111], T. F. Lin (1970)[130], S. F. Edwards (1970,1971)[48, 49], V. J. Emery (1975)[54], G. Grinstein and A. H. Luther (1976)[77], D. Jasnow and M. E. Fisher (1976)[108] らにより導入された解析手法である。そのレプリカトリックとは

$$\log Z = \lim_{n \rightarrow 0} \frac{Z^n - 1}{n} \quad (1.10)$$

となる恒等式を用いて $\log Z$ の平均 $[\log Z]$ を Z^n の平均 $[Z^n]$ を用いて評価する操作のことである²。式 (1.10) 以外に $\log Z$ の平均 $[\log Z]$ を求める方法として、

$$[\log Z] = \lim_{n \rightarrow 0} \begin{cases} \frac{[Z^n] - 1}{n} \\ \frac{1}{n} \log [Z^n] \\ \frac{\partial}{\partial n} \log [Z^n] \end{cases} \quad (1.11)$$

などが知られているが、いずれにしても上記の方針において $[Z^n]$ の評価を行わなければならない。特に一般の $n \in \mathbb{C}$ に対しては特殊な場合を除いて $[Z^n]$ の評価は難しいが、 $n \in \mathbb{N}$ に限り、(1) アニール平均と同様に J の平均と $\vec{S}$ の

²ただし $Z > 0$ であり、 Z^n を $n = 0$ の周りでマクローリン展開 $Z^n = e^{n \log Z} \simeq 1 + n \log Z + O(n^2)$ することで式 (1.10) は証明される [158]。

平均操作が入れ替えることができ、(2) また入れ替えた後の模型に対して、平均場近似などの統計力学で開発された計算手法を用いることができるという2つの性質を利用できるため、式(1.11)の右辺の解析的な評価が可能であり、対数平均の困難性は解消されることが分かる。またランダム磁性体に対するランダムネス平均のアナロジーを用いて、表1.1に示すように統計力学に「条件付き」という概念を取り入れることで、確率的情報処理の枠組みとして援用することが期待できる。スピングラス理論の解析方法の詳細については第2章で説明する。

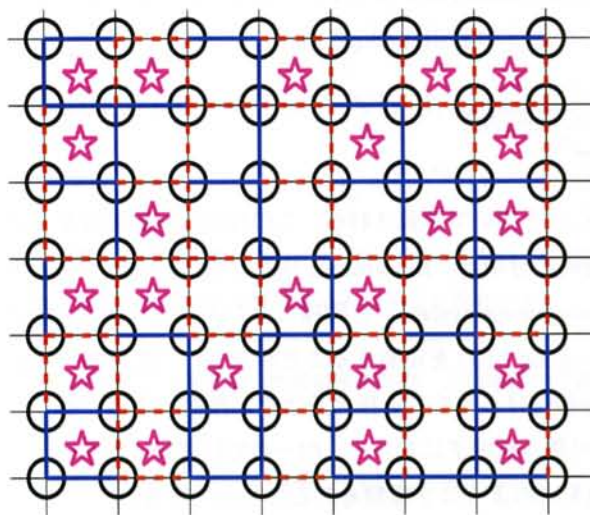


図 1.1: 格子にスピンの配置されているスピングラスの図。丸とエッジはそれぞれスピンと相互作用を表し、相互作用はランダムであると仮定する。スピン間の相互作用がランダムであるため、いたるところでスピンの要求を満足していないところが出てくる。ちなみに図において局所的にフラストレーションが起きているところを星印で表している。特に高温の場合は熱揺らぎによりスピンの向きを変えることができるが、低温ではエネルギー的に準安定な状態にスピンの向きが固まってしまう、フラストレーションが起きていることが分かる。これは均一系では見られない現象であり、均一系での秩序・無秩序転移とは違う転移現象であると解釈される。たとえば高温では空間的にも時間的にもバラバラな状態であるが、低温では空間的にはバラバラでも時間的には固まった(ランダムにそろった)状態になることが特徴的である [158]。

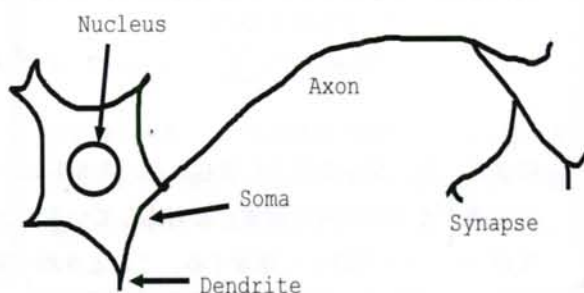


図 1.2: ニューロンの図。樹状突起 (dendrite) と細胞核 (nucleus) を持つ細胞体 (soma) から軸索 (axon) を伸ばして、シナプス (synapse) を用いて他のニューロンの樹状突起につながっている。

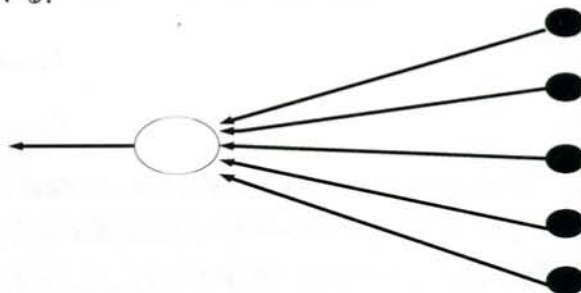


図 1.3: 形式ニューロンの図。白丸のニューロンに対して他の隣接するニューロン (黒丸) からくる影響によって発火するかどうかを判断する。図 1.2 と対応していることが分かる [134]。

1.1.2 機械学習

機械学習とは、(1) 人間の知的情報処理を模倣する学習器を設計したり、(2) 与えられた学習課題に対してその学習器の潜在的な学習効率を評価したり、(3) 柔らかな情報処理を実時間で実現する実効的なアルゴリズムの開発などを目的とする人工知能の1分野である。加えて高性能な計算機が安価になり、大規模なデータが扱えるようになりつつある今日において、大規模データに内在する規則性を抽出し利用することの重要性が増しており、それを可能とする機械学習は理論と応用の両面で広く注目されている。機械学習 (もしくは人工知能) の歴史は電子計算機の歴史と概ね符合しており、詳細については文献 [89, 179] などを参照されたい。ここでは機械学習の背景を概説するにあたり、神経科学 (Neuroscience) に起源をたどることにする。

脳は約 10^{11} 個の様々なタイプのニューロン (neuron) とそのおよそ 10 倍のグリア細胞 (glia cell) から構成されており、樹状突起 (dendrite) と細胞核 (nucleus) を持つ細胞体 (cell body, soma) 同士を神経線維でつながった「ネットワーク構造」を持つ。各ニューロンは、領域野 (cortex) によっても違うが大体 10^4 個の隣接するニューロンからシナプス (synapse) を通して入力を受ける。細胞体内部は細胞外に比べて通常およそ -70mV の電位差があり、

たくさんの隣接するニューロンから刺激が加わり、ある閾値を超えると、そのニューロンは軸索 (nerve fiber, axon) と呼ばれる長い繊維状の部分を通じて他のニューロンへパルス信号を流す。これはニューロンの発火 (fire) と呼ばれ、ニューロンは発火後、数 ms の間 (不応期) はどんな入力も来ても再び発火しない。この期間を過ぎるとまた発火前の状態に戻る (図 1.2)。一方もう一つの構成要素であるグリア細胞は、(1) 近傍のニューロンに対して電氣的に絶縁したり、(2) 構造を支持したり、(3) 栄養を与えることで脳機能の一役を担っていることが解明されている³。以上のように神経科学のめまぐるしい進歩により、脳の構成要素のメカニズムは解明されつつあるが、それらが複雑に絡み合った「脳」全体の機能については、未だ解明されていない現象が多い。

さて機械学習への橋渡しとして、知的情報処理としてのニューロンの次の性質に着目する。ニューロンの機能を数理的に単純化すれば、多数の入力を受けてその総和を計算し、閾値と比較して発火するかどうかを決定する「多入力・一出力の非線形素子」でモデル化することができる。1943 年に W. S. McCulloch と W. Pitts はニューロンの生理学的な機能を踏まえて、図 1.3 のようなニューロンの発火するか発火しないかを 2 状態で表し、閾値を持つ素子としての「形式ニューロン (formal neuron)」を提唱した [134]。彼らは、ニューロンを数理モデル化することで容易に計算機を用いてシミュレーションすることができ、また数理モデルの学習性能を評価して得られた知見を用いて、生理学的なニューロンの振る舞いの予測に使えることを述べている。この研究に触発され、以下で説明するように、多くの機械学習の研究が行われるようになった。本小節では機械学習の専門用語の中で、本学位論文に関係する用語とそれに対比する用語の一部を、「学習器」、「学習課題」、「学習過程」、「完全・不完全データ」の 4 つのカテゴリに分けて説明する。

♣ 学習器

ここでは代表的な学習器を紹介する。大きく分けて 2 種類のタイプに分けることができる。Hopfield 模型と Boltzmann マシンはシステム全体の中での入力と出力の役割があらわに分離していない、つまりあるニューロン状態はそれに相互作用するすべてのニューロンの出力を入力信号とするネットワークを用いて学習する「相互干渉型」の学習器である。一方パーセプトロンとニューラルネットワーク (無向・閉路無しのグラフで記述される) はシステム全体の中での入力と出力の役割があらわに分離している、つまりあるニューロンはネットワーク外部から与えられる入力信号によって、発火の有無を判定する学習器である。本学位論文では主に後者のパーセプトロンを研究対象にする。しかしながらここでは相互干渉型の学習器をはじめに説明する。

A 1. Hopfield 模型 J. J. Hopfield は 1982 年に神経回路網の特徴である、相互干渉型のネットワークを用いて、連想記憶 (associative memory) が行える模型を提案した [92]。これは統計力学的手法を用いて学習器の学習能力を評価した初めての例として知られている。Hopfield 模型を説明するために、はじめに W. S. McCulloch and W. Pitts (1943)[134] の形式ニューロンを用意する。そのニューロンモデルとは時刻 t の i 番目のニューロンの発火しているか、発火していないかの状態をそれぞれ $S_i(t) = \pm 1$ で表し、 i 番目のニューロンに隣接する j 番目のニューロンの時刻 $t-1$ の状態 $S_j(t-1)$ に相関の強さを表すシナプス加重 J_{ij} を掛け合わせた入力の総和と閾値 θ_i を比較して発火するかどうか決める次の離散時間力学系の模型

$$S_i(t) = \operatorname{sgn} \left(\sum_{j(\neq i)} J_{ij} (S_j(t-1) + 1) - \theta_i \right) \quad (1.12)$$

のことである [134]。ちなみに $\operatorname{sgn}(u) = u/|u| = \pm 1, (|u| > 0, u \in \mathbf{R})$ は符号関数である⁴。ただし $J := \{J_{ij}\} \in \mathcal{M}_{N \times N}$ の固有値分布はすべて非負であると仮定する。説明を簡単にするために閾値を $\theta_i = \sum_{j(\neq i)} J_{ij}$ とし、

³近年グリア細胞も脳の情報処理に関係している可能性があるという報告がある [61, 85, 126]。

⁴ $\operatorname{sgn}(u)$ において、線形分離面 ($u = 0$) にほとんど点が存在していない場合、つまり $\lim_{N \rightarrow \infty} P(u = 0) = 0$ であれば問題ないが、そうでない場合は注意が必要である。いくつかの方策を用いて、たとえば、前処理的に $u = 0$ に属する点集合と $u \neq 0$ に属する集合に分ける方法やサポートベクターマシンのようにマージン κ を付けることで回避することができる。

式 (1.12) は

$$S_i(t) := \text{sgn}(h_i(t-1)) \quad h_i(t-1) := \sum_{j(\neq i)} J_{ij} S_j(t-1) \quad (1.13)$$

となる。細かい話であるが式 (1.13) の力学系を実際に動かす場合、同期更新か非同期更新なのかによってアトラクタが異なることに注意。だがここでは $t \rightarrow \infty$ の状況、つまり力学系の固定点の安定性について議論を展開する ($S_i := \lim_{t \rightarrow \infty} S_i(t)$, $h_i := \lim_{t \rightarrow \infty} h_i(t)$)。彼のするどい洞察の一つとして、この力学系の式 (1.13) で表される「硬い決定」を、元の方程式を含む形で、確率的な解釈として「柔らかい決定」に拡張したことがある。つまり他の変数を固定したとき、 $S_i = 1$ になる確率を、 $\beta > 0$ を変数とし、

$$P(S_i = 1) := \frac{1}{1 + e^{-\beta h_i}} \quad (1.14)$$

と定義した。ちなみに式 (1.14) は $\beta \rightarrow \infty$ のとき硬い決定 (式 (1.13)) の固定点 (方程式) を再現する。また $P(S_i) \propto e^{\beta h_i S_i / 2}$ のため、自然な拡張として、他の変数を固定せずに全体の確率 $P(\vec{S})$ を次のハミルトニアン

$$\mathcal{H}(\vec{S}|J) := -\frac{1}{2} \vec{S}^T J \vec{S} \quad (1.15)$$

で定義される Boltzmann 分布であると仮定する。

$$P(\vec{S}) := \frac{e^{-\beta \mathcal{H}(\vec{S}|J)}}{Z} \quad (1.16)$$

ただし規格化定数を統計力学に習い、分配関数 $Z := \sum_{\vec{S}} e^{-\beta \mathcal{H}(\vec{S}|J)}$ と見なす。J. J. Hopfield はこの模型の学習課題として、 p 個のパターン $\vec{x}_\mu$ ($\mu = 1, 2, \dots, p$) を埋め込んだネットワークのシナプス加重 $J = \{J_{ij}\}$ を用いて「連想記憶」ができることを示した。特に彼は 1949 年に心理学者の D. O. Hebb によって提案された次の Hebb 則を採用した [86]⁵。

$$J := \frac{1}{N} \sum_{\mu=1}^p \vec{x}_\mu \vec{x}_\mu^T \in \mathcal{M}_{N \times N} \quad \left(J_{ij} := \frac{1}{N} \sum_{\mu=1}^p x_{i\mu} x_{j\mu} \right) \quad (1.18)$$

また p 個のパターン $\vec{x}_\mu \in \{\pm 1\}^N$ をランダムに選べば近似的に直交していることが望まれる。

$$\frac{1}{N} \vec{x}_\mu^T \vec{x}_\nu = \delta_{\mu\nu} + O\left(\frac{1}{\sqrt{N}}\right) \quad (1.19)$$

この入力間の直交性により、このシナプス加重 J において、埋め込んだパターン $\vec{x}_\nu$ は高い確率で式 (1.13) を満たすことも確認できる⁶。これより D. J. Amit, H. Gutfreund and H. Sompolinsky(1985,1987) はパターン数 p が有限の場合 $p \sim O(1)$ とパターン比 α が有限の場合 $\alpha = p/N \sim O(1)$ のそれぞれに対して解析を行った [8, 10]。

◎ $p \sim O(1)$ D. J. Amit, H. Gutfreund and H. Sompolinsky(1985) はパターン数 p が有限の場合について、埋め込んだパターンが想起できるかどうか検証した [8]。 β が十分大きいところで、対称性から 1 番目のパターンを考えるが、パターン $\vec{x}_1$ の近傍から解探索をスタートしたとき、高い確率で埋め込んだパターン

⁵Hebb 則はもともとシナプス加重 J_{ij} の更新則として提案されている。時刻 t のシナプス加重 J_{ij}^t 、入力 $x_i(t)$ と出力 $x_j(t)$ をそれぞれ表し、小さな学習率 $\varepsilon > 0$ を用いて

$$J_{ij}^{t+1} = J_{ij}^t + \varepsilon x_i(t) x_j(t) \quad (1.17)$$

で更新則を定義する。この更新則の派生形はいくつか報告されている。また実際海馬における運動野で行われていることが確認されている。(ある程度独立な) p 個のパターンを式 (1.17) を用いて学習し、 $\varepsilon = 1/N$ とおけば、式 (1.18) のシナプス加重は式 (1.17) で生成されることが分かる。

⁶以下の先行研究において本質ではないが、本学位論文では $J_{ii} = 0$ は要請しないことにする。なぜなら式 (1.18) の定義により J が半正定値 (SPD) であることが保証されるが、 $J_{ii} = 0$ では保証されないと考えられるからだ。

$\vec{x}_1$ を想起できることが示された。また $\beta \rightarrow \infty$ のとき、 p 個のパターンのうち奇数個のパターン ($2l+1$ 個) を同等に想起することのできる状態の自由エネルギー f_{2l+1} を用いて、 $f_1 < f_3 < f_5 \dots$ の単調非減少列になることが示された。一方偶数個の場合は $(\lim_{l \rightarrow \infty} f_{2l+1} <) f_2 < f_4 < \dots$ となることも示された。
 ◎ $\alpha = p/N \sim O(1)$ D. J. Amit, H. Gutfreund and H. Sompolinsky(1987) はパターン比 $\alpha = p/N$ が有限の場合において、典型的に埋め込んだパターンを完全に想起できる (偽のパターンを想起しない) パターン数の上限 p_C とシステムサイズとの比 $\alpha_C = p_C/N$ を解析的に調べた [10]。結果として $\beta \rightarrow \infty$ で $\alpha_C \simeq 0.138$ となることが示された。

A 2. Boltzmann マシン 説明を簡単にするためにひとまず隠れ素子のない場合を議論する (隠れ素子のある場合は後ほど触れる)。さて確率変数 $\vec{S} \in \{\pm 1\}^N$ のサンプル平均 $\langle \vec{S} \rangle_D$ や共分散行列 $\langle \vec{S} \vec{S}^T \rangle_D$ が与えられた状況で、ただしそれを生成する確率分布 $P_t(\vec{S})$ は未知である状況で、 $P_t(\vec{S})$ を推定する問題を考える。 $\langle \vec{S} \rangle_D$ や $\langle \vec{S} \vec{S}^T \rangle_D$ をもとに、 $P_t(\vec{S})$ を推定することは、一種の学習問題である。この問題に対して G. E. Hinton and T. J. Sejnowski は 1983 年に J. J. Hopfield の研究をもとに式 (1.15) で定義されたハミルトニアンのように

$$\mathcal{H}(\vec{S}|J, \vec{h}) := -\frac{1}{2} \vec{S}^T J \vec{S} - \vec{h}^T \vec{S} \quad (1.20)$$

で定義される温度 $1/\beta$ の Boltzmann 分布 $P(\vec{S}|J, \vec{h}) \propto e^{-\beta \mathcal{H}(\vec{S}|J, \vec{h})}$ で $P_t(\vec{S})$ を近似する学習器を提案した [91]。この学習器のことを Boltzmann マシンと呼び、この問題の推定パラメータは $J, \vec{h}$ である。ただし $J_{ij} = J_{ji}$ を仮定する。カルバックライブラー擬距離 (Kullback-Leibler divergence) を用いて $P_t(\vec{S})$ と $P(\vec{S}|J, \vec{h})$ の類似度 $KL := \sum_{\vec{S}} P_t(\vec{S}) \log \frac{P_t(\vec{S})}{P(\vec{S}|J, \vec{h})}$ を測ることができる。この KL をポテンシャルとする最急勾配法を用い (更新時刻 t の変数をそれぞれ $J^t, \vec{h}^t$ で表し、小さな学習率 $\varepsilon > 0$ とする), $J^{t+1} = J^t - \varepsilon \beta \left(\langle \vec{S} \vec{S}^T \rangle_D - \langle \vec{S} \vec{S}^T \rangle_{J^t, \vec{h}^t} \right)$ と $\vec{h}^{t+1} = \vec{h}^t - \varepsilon \beta \left(\langle \vec{S} \rangle_D - \langle \vec{S} \rangle_{J^t, \vec{h}^t} \right)$ の更新則の固定点である $J = \lim_{t \rightarrow \infty} J^t, \vec{h} = \lim_{t \rightarrow \infty} \vec{h}^t$ に収束するとき、 KL の極小値を与えることがわかる⁷。ちなみに $\langle A(\vec{S}) \rangle_{J^t, \vec{h}^t} = \sum_{\vec{S}} P(\vec{S}|J^t, \vec{h}^t) A(\vec{S})$ を用いた。さらに隠れ素子がある場合、つまり $\langle \vec{S} \rangle_D$ や $\langle \vec{S} \vec{S}^T \rangle_D$ に観測できない成分があったとしても、隠れ素子を J や $\vec{h}$ と同様にハイパーパラメータとして扱えば、上記のシナリオで解釈することができるため説明は省略する。ちなみに Boltzmann マシンの解析手法として EM アルゴリズムや変分ベイズ法などのアルゴリズムが提案されている [7, 42, 140, 173]。

B 1. パーセプトロン パーセプトロン (perceptron) とは、W. S. McCulloch と W. Pitts (1943) の形式ニューロンをもとに [134], F. Rosenblatt(1958) によって提案された、入力 $\vec{x}$ と出力 y の組を用いた「2 クラス判別のできる学習器」であり、高次元の入力空間を線形分離する「超平面 (hyperplane, その法線ベクトルを結合加重 $\vec{w}$ で表す)」を用いて入力パターンに対するクラス分類を行うことが出来る機能を持つ学習器である [174]。線形分離のため、入出力関係の表現能力は必ずしも高いわけではないが、構造が単純であるからこそ扱いやすく、過学習の危険性が比較的少なく、また実時間で実効的なアルゴリズムもいくつか存在しているという利点がある。さらにクラスを多クラスに拡張することも可能であり、パーセプトロン以外の様々な工夫を凝らした認識器が数多く提案されている現在でもパーセプトロンは機械学習の礎の1つをなしている。さて本学位論文でターゲットとするパーセプトロンを、2 クラス識別器として次のように定義する。

$$y = \text{sgn} \left(\frac{\vec{x}^T \vec{w}}{\sqrt{N}} - \kappa \right) \quad (1.21)$$

N 次元の入力 $\vec{x} \in \mathbf{R}^N$ に対して線形分離面を特徴付ける N 次元の結合加重 $\vec{w}$ を調整することで所望の出力 y を割り当てることができる。ただし $\kappa \sim O(1)$ は閾値を表し、各成分 x_k, w_k は $O(1)$ の量であると仮定し、前

⁷最急降下法以外にも KL の最小化のスケジュールとして、焼きなまし法 (simulated annealing) を用いたマルコフ連鎖モンテカルロ法などが考えられるが、統計誤差が大きく、収束に時間がかかるため、実用的ではない [99, 148]。

パーセプトロンアルゴリズム

ここでは議論を簡単にするために $\kappa = 0$ とし、逐次代入の反復時刻 t の結合加重を $\vec{w}_t := \{w_{1t}, w_{2t}, \dots, w_{Nt}\}^T \in \mathbf{R}^N$ で表す。この問題を解く際のコメントとして、分離面を特徴付ける法線ベクトル $\vec{w}$ は、原点からの「長さ」は意味を持たず、ベクトルの「方向」が重要であることに注意。また与えられた $(\vec{x}, y)$ に対して、入出力関係を満たさない $\vec{w}$ に対して、 $-y \times \text{sgn}\left(\frac{1}{\sqrt{N}} \vec{x}^T \vec{w}\right) > 0$ もしくは $-y \times \left(\frac{1}{\sqrt{N}} \vec{x}^T \vec{w}\right) \geq 0$ が成り立つので、この問題の評価関数を次のようにおく。

$$E(\vec{w}_t) := \sum_{\mu=1}^p E_{\mu}(\vec{w}_t), \quad E_{\mu}(\vec{w}_t) := -y_{\mu} \vec{x}_{\mu}^T \vec{w}_t. \quad (1.22)$$

線形分離可能問題のとき、この関数の値を下げるように線形分離面を特徴付ける結合加重 $\vec{w}$ を更新する。具体的には μ 番目の入出力関係 $(\vec{x}_{\mu}, y_{\mu})$ に着目し、 $E_{\mu}(\vec{w}_t)$ をポテンシャルと考え、小さな学習率 $\varepsilon > 0$ を用いた次の最急降下法が導かれる。

$$\vec{w}_{t+1} = \vec{w}_t - \varepsilon \times \begin{cases} \frac{\partial E_{\mu}(\vec{w}_t)}{\partial \vec{w}} & E_{\mu}(\vec{w}_t) \geq 0 \\ 0 & \text{otherwise} \end{cases} = \vec{w}_t + \varepsilon \times \begin{cases} y_{\mu} \vec{x}_{\mu} & E_{\mu}(\vec{w}_t) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1.23)$$

この式は入出力関係を間違ったときに $\varepsilon y_{\mu} \vec{x}_{\mu}$ だけ結合加重を更新することを表す。これは明らかに

$$E_{\mu}(\vec{w}_{t+1}) = E_{\mu}(\vec{w}_t) - \varepsilon \times \begin{cases} y_{\mu}^2 \vec{x}_{\mu}^T \vec{x}_{\mu} & E_{\mu}(\vec{w}_t) \geq 0 \\ 0 & \text{otherwise} \end{cases} \leq E_{\mu}(\vec{w}_t) \quad (1.24)$$

となり、 $E_{\mu}(\vec{w}_t)$ は t に対して単調非増加関数であり、 μ 番目の入出力関係を満たすまで式 (1.23) を繰り返す。そして入出力関係を満たしたとき、次の入出力関係の参照にうつる。こうしてもしすべての参照例題 ξ^p の入出力関係を再現することができれば、このアルゴリズムを停止することができる。またこのアルゴリズムは $t \rightarrow \infty$ で

$$\theta = \cos^{-1} \left(\frac{\vec{w}_t^T \vec{w}_{t+1}}{|\vec{w}_t| |\vec{w}_{t+1}|} \right) \rightarrow 0 \quad (1.25)$$

となるため、線形分離面は収束することが分かる。このアルゴリズムはパーセプトロンアルゴリズムもしくはパーセプトロン学習則と呼ばれる。

図 1.4: パーセプトロンアルゴリズムの導出の流れ

因子の $\frac{1}{\sqrt{N}}$ は $N \rightarrow \infty$ でも引数が有効に利く規格化因子である（前因子の議論は、中心極限定理から類推される）。ここでは p 個の入出力関係 $(\vec{x}_{\mu}, y_{\mu})$ の組 $\xi^p := \{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_p, y_p)\}$ が与えられているとき、特にここでは議論を限定して、線形分離可能な参照例題を与えられているとき、それを可能にする分離面を求めるアルゴリズムを考える。それは1960年にF. Rosenblattによって行われた、パーセプトロン学習のアルゴリズムであり、証明の概要を図1.4で説明する。

特筆すべきこととして、F. Rosenblattは学習の過程を、つまり手続きとしての「学習アルゴリズム」を導出したことである。これに触発されて、もしくはほぼ同時期に、B. Widrow(1962)がadatronを[212]、A. Gamba, L. Gamberini, G. Palmieri and R. Sanna (1961)がGamba perceptronを提案している[69]。これらを含む初期の研究はC. M. Bishop[21]の本とN. J. Nilsson[156]の本に詳しくまとめられている。さてたとえば図1.5と図1.6から分かるように、論理和ORと論理積ANDは単層パーセプトロンを用いて実現することができる。しかしながら表1.7から分かるように、排他的論理和XORは単層パーセプトロンを用いて、どんな結合加重と閾値を用意しても表現することができない。つまりたとえ単純な入出力関係であったとしても、線形分離できない参照例題に関しては単層パーセプトロンで表現することができない。機械学習の分野の中でもっとも有名な書籍の1つであるM. L. MinskyとS. A. Papertにより1969年に出版された「Perceptrons」の中で[139]、彼らは上記のパーセプトロンで表現できるのは、あくまで線形分離可能な参照例題に限られることを数学的に証明を与えた。また下で説明する多層のネットワークに対しても収束性が数学的に保証された効率のよ

入力 $\vec{x} := (x_1, x_2)$, 結合加重 $\vec{w} := (w_1, w_2)$, 閾値 κ とする. 出力は

$$y = \text{sgn} \left(\frac{x_1 w_1 + x_2 w_2}{\sqrt{2}} - \kappa \right)$$

で与えられる. 論理和 OR と論理積 AND は (単層) パーセプトロンを用いて実現可能である.

OR	0	1
0	0	1
1	1	1

図 1.5: 論理和 OR. $w_1 = w_2 = \sqrt{2}$ と $\kappa = 1.5$ により OR が実現する.

AND	0	1
0	0	0
1	0	1

図 1.6: 論理積 AND. $w_1 = w_2 = \sqrt{2}$ と $\kappa = 0.5$ により AND が実現する.

XOR	0	1
0	0	1
1	1	0

図 1.7: 排他的論理和 XOR. (単層) パーセプトロンで表現することができない.

い学習アルゴリズムが存在しないことに証明を与え, これより先 J. J. Hopfield(1982) の Hopfield 模型や D. E. Rumelhart, G. E. Hinton and R. J. Williams (1986)[176, 177] が世に出るまで機械学習の氷河期が訪れることになる⁸.

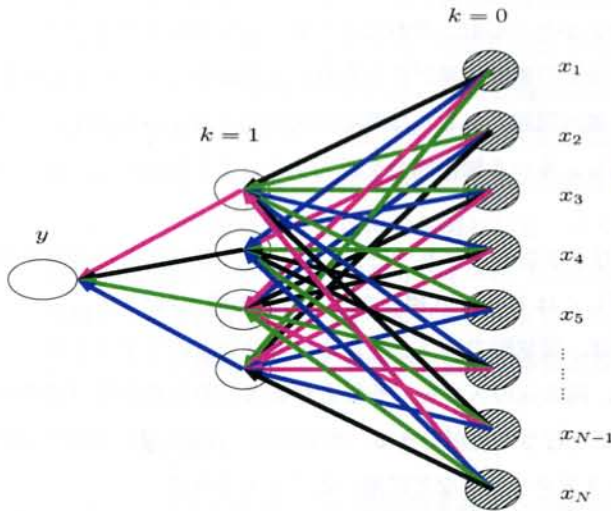


図 1.8: 多層ニューラルネットワークの図. 左図の有向グラフは右の入力層と左の出力層の間に中間層 (隠れ層) 1 つ ($K = 1$) を持つ多層ニューラルネットワークを表す (3 層パーセプトロンとも呼ばれる). 回帰問題を学習させる場合は線形関数 $g_k(u) = u$ を用い, 判別問題を学習させる場合は階段関数や sigmoid 関数 $g_k(u) = \tanh(u)$ もしくは $g_k(u) = (1 + e^{-u})^{-1}$ を用いる. $K = 1$ のとき, つまり中間層が 1 つのとき, XOR を再現できることが知られている.

B 2. 多層ニューラルネットワーク 多層ニューラルネットワーク (multilayered neural network) とは図 1.8 にあるように入力 $\vec{x}$ と出力 y の間に「隠れ層 (図 1.8 の真ん中の層)」が階層的にはさまれている学習器のことである. 各層の間ではパーセプトロンと同じ働きをする関数か, もしくは任意の非線形関数 $g_k \left(\frac{1}{\sqrt{N_k}} \sum_{j=1}^{N_k} x_{\mu j, k} w_{j, k} \right)$ を用いて $k + 1$ 番目の隠れ層の入力を $\vec{x}_{\mu, k+1} := \{x_{\mu 1, k+1}, x_{\mu 2, k+1}, \dots, x_{\mu N_{k+1}, k+1}\}^T \in \mathbf{R}^{N_{k+1}}$ で表したとき, $x_{\mu l, k+1} = g_k \left(\frac{1}{\sqrt{N_k}} \sum_{j=1}^{N_k} x_{\mu j, k} w_{j, k} \right)$ に従って入力 $\vec{x}_{\mu, k}$ を更新していく ($k = 0, 1, 2, \dots, K$ かつ $\vec{x}_{\mu, 0} = \vec{x}_{\mu}$ とする. さらに $N_{K+1} = 1$ として $\vec{x}_{\mu, K+1}$ を y_{μ} と見なす). ただし k 番目の結合加重を N_k 次元の $\vec{w}_k := \{w_{1, k}, w_{2, k}, \dots, w_{N_k, k}\}^T \in \mathbf{R}^{N_k}$ で表し, K 個の結合加重 (総成分数は $\sum_{k=0}^K N_k$) を調整することで与えられた参照例題 ξ^p を学習する. この学習器は単にニューラルネットワーク, 特に中間層が K 個であるので, $K + 2$ 層ニューラルネットワークと呼ばれる (+2 は入力層と出力層もカウントに入れている). この仕組みにより $K = 1$ で任意の非線形関数を表現できることが知られている [21].

⁸彼らは「Perceptrons(1969)」の改訂版として, 1990 年に同タイトルの書籍の中で, PDP に代表される「コネクショニズム」に対しても警鐘を鳴らしている. また皮肉にも著者の一人である M. L. Minsky は, D. Edmonds(1951) と一緒に世界で初めての SNARC と呼ばれるニューラルネットコンピュータを作成している.

歴史的には1981年のG. E. Hinton and J. A. Anderson[90]と1986年に刊行されたParallel Distributed Processing (PDP)の1巻と2巻によって、コネクショニズムの人氣に火がつけられたと言われている。特にPDPの2巻におさめられている、誤差逆伝搬法(back propagation)を扱ったD. E. Rumelhart, G. E. Hinton and R. J. Williams (1986)[176]や彼らのNatureの論文(1986)[177]は多くの研究者の注意をひきつけた。実際現在も活発に行われている機械学習の源の一つとしてあげることができ、我々が行った研究に関連する、E. Gardner(1988)[71]やE. Levin, N. Tishby and S. A. Solla(1990)[129]などもこのコネクショニズムに触発された研究である。

♣ 学習課題

次に典型的な学習課題を紹介する。本学位論文では特に「記憶容量問題」と「教師推定問題」を取り扱う。

1. **記憶容量問題** 入力と出力をそれぞれ $\vec{x}$ と y で表し、入出力関係 $(\vec{x}, y)$ が与えられたとき、特に入力とはまったく無関係にランダムなラベルを与えたとき、学習器を用いてどれくらい多くの入出力関係 $(\vec{x}, y)$ を記憶することができるか、という学習問題のことを記憶容量問題と呼ぶ。ランダムラベルを再現できることは、「乱数」を模倣していることに相当し、この学習問題では学習器の「表現能力」を評価することに対応する。第2章で詳しく議論する。
2. **教師推定問題** 教師ネットワーク(オラクル; oracle)が与えられ、それにより各入力に対してクラスラベルを割り当てた入出力関係 $(\vec{x}, y)$ を用いて、どれくらい多くの参照例題が与えられると教師ネットワークを同定することができるか、という学習問題のことを規則性抽出問題、もしくは教師推定問題と呼ぶ。教師ネットワークを同定できることは規則性を抽出することに相当し、ルール抽出能力を評価することに対応する。第2章で詳しく議論する。
3. **連想記憶** 連想記憶とは、Hopfield 模型のような相互干渉型のネットワークに対して、埋め込んだ情報を想起できるかどうか議論する学習課題のことである。詳しくはHopfield 模型のところを参照されたい。
4. **自己組織化** 自己組織化(self organizing maps)とは、視覚野のニューラルネットワークのモデル化の一つとして有名な学習課題である[121, 122]。具体的には、高次元の入力を、それより低次元の位相空間(近傍や距離が定義されている空間)に射影し、クラス分類などを行うことができる。次元削除(pruning)を用いてデータを可視化することができ、コミュニティ検出やクラスタリングなどに用いることもできる。

♣ 学習過程

学習器が例題を参照する際にはいくつかの方法が考えられる。本学位論文では主にバッチ学習を扱う。

1. **バッチ学習** バッチ学習(batch learning, もしくは一括学習)とは、 p 個の入出力関係の組 ξ^p のすべてを使って学習を行い、パラメータを決定する学習方法である。情報理論における「ブロック符号・復号化」に相当し、一般的に計算量やメモリの面で以下で説明するオンライン学習に比べて不利な面を持ち合わせている。だが学習後の学習器が、得られたデータに関しては完全に入出力関係と、整合性が取れることが特徴である。
2. **オンライン学習** オンライン学習(on-line learning, 逐次学習(sequence learning)とも言う)とは、入出力関係を1つずつ参照してパラメータを更新し、それより前の学習例題の情報は推定したパラメータに繰り込まれているとして、学習を行う方法のことである。オンライン学習は、動的な変化をするオラクルに対応することができ、使用するメモリが少なく済む。一方パラメータの更新則が自明でない場合も多く、また学習の済んだ例題の入出力関係を満たさない場合もある。
3. **質問学習** 質問学習(query learning)とは学習器がオラクルに質問し、それに答えた答えを用いて学習する方法である。一般的にランダムな入出力関係を用いてバッチ学習を行うよりも、学習器にとって、現在の参照例題

だけではクラス判別しづらい入力を質問することができるため効率がよいことが知られている。質問学習には大きく分けて2つあり、それぞれ「属性質問 (membership query)」、「等価性学習 (equivalence query)」がある。「属性質問」とは、オラクルに新たな入力 $\vec{x}$ を質問し、出力を聞く学習方法である。「等価性質問」とは、オラクルの結合加重が推定結果の $\vec{w}$ であるかを質問し、もし違っていれば、反例を要求することができる学習方法である。属性質問に答えるオラクルを仮定することは、実験などでデータを採取することを考えれば自然であるが、等価性質問に答えられるオラクルは反例を探さないといけなく、計算量的に困難であるため現実的ではない。

4. 強化学習 強化学習 (reinforcement learning) とは、もともと心理学で用いられていた概念であるが、質問学習と同様に現在の状況を下に学習器が意思決定 (action) を行ったとき、オラクルはそれに対して賞罰 (reward and penalty) を与えることで、知識を獲得する学習方法のことである。逐次的に報酬を与える場合もあるが、意思決定の行動列 (policy) に対して報酬を与えることもある。

♣ 完全・不完全データ

1. ラベルあり学習 入力パターンに対して、そのパターンが属するクラスを示すクラスラベルが完全に割り当てられている参照例題を用いて様々な学習課題を議論することを「ラベルあり学習 (Supervised learning)」と呼ぶ。一般的には「教師あり学習」と呼ばれる場合があるが、「教師推定問題」と区別するために本学位論文では「ラベルあり学習」と呼ぶことにする。制御の分野で入力と出力が与えられているとき、システムの内部状態を同定することもこの学習に含まれる。本学位論文では一貫してこのタイプの学習について議論する。
2. ラベルなし学習 ラベルあり学習に対して、入出力関係に欠損がある場合、つまり入力は与えられているがクラスラベルが与えられていない欠損データ (逆の場合もある) を用いての学習を「ラベルなし学習 (Unsupervised learning)」と呼ばれる。具体的な例として、いくつかの尺度に基づいて EM アルゴリズム (Expectation-Maximization algorithm) や変分ベイズ法 (Variational Bayes approach) などの近似アルゴリズムを用いて、欠損データを補完しながらモデルパラメータと平均値を推定する方法である [7, 42, 140, 173]。この問題は奥が深く、また実問題としても興味深い話題が多数あるが、これ以上は立ち入らない。

1.1.3 3つの論文

この小節では我々の研究に深く関係のする3つの論文について紹介する。特に解析的な操作は第2章や第3章で詳しく行うため、ここでは問題意識と解析結果、そしてそれらの論文による我々の研究との関係について触れる。

1. E. Gardner (1988)[71] 1988年にE. Gardnerはレプリカ法を用いてHopfieldモデルの記憶容量を解析した[71]。しかしながら彼女の導出した結果は、今日パーセプトロンと呼ばれる学習器の記憶容量に相当するため、ここではパーセプトロン研究の文脈で解説する。さて N 次元の入力と2値のクラスラベルをそれぞれ $\vec{x} := \{x_1, x_2, \dots, x_N\}^T \in \mathbf{R}^N$ と $y = \pm 1$ で表し、とくに $\vec{x}$ に対してまったく無関係にランダムなクラスラベルを割り当てた入出力関係 $(\vec{x}, y)$ を p 個、 $\xi^p := \{(\vec{x}_1, y_1), \dots, (\vec{x}_p, y_p)\}$ を用意する。また議論を簡単にするため各入力は各々独立同分布から生成されると仮定する。このとき、典型的に式 (1.21) で定義されるパーセプトロンでどのくらいのパターンを記憶することができるかという問題を考える。これは上述した記憶容量問題である。学習器の性能を示す指標として、パターン数 p をシステムサイズ N で割ったパターン比 $\alpha = p/N$ を用いることにする。ここでは N 次元の結合加重 $\vec{w} := \{w_1, w_2, \dots, w_N\}^T \in \mathbf{R}^N$ の各成分 w_k が連続値でかつ $\sum_{k=1}^N w_k^2 = N$ という拘束条件の下で解析を行う。これは直感的に、F. Rosenblatt のパーセプトロンアルゴリズム (もし線形分離可能な参照例題であれば、結合加重を式 (1.23) を用いて更新することで分離可能面の法線ベクトルである結合加重を求めることができるアルゴリズム) を用いて、典型的に記憶ができなくなる (もしくは線形分離不可能になる) パターン比 α_C を用いてパーセプトロンアルゴリズムの性能限界を特徴付

ることができる。彼女は $\alpha_C = 2$ であることをレプリカ法を用いて解析的に示した [71]。この評価値が組み合わせ論的に求めた T. M. Cover(1965) や S. S. Venkatesh(1986) などの先行研究 [39, 207] の解析結果と一致することに、当時の研究者は驚いた。

また [39, 207] の解析手法は厳密ではあるが、解析できるモデルが限られており、一方彼女の行った解析方法はある程度汎用性があったため、いくつかの学習問題の解析に自然に拡張することができる。たとえば W. Krauth and M. Mézard (1989) は結合加重が $w_k = \pm 1$ で等確率のときの記憶容量を $\alpha_C \simeq 0.833$ と求め [124]、ランダムラベルと同様に $\vec{w}_0$ で特徴付けられる教師ネットワークによりクラスラベルが与えられた場合、 w_k が離散値のとき G. Györgyi (1990) により [80]、 w_k が連続値のとき G. Györgyi and N. Tishby (1990) により解析結果が得られている [81]。上で挙げた既存結果については、我々の研究と直接関係するため、第3章で詳しく解説する。

ここで2点コメントする。第1に彼女が導入した手法は学習器の構造にある程度依存しない解析が可能であり、また Hopfield 模型と違い、スピングラスの数理模型を経由せず、統計的学習理論の立場で意味のある量、今日「Gardner 体積」と呼ばれる量、を直接評価したという点である。彼女のアプローチは当時として斬新な方法であり、機械学習に与えた影響は、E. Gardner (1988) の仕事に続く研究の数からも明らかである。次にこれまで上で挙げた先行研究も含め多くの場合、入力間に独立同分布を暗黙のうちに仮定しており、それについては多くの知見が得られている。一方、独立同分布以外の例題に対しての学習器の性能評価やアルゴリズムなどは、独立同分布に比べて知見が得られているとは言い難い。

2. E. Levin, N. Tishby and S. A. Solla (1990)[129] 1990年にそのあとの研究潮流に大きな影響を与える重要な論文が出版された [129]。これまで紹介してきた学習器や学習課題などに関連する「統計的学習理論」の枠組みと統計力学的手法の対応関係に対して、E. Levin, N. Tishby and S. A. Solla は以下の考察を行っている。彼らより以前でもたしかに、統計力学的手法を学習理論の問題に援用し解析を行う研究は存在したが、統計力学の枠組みを学習理論に持ち込んだときに得られた結果の正当性に対して、[129]の研究以前まで明確な根拠を与えられていなかった。しかしながら彼らは、統計的学習理論で用いる「Bayes 推論」とスピングラス理論で用いる「Gibbs 分布（本学位論文では Boltzmann 分布と呼ぶ）」が、構造的に同一の枠組みで記述されていることを指摘したのである。このときハミルトニアンは条件付き確率の対数のマイナスに比例した量として定義される。

$$\boxed{\text{統計力学}} \quad P(\vec{S}|J) = \frac{e^{-\beta \mathcal{H}(\vec{S}|J)}}{\sum_{\vec{S}} e^{-\beta \mathcal{H}(\vec{S}|J)}} \quad \Longleftrightarrow \quad \boxed{\text{Bayes 推定}} \quad P(\vec{w}|\xi^p) = \frac{P(\vec{w}) \prod_{\mu=1}^p P(y_{\mu}|\vec{x}_{\mu}, \vec{w})}{\sum_{\vec{w}} P(\vec{w}) \prod_{\mu=1}^p P(y_{\mu}|\vec{x}_{\mu}, \vec{w})}$$

さらに彼らは、既存研究で議論されていた「確率的複雑さ (stochastic complexity)」と自由エネルギーが対応することを指摘している⁹。当時の機械学習と統計力学の研究者に与えたインパクトは大きいものであったと推測される。この研究を受けて、また J. J. Hopfield[92] や G. E. Hinton and T. J. Sejnowski[91] らの研究もきっかけとなり、現在もなお機械学習と統計力学の両分野で活発に、学際領域的な研究が行われている。厳密な数字は不正確であるが、一説には [129] の研究以前の10年間で出版された領域横断的な研究に関する論文数とそれ以降の10年間の論文数を比較して、10倍にも20倍にも増えているという報告もある [179]。ただし計算機の向上や研究環境が整備されたため、直接的な因果関係が成立するわけではないが、それだけ学際領域に与えた影響力は大きいものであったと伺える。

3. M. Oppor and O. Winther (2001)[167] M. Oppor and O. Winther(2001) は Boltzmann マシンを用いて適応的に統計量を計算する手法を提案した。この手法では特に相互作用の従う確率分布が未知の場合でも解析できると

⁹E. Levin, N. Tishby and S. A. Solla (1990) は、アニール平均により確率的複雑さ f の漸近形を参照例題数を p としたとき $f \propto \frac{1}{p}$ であることを示しており、既存研究の結果と一致することを確認している [129]。

ころが重要である。彼らの手法の導き方は2通りあり、(1)Cavity法と呼ばれる、周辺化確率測度を Gaussian で近似する方法と(2)球形模型を用いる方法である。Cavity法は各々の確率変数に含まれる相関がある程度弱ければ、正攻法の解析手法と成り得るが、相互作用の統計性が分からない場合は数学的な保証がない。また計算過程で2次のモーメント(分散共分散行列のこと)や逆行列を計算する必要があり、システムサイズの3乗に比例する計算量が必要であるため実効的ではない。そこで我々は彼らの第2の方法である球形模型の方法を採用する。Sherrington-Kirkpatrick模型を含む、ハミルトニアンが2次形式でかつ、熱変数 $\vec{S} := \{S_1, \dots, S_N\}$ が $\sum_{k=1}^N S_k^2 = N$ の拘束条件が課された模型(これを球形模型と名付ける)に対するレプリカ解析は J. M. Kosterlitz, D. J. Thouless and R. C. Jones(1976)によって[123], 相互作用が固定された下での解析は B. W. Southern and A. P. Young(1977)によって初めて行われているが[190], M. Opper and O. Wintherはその解析手法に対して、(1) $\vec{S}$ の事前確率に任意性を持たせ、(2)また第2章で導出する $G(x)$ 関数と呼ばれる、相互作用の漸近固有値分布で定義される汎関数を用いて相互作用の相関の効果を取り込む形で、実効的な時間で統計量を計算するアルゴリズムを導出している。重要な点として、[123, 190]でも潜在的に $G(x)$ 関数に相当する関数を用いて解析を行っていたが、明示的に、特に一般の固有値分布に対する Onsager の反跳場を議論したのは初めてである。また我々の扱う問題に対する解析法の多くは彼らの仕事を雛形としている。

1.2 研究目的

これまで時代背景や研究対象の大まかな位置づけを概説してきたが、機械学習が対象としている学習問題の中で、我々は次の2点のことに注目した。

1. 機械学習の中でもっとも簡単なパーセプトロンにおいて、与えられた参照例題の入力間に暗黙のうちに独立同分布(independently and identically distributed)を仮定していることが数多くない。その理由として、入力 が各々独立である課題(examples)は一般的に言って、入力に相関が含まれる課題に比べて解析的に扱いやすく、また確率伝搬法(belief propagation)などの実効的なアルゴリズムが構成しやすいという利点が挙げられる。しかしながら実用の面で考えると、実データに対して多くの場合、入力の「独立同分布の仮定」の妥当性が保証されているとは言い難い。そのため独立同分布の仮定のままで実データに適用した場合、パーセプトロンを用いて各種の学習課題を学習した後のクラス識別における、「系統的な誤り(相関が含まれるパターンに対して独立パターンだと仮定することによる、モデルミスマッチングに起因する予測誤差に含まれる誤りなど)」が発生する可能性がある。つまりこのような状況では、i.i.d. パターンに対する学習器のクラス識別に対する知見やアルゴリズムなど、クラス識別器の性能を十分に評価できているとは言い難い。
2. またもう1つの例として、実データ解析において、入力が各々直交化する「前処理」をあらかじめデータに施す場合がある。問題にも依存するが、適切な「相関」を積極的に取り入れることで、クラス分類能力が向上することが知られているため、「前処理」がクラス分類能力に与える影響を評価する必要がある。

以上2点から、本学位論文では入力間に相関が含まれるデータに対する、パーセプトロンのデータ解析能力の典型的な性能評価や実効的なアルゴリズムの開発など理論的整備を行い、得られた知見を用いて情報理論の問題に適用し、提案手法の有効性を検証する。

1.3 アウトライン

この学位論文はこの章も含めて6つの章で構成される。第2章では本学位論文に関係するパーセプトロン学習とスピングラス理論とランダム行列理論の前提知識を説明する。また具体的な解析方法についてもいくつか紹介する。第3章では本研究との比較対象として、既存研究で議論されていた「独立同分布」に従うパターンに対するパーセプト

ロン学習の結果について簡単にまとめる。第4章では入力間に含まれる相関の特徴付けと定式化を行い、既存の解析手法の拡張を行う。その際いくつかの既存研究の結果が再現されることも示す。第5章では拡張された解析手法を用いて相関の含まれる例題の1つである「直交例題」についてパーセプトロンの典型的な問題である「記憶容量問題」と「教師推定問題」についての学習効率を検証する。その解析結果をサポートする数値実験もいくつか行っている。最後に第6章は全体のまとめと今後の抱負に当てられる。

第2章 予備知識

本章では我々が議論するテーマについての予備知識を詳しく説明する。本章の構成は以下の通りである。はじめに第1章で紹介した用語を用いて、パーセプトロン学習を議論するための数学的な準備を行う。次にスピングラス理論で開発された手法を2つ説明する。第3節では、ランダム行列理論、特に本学位論文で用いる積分公式を導出する。最後にそれらをまとめる。

2.1 パーセプトロン学習

近年確率推論を行う際の有力な方法である「Bayesの枠組み」が注目を集めており、我々はそれを先に説明する。ここでは第1章との重複を避けるため、簡潔に定式化のみを行う。対比する学習器や学習問題については第1章を参照されたい。

2.1.1 Bayes推論

Bayes推論 (Bayesian inference) とは、与えられた証拠 (evidence) を用いて確率的な現象の原因を推定するための方法論のことである [21, 22, 59, 136, 153]。具体的には事後確率を用いて行うため、まず事後確率について簡単に復習する。さて L 個の可能性を取り得る確率変数 $A(= a_1, a_2, \dots, a_L)$ と M 個の可能性を取り得る確率変数 $B(= b_1, b_2, \dots, b_M)$ を用意する (ここでは説明を簡単にするために確率事象と確率変数を同一視し、可算個であるとする)。事象 a_i の事前確率 $P(a_i)$ と、事象 a_i が起こった下で事象 b_j の発生する条件付き確率 $P(b_j|a_i)$ が与えられたとき、 b_j を観測したときの a_i の事後確率は次のように表される。

$$P(a_i|b_j) = \frac{P(b_j|a_i)P(a_i)}{\sum_{k=1}^L P(b_j|a_k)P(a_k)}. \quad (2.1)$$

ただし恒等式 $P(a_i, b_j) = P(a_i)P(b_j|a_i) = P(b_j)P(a_i|b_j)$ と周辺確率 $P(b_j) = \sum_{i=1}^L P(a_i, b_j)$, $P(a_i) = \sum_{k=1}^M P(a_i, b_k)$ を用いた。式 (2.1) は Bayes の定理 (Bayes' theorem) と呼ばれる。また変数が複数の場合、たとえば3つの確率変数 A, B, C を用意し、 $P(A, C) = P(A)P(C)$ の関係が成り立つとき、 B, C を観測したときの A の起こる事後確率は

$$P(A|B, C) = \frac{P(B|A, C)P(A, C)}{\sum_A P(B|A, C)P(A, C)} = \frac{P(B|A, C)P(A)}{\sum_A P(B|A, C)P(A)} \quad (2.2)$$

のように表される。推定方法の1つとして知られている最大事後確率 (Maximum A Posteriori probability, MAP) 推定は式 (2.1) の事後確率を用いて

$$a_i : = \max_{a_k} P(a_k|b_j) \quad (2.3)$$

によって決められる。しかしながら多くの問題において式 (2.3) は a_i の可能性が指数関数的に増えるため、いくつかの特殊な場合を除いて、計算量的に実行困難となる。このため、高性能の計算資源を用いても厳密な処理は一般に難しい。それに対して E. Levin, N. Tishby and S. A. Solla (1990) により指摘された、統計力学との対応関係を用いることで、実際の近似により推論が容易になり、確率推論の枠組みを広げることとなる。我々もパーセプトロン学習を確率推論の枠組みで定式化し、統計力学的解析を通して一定の知見が得られると期待される。

2.1.2 パーセプトロン

パーセプトロンは N 次元の入力 $\vec{x} := \{x_1, x_2, \dots, x_N\}^T \in \mathbf{R}^N$ に対して、次の非線形関数でクラスを割り振ることのできる学習器のことである。

$$y = \text{sgn} \left(\frac{1}{\sqrt{N}} \sum_{k=1}^N x_k w_k - \kappa \right) \quad (2.4)$$

ただし $\text{sgn}(u) = u/|u|$ は符号関数であり、線形分離面を特徴付ける法線ベクトル $\vec{w} := \{w_1, w_2, \dots, w_N\}^T$ を調整することで課題を熟すことができる。この法線ベクトルのことを結合加重ベクトル、もしくは単に結合加重（または混乱の生じない限り結合加重の各成分も結合加重）と呼ぶことにする。本学位論文では断りがない限り閾値 κ は 0 と見なす。さて p 個の参照例題 $\xi^p := \{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_p, y_p)\}$ が与えられたとき（ただし $\vec{x}_\mu = \{x_{1\mu}, x_{2\mu}, \dots, x_{N\mu}\}^T \in \mathbf{R}^N, (\mu = 1, 2, \dots, p)$ を用いた）、以下の学習問題における学習器の学習能力を評価することを考えよう。

2.1.3 記憶容量問題

記憶容量問題とは、各入力 $\vec{x}_\mu$ に対してまったく無関係に等確率でランダムにクラスラベル y_μ を割り当てたとき、どのくらい多くのパターン数を学習器、この場合パーセプトロンに覚えさせることができるか、という学習問題である。特に誤りなく入出力関係を再現できる最大のパターン数 p_C と入力次元 N の比（パターン比） $\alpha_C = p_C/N$ の典

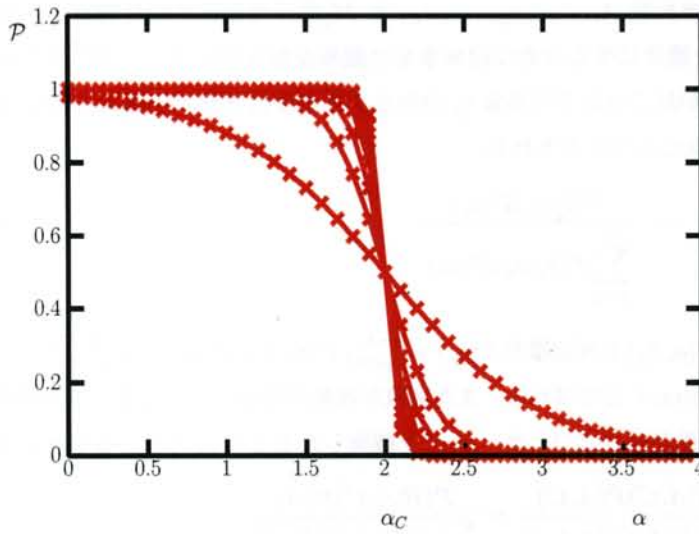


図 2.1: パターン比 α と記憶できる確率 P のグラフ。極端にパターン数が少ない場合は典型的に参照例題を学習器を用いて記憶することができる。一方極端に多い場合は記憶できる確率が典型的に 0 になることが予測される。つまり記憶できる確率 P があるパターン比 α_C を境に 1 から 0 に切り替わることが容易に予測できる。さらに次元数 N が大きくなるにつれて切り替わりが鋭敏になることが知られている。

型値により学習器の性能を特徴付けることができる。たとえば極端な状況として、パターン数が少ない場合は、典型的に参照例題 ξ^p を記憶することが可能であり、一方パターン数が極端に多い場合は典型的に参照例題を記憶することが困難になることが直感的に分かる。その典型的に「記憶可能・不可能」の性能が切り替わる α_C によって学習器の性能を表すことができ、また α_C が大きい学習器ほど、学習能力が高いことを示す（図 2.1）。

2.1.4 教師推定問題

教師推定問題とは、各入力 $\vec{x}_\mu$ に対して、 $\vec{w}_0$ で特徴付けられる教師ネットワーク（オラクル）により式 (2.4) を用いてクラスラベル y_μ が割り当てられたとき、どのくらい多くのパターン数を用いてパーセプトロンが教師ネットワークのベクトル $\vec{w}_0$ を同定することができるか、という学習問題である。

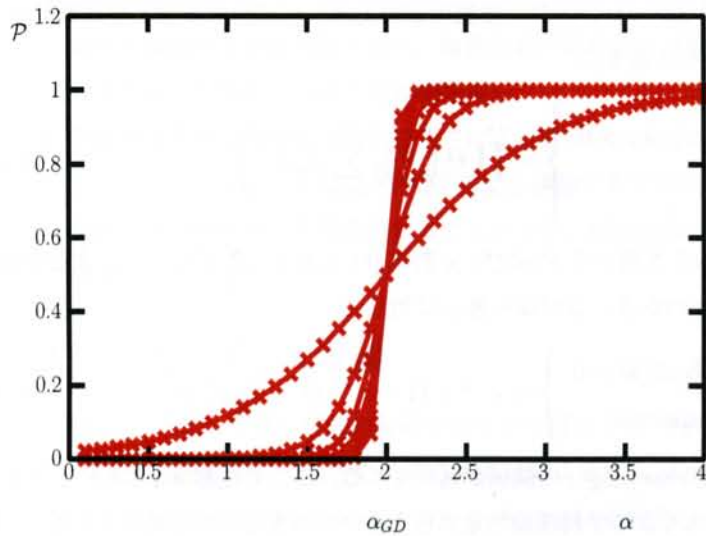


図 2.2: パターン比 α とオラクルを同定できる確率 P のグラフ。極端にパターン数が少ない場合は典型的に参照例題を学習器を用いて同定することは不可能である。一方極端に多い場合は同定できる確率が典型的に 1 になることが予測される。つまり同定できる確率 P があるパターン比 α_{GD} を境に 1 から 0 に切り替わることが容易に予測できる。さらにこの学習問題に対しても次元数 N が大きくなるにつれて切り替わりが鋭敏になることが知られている。

結合加重が離散値の場合 特にオラクルが同定できる最小のパターン数 p_{GD} と入力次元 N の比（パターン比） $\alpha_{GD} = p_{GD}/N$ の典型値により学習器の性能を特徴付けることができる。上記と同様に極端な状況として、パターン数が少ない場合は、典型的に $\vec{w}_0$ を同定することは不可能であり、一方パターン数が極端に多い場合は典型的にオラクルを同定することは可能であることが直感的に分かる。その典型的に「同定不可能・可能」の性能が切り替わる α_{GD} によって学習器の性能を表すことができ、また α_{GD} が小さい学習器ほど、学習能力が高いことを表す（図 2.2）。

結合加重が連続値の場合 一方結合加重が連続値の場合、変数の特性上オラクルを完全に同定することは難しい。そこで α_{GD} とは違う尺度を導入する。議論を簡単にするために結合加重 $\vec{w}$ を半径 $\sqrt{N}$ の N 次元ベクトル（ N 次元超球の表面に拘束されている）で表し、教師ベクトル $\vec{w}_0$ と生徒ベクトル $\vec{w}$ との内積、もしくは適当に規格化した内積を

$$q := \frac{\vec{w}^T \vec{w}_0}{|\vec{w}| |\vec{w}_0|} = \frac{1}{N} \vec{w}^T \vec{w}_0 \quad (2.5)$$

のように定義する。明らかに $|q| \leq 1$ であり、 $q = 1$ のとき $\vec{w} = \vec{w}_0$ を満たす。この q を用いて、完全同定 ($q = 1$) からのズレとして $\varepsilon = 1 - q$ を定義し、 ε のパターン比 α が大きい極限での漸近的な振る舞い¹を議論することで学習器の学習能力を評価することができる。

2.1.5 コスト関数の導入

p 個の入出力関係（参照例題） ξ^p を用意し、ある結合加重 $\vec{w}$ に着目する。クラス識別問題において μ 番目の入出力関係 $(\vec{x}_\mu, y_\mu)$ に対して $y_\mu \neq \text{sgn}\left(\frac{1}{\sqrt{N}} \vec{x}_\mu^T \vec{w}\right)$ のとき「誤り」であるので、 ξ^p に対する結合加重 $\vec{w}$ の誤りの個数（コスト関数）を次の表現

$$\text{Error}(\vec{w}|\xi^p) = \sum_{\mu=1}^p \Theta\left(-\frac{y_\mu}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} w_k\right) \quad (2.6)$$

で表すことができる。ただし階段関数 $\Theta(u) := \frac{1+\text{sgn}(u)}{2}$ を用いた。ここで $\text{Error}(\vec{w}|\xi^p) > 0$ となる結合加重 $\vec{w}$ は、学習問題（記憶容量問題や教師推定問題など）のタスクを達成できていないため、その確率を 0 になる表現形を用

¹ ε を α の関数であるため、学習曲線（learning curve）と呼ぶことにする。

意することができれば、上記の確率推論の枠組みでパーセプトロン学習の学習問題を定式化することが可能になる。ここではその方法を紹介する。そのため次の関数を定義する。

$$\mathcal{I}(\xi^p|\vec{w}) := \begin{cases} 1 & \forall \mu, \frac{y_\mu}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} w_k > 0 \\ 0 & \text{otherwise} \end{cases} = \prod_{\mu=1}^p \Theta \left(\frac{y_\mu}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} w_k \right) \quad (2.7)$$

これより結合加重 $\vec{w}$ の事後確率 $P(\vec{w}|\xi^p)$ は $\mathcal{I}(\xi^p|\vec{w})$ を用いて $P(\vec{w}|\xi^p) \propto \mathcal{I}(\vec{w}|\xi^p)$ と表すことができる。また定義から $\mathcal{I}(\xi^p|\vec{w}) = \lim_{\beta \rightarrow \infty} e^{-\beta \text{Error}(\vec{w}|\xi^p)}$ であることがわかる。なぜなら各 μ に対して

$$e^{-\beta \Theta \left(-\frac{y_\mu}{\sqrt{N}} \vec{x}_\mu^T \vec{w} \right)} = \begin{cases} 1 & \frac{y_\mu}{\sqrt{N}} \vec{x}_\mu^T \vec{w} > 0 \\ e^{-\beta} & \text{otherwise} \end{cases} = e^{-\beta} + (1 - e^{-\beta}) \Theta \left(\frac{y_\mu}{\sqrt{N}} \vec{x}_\mu^T \vec{w} \right) \quad (2.8)$$

となり、 $\beta \rightarrow \infty$ で事後確率 $P(\vec{w}|\xi^p)$ と誤り関数 $\text{Error}(\vec{w}|\xi^p)$ の関係が見えてくる。ここで注意を1つ行う。上記の枠組みは統計力学を意識してはいるが、ここではあくまで学習理論的定式化を用いて事後確率を定義している。つまり、誤り関数を用いて尤度関数、この場合 $e^{-\beta \Theta \left(-\frac{y_\mu}{\sqrt{N}} \vec{x}_\mu^T \vec{w} \right)}$ を定義することで、自然な形として事後確率が定式化でき、以下で説明する統計力学との関係もはっきりすることが分かる

2.1.6 事後確率と Boltzmann 分布

式 (2.2) を用いてパーセプトロン学習に限定せず、上記の学習問題に合わせて、

$$A \rightarrow \vec{w} \in \mathbf{R}^N, \quad B \rightarrow \vec{y} := \{y_1, y_2, \dots, y_p\} \in \mathbf{R}^p, \quad C \rightarrow X := \frac{1}{\sqrt{N}} \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_p\}^T \in \mathcal{M}_{p \times N} \quad (2.9)$$

とおけば、上記の典型的な学習問題に対する結合加重 $\vec{w}$ の事後確率は

$$P(\vec{w}|\vec{y}, X) = \frac{\prod_{\mu=1}^p P(y_\mu|\vec{x}_\mu, \vec{w}) P(\vec{w})}{\sum_{\vec{w}} \prod_{\mu=1}^p P(y_\mu|\vec{x}_\mu, \vec{w}) P(\vec{w})} \quad (2.10)$$

と定式化される。ただし $P(B|A, C) \propto \prod_{\mu=1}^p P(y_\mu|\vec{w}, \vec{x}_\mu)$ とし、取り扱う学習問題に依存して尤度関数 $P(y_\mu|\vec{x}_\mu, \vec{w})$ が定義される。式 (2.10) は第1章で指摘したとおり Boltzmann 分布と見なすことができ、統計力学的な手法を用いることが可能である。すなわち本学位論文ではハミルトニアンを与えられた参照例題 ξ^p の入出力関係を再現できないパターン数として

$$\mathcal{H}(\vec{w}|\xi^p) := \sum_{\mu=1}^p \Theta \left(-\frac{y_\mu}{\sqrt{N}} \vec{x}_\mu^T \vec{w} \right) \quad (2.11)$$

と定義する (F. Rosenblatt (1958) のパーセプトロンアルゴリズムなどを参照 [174])。ここで有限温度 ($\beta < \infty$) の場合 $P(y_\mu|\vec{x}_\mu, \vec{w}) \propto e^{-\beta \Theta \left(-\frac{y_\mu}{\sqrt{N}} \vec{x}_\mu^T \vec{w} \right)}$ であり、絶対零度 $\beta \rightarrow \infty$ の場合は $P(y_\mu|\vec{x}_\mu, \vec{w}) := \Theta \left(\frac{y_\mu}{\sqrt{N}} \vec{x}_\mu^T \vec{w} \right)$ である。有限温度はオラクルの信号が外的雑音により変化することを考慮に入れている場合であり、絶対零度は外的雑音がない場合をそれぞれ表す [55, 158]。

2.1.7 まとめ

本節では以下のことを説明した。

- Bayes 統計の枠組みにより、事後確率を用いて確率的な因果関係の推論を行うことができ、パーセプトロン学習を事後確率を用いて定式化することができる。
- 入力空間を結合加重を法線ベクトルとする超平面により分割することで、与えられた参照例題のクラスを識別する学習器であるパーセプトロンを定義した。

- 記憶容量問題とはクラスラベルがランダムに割り当てられた参照例題をどのくらいの数を記憶することのできるかを議論する学習問題であり、結合加重が離散値と連続値のどちららにおいても、典型的なパターン比により学習能力を特徴付けることができる。
- 教師推定問題とは教師ネットワークによってクラスラベルが割り当てられた参照例題をどのくらいの数で教師ネットワークを同定することができるかを議論する学習問題であり、結合加重が離散値の場合、典型的なパターン比により学習能力を特徴付けることができ、結合加重が連続値の場合は、学習曲線により学習能力を特徴付けることができる。

2.2 スピングラス理論

この節ではスピングラス理論の解析手法を2つ紹介する。特に第1章と重複する部分もあるが、新しい視点を導入する。また議論を簡単にするために次の小節以降 J. M. Kosterlitz, D. J. Thouless and R. C. Jones (1976) [123] や B. W. Southern and A. P. Young (1977) [190] の手法を拡張した, M. Oppen and O. Winther (2001)[167] による議論に従ってレプリカ法 (2.2.1 小節) と Thouless-Anderson-Palmer (TAP) 法 (2.2.2 小節) を説明する。2.2.1 小節のレプリカ法と 2.2.2 小節の TAP 法は大自由度複雑系を解析する手続きとして、いくつかの既存研究により有効で強力な方法であることが示されているが、すべての場合に対してうまくいくわけではなく、一般に長所と短所があることが知られている [55, 63, 138, 158, 162, 163]。両手法の適応範囲を詳細に議論することはこの学位論文のレベルを超えるため省略する。気になる方は [55, 63, 138, 158, 162, 163] などを参照されたい。

2.2.1 レプリカ法

ここでは具体的な表現の一つを用いてレプリカ法を説明する。

Sherrington-Kirkpatrick 模型のレプリカ解析 さてより教育的で基本的な模型として, Sherrington-Kirkpatrick (SK) 模型を取り上げる。\$N\$ 次元のイジングスピン \$\vec{S} := \{S_1, S_2, \dots, S_N\}^T \in \{\pm 1\}^N\$ のハミルトニアンを次のように定義する。

$$\mathcal{H}(\vec{S}|J) := -\frac{1}{2} \vec{S}^T J \vec{S} - \vec{h}^T \vec{S}, \quad (2.12)$$

ただし外部磁場と相互作用をそれぞれ \$\vec{h} := \{h_1, h_2, \dots, h_N\}^T \in \mathbf{R}^N, (h_k = h)\$ と \$J := \{J_{ij}\} \in \mathcal{M}_{N \times N}, (J_{ij} = J_{ji}, J_{ii} = 0)\$ で表す。また \$J_{ij}\$ は各々独立に平均 \$J_0/N\$, 分散 \$J^2/N\$ の正規分布に従うと仮定する。便宜上相互作用 \$J\$ の確率測度を \$P(J)\$ で表し、分配関数 \$Z\$ の \$n(\in \mathbf{N})\$ 次モーメントは

$$\begin{aligned} [Z^n] &= \sum_J P(J) \sum_{\{\vec{S}_a\}} \exp \left[-\beta \sum_{a=1}^n \mathcal{H}(\vec{S}_a|J) \right] \\ &\simeq \sum_{\{\vec{S}_a\}} \exp \left[\frac{N\beta^2 J^2}{2} \sum_{a>b} \left(\frac{1}{N} \sum_{k=1}^N S_{ka} S_{kb} \right)^2 + \frac{N\beta J_0}{2} \sum_{a=1}^n \left(\frac{1}{N} \sum_{k=1}^N S_{ka} \right)^2 + N\beta h \sum_{a=1}^n \left(\frac{1}{N} \sum_{k=1}^N S_{ka} \right) + \frac{Nn\beta^2 J^2}{4} \right] \\ &\simeq \text{Extr}_{\theta, \tilde{\theta}} \left\{ \sum_{\{\vec{S}_a\}} \exp \left[\frac{N\beta^2 J^2}{2} \sum_{a>b} q_{ab}^2 + \frac{N\beta J_0}{2} \sum_{a=1}^n m_a^2 + N\beta h \sum_{a=1}^n m_a + \frac{Nn\beta^2 J^2}{4} \right. \right. \\ &\quad \left. \left. + \sum_{a=1}^n \tilde{m}_a \left(\sum_{k=1}^N S_{ka} - Nm_a \right) + \sum_{a>b} \tilde{q}_{ab} \left(\sum_{k=1}^N S_{ka} S_{kb} - Nq_{ab} \right) \right] \right\} \end{aligned} \quad (2.13)$$

となる²。ただし a, b は $a, b = 1, 2, \dots, n$ となるレプリカ添え字を表し、秩序変数を (定義する関数 $\theta(\vec{S})$ を用いて) 次のように定義する。

$$m_a = \frac{1}{N} \sum_{i=1}^N S_{ia}, \quad q_{ab} = \frac{1}{N} \sum_{i=1}^N S_{ia} S_{ib}. \quad (2.14)$$

また Legendre 変換で用いる共役な秩序変数を $\tilde{m}_a, \tilde{q}_{ab}$ とする。式 (2.13) において $\theta := \{m_a, q_{ab}\}$ と $\tilde{\theta} := \{\tilde{m}_a, \tilde{q}_{ab}\}$ とおいた。 $N \rightarrow \infty$ の極限で式 (2.13) は、積分に寄与する部分自由エネルギーで近似することができるため、

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log [Z^n] = \text{Extr}_{\theta, \tilde{\theta}} \left\{ \frac{\beta^2 J^2}{4} \text{Tr} Q^2 + \frac{\beta J_0}{2} \vec{m}^T \vec{m} + \beta \vec{h}^T \vec{m} - \vec{m}^T \vec{m} - \frac{1}{2} \text{Tr} \tilde{Q} Q + \log \sum_{\vec{S}} e^{\vec{S}^T \vec{m} + \frac{1}{2} \vec{S}^T \tilde{Q} \vec{S}} \right\} \quad (2.15)$$

となる。ただし n 次正方行列 $Q := \{q_{ab}\}$, $\tilde{Q} := \{\tilde{q}_{ab}\} \in \mathcal{M}_{n \times n}$, ($q_{aa} = 1, \tilde{q}_{aa} = 0$) と n 次元ベクトル $\vec{h} := \{h, h, \dots, h\}^T$, $\vec{m} := \{m_1, m_2, \dots, m_n\}^T$, $\vec{\tilde{m}} := \{\tilde{m}_1, \tilde{m}_2, \dots, \tilde{m}_n\}^T \in \mathbf{R}^n$ を用いた。ここで再度注意するが、Legendre 変換のところで指摘したように秩序変数 $\theta, \tilde{\theta}$ は実数 (もしくは共役な変数同士の掛け算が実数、たとえば $m_a \tilde{m}_a \in \mathbf{R}$) でかつ、定義域がそれぞれ凸集合でなければならない。de Almeida-Thouless(AT) 解析に代表されるように [5], また以下で説明する TAP 法でも威力を発揮するが、この指摘はいくつかの問題で解候補を選択するための指標になり、我々も多くの困難を解くためのヒントとして用いていることができる。

さて式 (2.15) において適当なレプリカ添え字 a, b の入れ替え ($a \leftrightarrow b$) に対して関数の中身の表現が変わらないことから、上記の凸最適化問題の最適解の解表現として、すべてのレプリカ添え字 a, b に対して $m_a = m, \tilde{m}_a = \tilde{m}, q_{ab} = q, \tilde{q}_{ab} = \tilde{q}$ となる対称解が、最適解の 1 つの候補になりうるということが知られている。このように変数を仮定することを「レプリカ対称仮定 (replica symmetry assumption)」と呼び、最適解を満たす $m, q, \tilde{m}, \tilde{q}$ をレプリカ対称解 (RS 解) と呼ぶ。これより 1 自由度あたりの自由エネルギーは

$$-\beta f := \text{Extr}_{m, q, \tilde{m}, \tilde{q}} \left\{ \frac{\beta^2 J^2}{4} - \frac{\beta^2 J^2}{2} q^2 + \frac{\beta J_0}{2} m^2 + \beta h m - \tilde{m} m - \frac{\tilde{q}}{2} (1 - q) + \int_{-\infty}^{\infty} Dz \log 2 \cosh \left(\tilde{m} + z \sqrt{\tilde{q}} \right) \right\}, \quad (2.16)$$

となる。ただし $Dz := \frac{dz e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$ は正規分布の確率測度を表す。

球形モデルのレプリカ解析 次に J. M. Kosterlitz, D. J. Thouless and D. J. Jones (1976) の扱った球形モデルを紹介する [123]。議論を簡単にするために $\vec{S}^T \vec{S} = N$ を満たす N 次元ベクトル $\vec{S} \in \mathbf{R}^N$ とし、ハミルトニアンは式 (2.12) を用い、ここでは $h = 0$ とする。相互作用を自明な置換を除いて

$$J := \{J_{ij}\} = O D O^T \in \mathcal{M}_{N \times N} \quad (2.17)$$

と分解できる。ただし直交行列 $O \in \mathcal{M}_{N \times N}$, ($O O^T = O^T O = I$) と固有値 λ_k , ($k = 1, 2, \dots, N$) で定義される対角行列 $D = \text{diag} \{\lambda_k\} \in \mathcal{M}_{N \times N}$ を用いた。ここでは相互作用のアンサンブルとして、固有値分布

$$\rho(\lambda) := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \delta(\lambda - \lambda_k) = \lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr} \delta(\lambda - J) \quad (2.18)$$

が一定のランダムネスの集合における典型的な振る舞いを議論する。新たな変数 $\vec{t}_a := O^T \vec{S}_a$ を定義すると、以下の関係が成り立つ。

$$\vec{t}_a^T \vec{t}_b = \vec{S}_a^T \vec{S}_b = N \quad (2.19)$$

² $\{\vec{S}_a\} := \{\vec{S}_1, \vec{S}_2, \dots, \vec{S}_n\}$ とする。

この関係を用いて $\vec{S}_a$ での平均を適当に座標を取り直した $\vec{t}_a$ による平均で置き換えることができる。分配関数は事前分布 $P(\vec{S}) \propto \delta(\vec{S}^T \vec{S} - N)$ を用いて $Z(\beta) := \int d\vec{S} P(\vec{S}) e^{-\beta \mathcal{H}(\vec{S})}$ と定義されることから分配関数の $n (\in \mathbf{N})$ 乗は

$$\begin{aligned} Z^n(\beta) &:= \int \prod_{a=1}^n d\vec{S}_a P(\vec{S}_a) \sum_{\{\vec{t}_a\}} P(\{\vec{t}_a\} | \{\vec{S}_a\}) \exp\left(-\beta \sum_{a=1}^n \mathcal{H}(\vec{S}_a | J)\right) \\ &\propto \text{Extr}_{\Lambda, \tilde{Q}} \int \prod_{a=1}^n \frac{d\vec{S}_a d\vec{t}_a}{(2\pi)^n} \exp\left[-\frac{1}{2} \sum_{a=1}^n \Lambda_a \left(\sum_{k=1}^N t_{ka}^2 - N\right) + \sum_{a>b} \tilde{q}_{ab} \sum_{k=1}^N (t_{ka} t_{kb} - S_{ka} S_{kb}) + \frac{\beta}{2} \sum_{a=1}^n \vec{t}_a^T D \vec{t}_a\right] \\ &= \text{Extr}_{\Lambda, \tilde{Q}} \exp\left\{-\frac{1}{2} \sum_{k=1}^N \log \det(\Lambda - \tilde{Q} - \beta \lambda_k I) + \frac{N}{2} \text{Tr} \Lambda - \frac{N}{2} \log \det \tilde{Q}\right\} \end{aligned} \quad (2.20)$$

となる。式 (2.20) は直交行列 O に依存しない表現のため $Z^n(\beta) = [Z^n(\beta)]$ であることが分かる。これより

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \log [Z^n(\beta)] &= \text{Extr}_{\Lambda, \tilde{Q}} \left\{ -\frac{1}{2} \int d\lambda \rho(\lambda) \log \det(\Lambda - \beta \lambda I) + \frac{1}{2} \text{Tr} \Lambda + \frac{1}{2} \text{Tr} \tilde{Q} - \frac{1}{2} \log \det \tilde{Q} \right\} + \text{const.} \\ &= \text{Extr}_{\Lambda} \left\{ \frac{1}{2} \text{Tr}(\Lambda - I) - \frac{1}{2} \int d\lambda \rho(\lambda) \log \det(\Lambda - \beta \lambda I) \right\}. \end{aligned} \quad (2.21)$$

となる。定義から $Z^n(0) = 1$ であり、また $\Lambda - I$ と $\Lambda - \beta \lambda I$ の固有ベクトルが一致するため、

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log [Z^n(\beta)] = nG(\beta) \quad (2.22)$$

となる。ただし補助変数 $\Lambda \in \mathbf{R}$ を用いて

$$G(x) := \text{Extr}_{\Lambda} \left\{ \frac{\Lambda - 1}{2} - \frac{1}{2} \int d\lambda \rho(\lambda) \log(\Lambda - x\lambda) \right\} \quad (2.23)$$

で定義される関数 $G(x)$ を用いた。これよりクエンチ平均は

$$-\beta f = \lim_{N \rightarrow \infty} \frac{1}{N} [\log Z(\beta)] = G(\beta) \quad (2.24)$$

となる。この球形模型はレプリカ法が厳密な解を再現する模型の1つとして知られている [43, 44]。また本研究のテクニックの1つとして、この球形模型のように被積分関数の指数部が2次形式もしくは双1次形式のとき、ガウス積分が実行できるという事実を用いている。

SK 模型の一般化 最後に SK 模型の一般化を行う。 N 次元のベクトル $\vec{e} := \{1, 1, \dots, 1\}^T \in \mathbf{R}^N$ を用いて、相互作用を $J := \frac{J_0}{N} (\vec{e}\vec{e}^T - I) + JODO^T$ と定義する³。これよりレプリカ系のハミルトニアンは

$$\begin{aligned} -\beta \sum_{a=1}^n \mathcal{H}(\vec{S}_a | J) &= \frac{\beta}{2} \sum_{a=1}^n \vec{S}_a^T \left(\frac{J_0}{N} (\vec{e}\vec{e}^T - I) + JODO^T \right) \vec{S}_a + \beta h \vec{e}^T \sum_{a=1}^n \vec{S}_a \\ &\simeq \frac{N\beta J_0}{2} \sum_{a=1}^n m_a^2 + \frac{\beta J}{2} \sum_{k=1}^N \lambda_k \sum_{a=1}^n t_{ka}^2 + N\beta h \sum_{a=1}^n m_a \end{aligned} \quad (2.25)$$

となる。よって今までの議論から

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log [Z^n(\beta)] = \text{Extr}_{\vec{m}, \vec{m}, Q, \tilde{Q}} \left\{ \frac{\beta J_0}{2} \vec{m}^T \vec{m} + \beta \vec{h}^T \vec{m} + \mathcal{G}(\beta J Q) - \vec{m}^T \vec{m} - \frac{1}{2} \text{Tr} Q \tilde{Q} + \log \sum_{\vec{s}} e^{\vec{s}^T \vec{m} + \frac{1}{2} \vec{s}^T \tilde{Q} \vec{s}} \right\} \quad (2.26)$$

となる。ここで行列 $D \in \mathcal{M}_{N \times N}$ の固有値分布 $\rho(\lambda)$ を用いて

$$\mathcal{G}(\chi) := \text{Extr}_{\Lambda} \left\{ \frac{1}{2} \text{Tr}(\Lambda - I) - \frac{1}{2} \int d\lambda \rho(\lambda) \log \det(\Lambda - \chi \lambda) \right\} \quad (2.27)$$

³ $\vec{e}\vec{e}^T - I$ と ODO^T はランク N の行列であり、お互い自由独立である [202]。

を定義する。ただし $\chi, \Lambda \in \mathcal{M}_{n \times n}$ とした。さらにコメントとして、 $\Lambda - I$ と $\Lambda - \chi\lambda$ が同時対角化される、つまり χ の固有ベクトル方向にとると、 Tr と $\det$ は固有ベクトル方向で分解されるため、 χ の n 個の固有値 $\chi_a, (a = 1, 2, \dots, n)$ を用いて

$$\mathcal{G}(\chi) = \sum_{a=1}^n G(\chi_a) \quad (2.28)$$

が得られる [195]。特に SK 模型では固有値分布が

$$\rho(\lambda) = \begin{cases} \frac{\sqrt{2^2 - \lambda^2}}{2\pi} & |\lambda| \leq 2 \\ 0 & \text{otherwise} \end{cases} \quad (2.29)$$

であり

$$\mathcal{G}(\chi) = \frac{1}{4} \text{Tr} \chi^2 \quad (2.30)$$

となり、式 (2.15) が再現される。固有値分布 (式 (2.29)) と $G(\chi)$ 関数 (式 (2.30)) の導出は 2.3 節で行う。

2.2.2 Thouless-Anderson-Palmer 法

レプリカ法は相互作用に関して平均的なシステムの性質を調べる手法と位置付けることができる。ここではある個別の相互作用 $J := \{J_{ij}\} \in \mathcal{M}_{N \times N}$ と外部磁場 $\vec{h} := \{h_1, h_2, \dots, h_N\} \in \mathbf{R}^N$ が与えられた下で、特に N が十分大きい、有限の場合に成り立つ、別の理論的な方法を紹介する。この方法はもともと D. J. Thouless, P. W. Anderson and R. G. Palmer (1977) によって初めて導入された方法であり [198], Thouless-Anderson-Palmer 法、もしくは単に TAP 法と呼ぶことにする。しかしながらここではオリジナルの方法ではない M. Oppen and O. Winther (2001) が導出した方法を説明する [167]。オリジナルの導出方法は相互作用の性質を独立な正規乱数に限定しており、汎用性の低い導出方法であり、一方 M. Oppen and O. Winther の方法は正規分布に限定していないため、我々は後者の方法を採用する。さらに M. Oppen and O. Winther の議論は相互作用を定行列 $\vec{e}\vec{e}^T - I \in \mathcal{M}_{N \times N}$ と固有値分布が $\rho(\lambda)$ で定義されるランダム行列 $ODO^T \in \mathcal{M}_{N \times N}$ を用いて $J = \{J_{ij}\} = \frac{J_0}{N} (\vec{e}\vec{e}^T - I) + JO DO^T \in \mathcal{M}_{N \times N}$ で表して高次の「相関」を特徴付けられており、我々のパーセプトロン学習の議論に自然に援用することができる点でよい。

さてハミルトニアン $\mathcal{H}(\vec{S}|J)$ は式 (2.12) で定義されているとする。TAP 法の議論の前に、重要なコメントとして、まず十分大きな N に対しては、個別の J で特徴付けられるシステムの 1 自由度あたりの自由エネルギーは、1 自由度あたりの典型的な自由エネルギーに確率 1 で確率収束することが知られており、レプリカ法の結果と対応することが期待できる [32]。次に統計力学の観点から言えば、Helmholtz 自由エネルギーは直接評価することが、いくつかの特殊な場合を除いて難しいことが知られているため、我々は「Gibbs 自由エネルギー」、つまり Helmholtz 自由エネルギーの Legendre 変換を用いて、さらに凸最適化の知見を十分に活用し、定式化を行う。

導出方法は以下の通りである。まず $\vec{m} := \{m_1, m_2, \dots, m_N\}^T \in \mathbf{R}^N$ と $\chi \in \mathbf{R}$ のそれぞれに Legendre 変換の共役な変数として $\vec{\tilde{m}} := \{\tilde{m}_1, \tilde{m}_2, \dots, \tilde{m}_N\}^T \in \mathbf{R}^N$ と $\tilde{\chi} \in \mathbf{R}$ を採用する。ここで上記のコメントより Gibbs 自由エネルギーを制御変数 ε を用いて、

$$Ng(\varepsilon, \vec{m}, \chi) := \text{Extr}_{\vec{\tilde{m}}, \tilde{\chi}} \left\{ \log \sum_{\vec{S}} e^{-\beta \varepsilon \mathcal{H}(\vec{S}) + \vec{\tilde{m}}^T (\vec{S} - \vec{m}) - \frac{\tilde{\chi}}{2} (\vec{S}^T \vec{S} - N\chi - \vec{m}^T \vec{m})} \right\}. \quad (2.31)$$

と定義し、それを用いて Helmholtz 自由エネルギーは

$$-\beta f = \lim_{N \rightarrow \infty} \frac{1}{N} \text{Extr}_{\vec{\tilde{m}}, \chi} Ng(1, \vec{m}, \chi), \quad (2.32)$$

と表される。ここで注意として、定義より任意の ε において、 $\vec{m}$ と χ の凸集合上で $Ng(\varepsilon, \vec{m}, \chi)$ は凸関数である。加えて $\varepsilon = 0$ を除いて式 (2.31) の $Ng(\varepsilon, \vec{m}, \chi)$ を評価することは難しいために、我々は次の 2 つの性質を満たす関数 $N\tilde{g}(\varepsilon, \vec{m}, \chi)$ を導入し、その困難を回避する。

関数 $N\tilde{g}(\varepsilon, \vec{m}, \chi)$ の要請

1. N の大きい極限で次の関係式が成り立つ。

$$N \frac{\partial g(\varepsilon, \vec{m}, \chi)}{\partial \varepsilon} \simeq N \frac{\partial \tilde{g}(\varepsilon, \vec{m}, \chi)}{\partial \varepsilon} \implies Ng(1, \vec{m}, \chi) - Ng(0, \vec{m}, \chi) \simeq N\tilde{g}(1, \vec{m}, \chi) - N\tilde{g}(0, \vec{m}, \chi) \quad (2.33)$$

2. $N\tilde{g}(\varepsilon, \vec{m}, \chi)$ は任意の ε に対して解析的に求まる関数である。

式 (2.33) の意味として、 $Ng(\varepsilon, \vec{m}, \chi)$ と $N\tilde{g}(\varepsilon, \vec{m}, \chi)$ は $\vec{m}$ と χ の関数として展開することができ、具体的に言えば、両関数の ε の微分係数 (絶対温度 $1/(\beta\varepsilon)$ の内部エネルギーを表す) は N の大きい極限で漸近的に一致することを要請する。さらに両辺を $\varepsilon = 0$ から $\varepsilon = 1$ まで積分すると関係式 (2.33) が導出される。加えて 2 の条件から、任意の ε に対して解析的に求まる関数として次の関数を定義する。

$$N\tilde{g}(\varepsilon, \vec{m}, \chi) := \text{Extr}_{\vec{m}, \tilde{\chi}} \left\{ \log \int_{-\infty}^{\infty} \frac{d\vec{S}}{(2\pi)^{\frac{N}{2}}} e^{-\beta\varepsilon \mathcal{H}(\vec{S}) + \vec{m}^T (\vec{S} - \vec{m}) - \frac{\tilde{\chi}}{2} (\vec{S}^T \vec{S} - N\chi - \vec{m}^T \vec{m})} \right\}, \quad (2.34)$$

定義から Gauss 積分に帰着されているため、式 (2.31) の計算よりも計算が簡単になることは明らかである。これより

$$N\tilde{g}(\varepsilon, \vec{m}, \chi) = \beta\varepsilon \vec{h}^T \vec{m} + \frac{\beta\varepsilon}{2} \vec{m}^T J \vec{m} + N \text{Extr}_{\tilde{\chi}} \left\{ \frac{\tilde{\chi}\chi}{2} - \frac{1}{2} \int d\lambda \rho(\lambda) \log(\tilde{\chi} - \beta\varepsilon\lambda) + O\left(\frac{1}{N^\gamma}\right) \right\} \quad (2.35)$$

となる。ただし十分大きな N での補正項は $\gamma > 0$ を用いて、 $O(N^{-\gamma})$ と表した。よって十分大きな N において主要項

$$N\tilde{g}(1, \vec{m}, \chi) - N\tilde{g}(0, \vec{m}, \chi) = \beta \vec{h}^T \vec{m} + \frac{\beta}{2} \vec{m}^T J \vec{m} + NG(\beta\chi) \quad (2.36)$$

が得られる。 $G(x)$ は前の小節で定義したが、詳しくは次の小節のランダム行列理論で取り扱うことにし、 $\Lambda := \chi\tilde{\chi}$ とおいた。一方 $Ng(0, \vec{m}, \chi)$ は解析的に評価できるため、

$$Ng(0, \vec{m}, \chi) = \log 2 \sum_{k=1}^N \mathcal{H}\left(\frac{1+m_k}{2}\right) + N \text{Extr}_{\tilde{\chi}} \frac{\tilde{\chi}}{2} \left(\chi + \frac{\vec{m}^T \vec{m}}{N} - 1 \right), \quad (2.37)$$

となる。ただし Shannon の 2 値エントロピー関数 $\mathcal{H}(p) := -p \log_2 p - (1-p) \log_2 (1-p)$ を用いた。結論として

$$Ng(1, \vec{m}, \chi) \simeq \beta \vec{h}^T \vec{m} + \frac{\beta}{2} \vec{m}^T J \vec{m} + NG(\beta\chi) + \log 2 \sum_{k=1}^N \mathcal{H}\left(\frac{1+m_k}{2}\right) + N \text{Extr}_{\tilde{\chi}} \frac{\tilde{\chi}}{2} \left(\chi + \frac{\vec{m}^T \vec{m}}{N} - 1 \right) \quad (2.38)$$

が導出される。個別の相互作用で定義される経験分布は、 N が大きい極限で漸近固有値分布に確率 1 で漸近することから、上記のレプリカ解析の結果と一致することが期待される (自己平均性という)。また実際のアルゴリズムは表 2.1 に示す。

2.2.3 まとめ

本節では大自由度複雑系の解析の雛型として代表的なレプリカ法と TAP 法について議論し、以下のことが解明された。

1. ランダム系の強力な解析手法の 1 つとして、レプリカ解析を紹介した。またそれと対をなすものとして、TAP 法を導出した。両手法とも大偏差解析と Legendre 変換 (凸解析) により数学的にサポートされていることも解説した。
2. 本学位論文で導入する手法の架け橋として、[123, 190] を拡張した M. Oppen and O. Winther の方法を解説した [167]。

TAP アルゴリズム

極値方程式は以下のようになる。

$$m_i = \tanh \left(\beta \sum_{j=1}^N J_{ji} m_j + \beta h_i + \tilde{\chi} m_i \right), \quad (2.39)$$

$$\chi = 1 - \frac{1}{N} \sum_{k=1}^N m_k^2, \quad (2.40)$$

$$\tilde{\chi} = - \left\langle \frac{\beta \lambda}{\Lambda - \beta \chi \lambda} \right\rangle, \quad (2.41)$$

$$\Lambda = 1 + \left\langle \frac{\beta \chi \lambda}{\Lambda - \beta \chi \lambda} \right\rangle = 1 - \chi \tilde{\chi}, \quad (2.42)$$

$$\langle A(\lambda) \rangle = \int d\lambda \rho(\lambda) A(\lambda). \quad (2.43)$$

フローチャート

同期更新でかつ逐次過緩和法 (Successive Over Relaxation, SOR) などを用いて実効的な時間で解析可能である。

$$\vec{m}, \tilde{\chi} \rightarrow \vec{m}$$

$$\vec{m} \rightarrow \chi$$

$$\Lambda, \chi \rightarrow \tilde{\chi}$$

$$\Lambda, \chi \rightarrow \Lambda$$

表 2.1: $\beta, \vec{h}, J$ が与えたとき、上記のアルゴリズムを用いて平均磁化 m_i を求めることができる。

2.3 ランダム行列理論

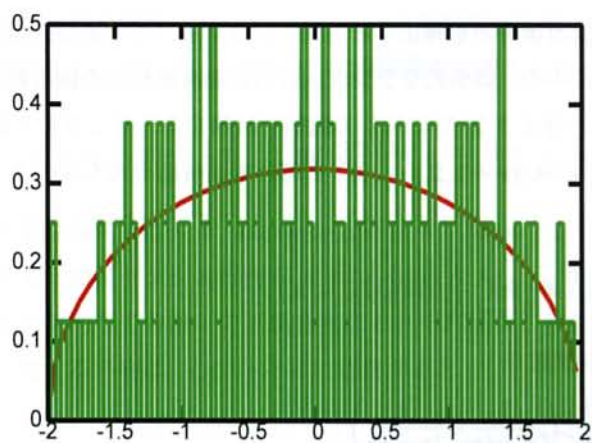
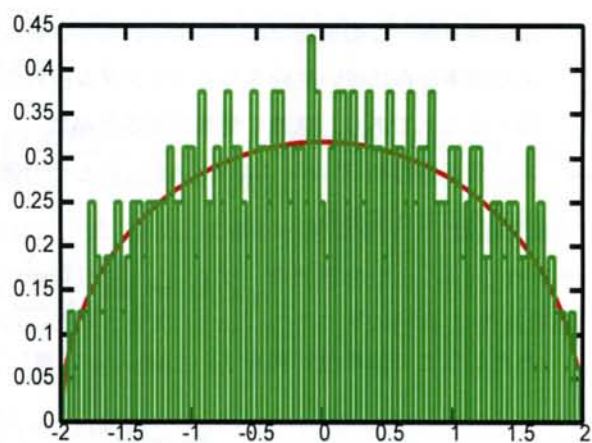
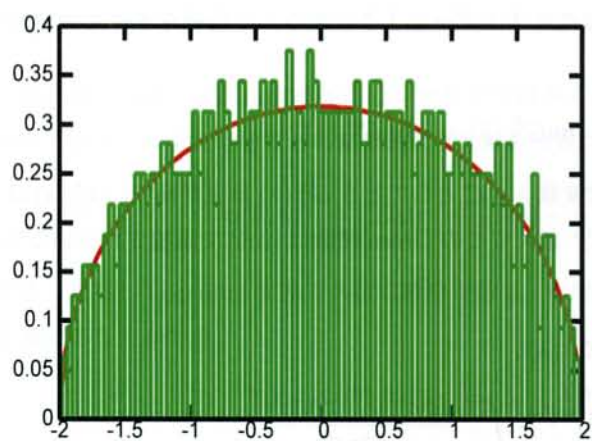
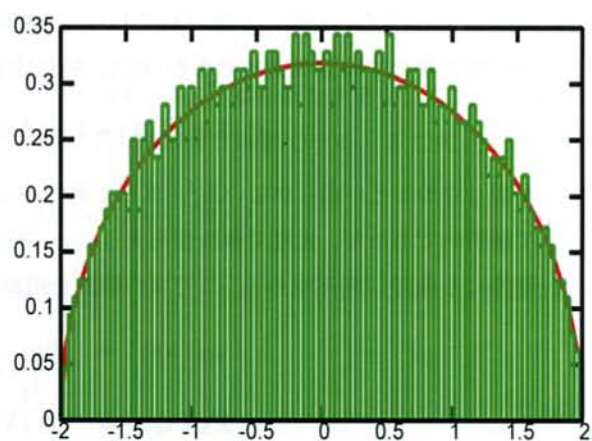
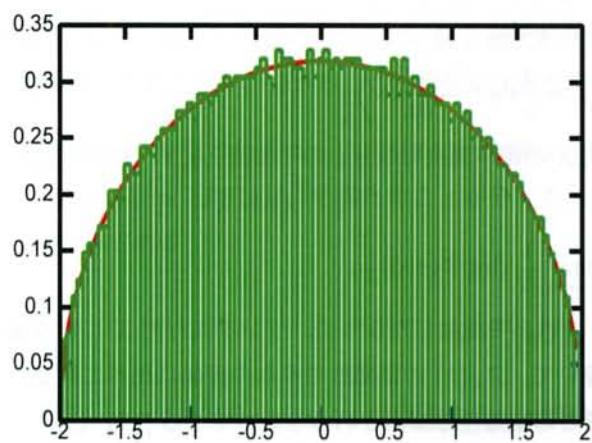
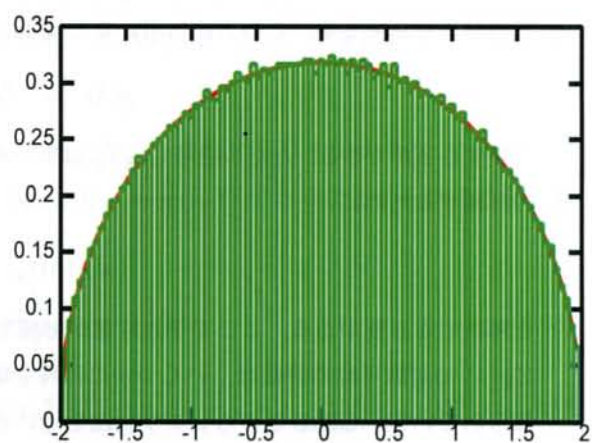
ランダム行列とは、確率的に生成された成分を持つ行列のことであり、この行列は数理統計学、原子核物理学、量子カオス、通信理論と最適化問題に関連する研究分野などで活発に研究されており、数学的にも興味深く、もともと魅力的な研究課題の一つである。ランダム行列理論において、いくつかの重要な性質が解明されており、特に我々の研究に関連し、J. Wishart (1928)[214], E. Wigner (1955)[215] や V. A. Marčenko and L. A. Pastur (1967)[132] に代表される、ランダム行列の固有値分布は、現在も活発に研究が行われている。この節では、既存の解析手法を自然に拡張することで得られる、我々の提案手法に繋げるために、ランダム行列の集合に対する典型的な固有値分布を導出し、その固有値分布で定義される積分公式を紹介する。この節の構成は以下の通りである。次の小節でランダム行列の固有値分布の性質について説明する。さらに 2.3.2 小節でランダム正方行列に対する積分公式を議論する。この積分公式は C. Itzykson and J. B. Zuber[107] により最初に導出されたものであり、我々が開発したものではないが、重要であるため簡単に触れる。またこの公式は前章のレプリカ法や TAP 法で用いた $G(x)$ 関数と関連することも説明する。続いて 2.3.3 小節では、ランダム長方行列に対する積分公式を導出する。これは我々が解析する問題に対する強力な重要な道具であり、詳しく解説する。

2.3.1 ランダム行列の固有値分布

多くの先行研究によりランダム行列の固有値分布のさまざまな性質が明らかになっている。特に Wigner らにより行列サイズ N が十分大きな、ランダム正方行列の固有値分布の普遍性が存在することが解明されている [215]。詳しく言えば $X := \left\{ \frac{x_{ij}}{\sqrt{N}} \right\} \in \mathcal{M}_{N \times N}$ で、対角成分 x_{ii} が 0 に固定され、非対角成分 $x_{ij} = x_{ji}$ が平均 0 で分散 1 の独立同分布にそれぞれ従い (ただし前因子 $\sqrt{N}$ は $N \rightarrow \infty$ でも有効に働くためのスケールである)、さらに $\sqrt{N} x_{ij}$ の $m(> 2)$ 次モーメントが有界であるとき、このランダム行列 X の固有値分布 $\rho(\lambda)$ は次の半径 2 の Wigner の半円則に漸近することが知られている。

$$\rho(\lambda) = \begin{cases} 0 & |\lambda| \geq 2 \\ \frac{1}{2\pi} \sqrt{4 - \lambda^2} & \text{otherwise} \end{cases} \quad (2.44)$$

図 2.3 から図 2.8 はそれぞれ $N = 200$ から $N = 6400$ のランダム行列の経験固有値分布を実際数値的に導き、Wigner の半円則と比較してみた。 N が大きくなるに従い、Wigner の半円則に漸近していることが見て取れる。

図 2.3: $N = 200$ 図 2.4: $N = 400$ 図 2.5: $N = 800$ 図 2.6: $N = 1600$ 図 2.7: $N = 3200$ 図 2.8: $N = 6400$

固有値分布 この小節では、与えられたランダム行列の固有値分布を導出する方法を議論する。ただ最終的な目的はある確率分布に従い生成される「ランダム行列の集合」の中の、典型的なランダム行列の固有値分布の振る舞いを議論することであり、議論を簡単にするために、ここでは十分大きな N に対して説明を展開する。さて与えられた N 次元の正方行列 $X := \left\{ \frac{x_{ij}}{\sqrt{N}} \right\} \in \mathcal{M}_{N \times N}$ とその固有値 $\lambda_i \in \mathbf{R}, (i = 1, 2, \dots, N)$ を用いて固有値分布を次のように定義することができる。

$$\rho_X(\lambda) := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta(\lambda - \lambda_i) = \lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr} \delta(\lambda I_N - X),$$

ただしここでは単位行列を $I_N \in \mathcal{M}_{N \times N}$ で表し、デルタ関数を

$$\delta(x) := \lim_{\varepsilon \rightarrow +0} \frac{1}{2\pi i} \left(\frac{1}{x - i\varepsilon} - \frac{1}{x + i\varepsilon} \right).$$

で表すことができる。ここで2種類の Green 関数を定義する。

$$G_X^R(\lambda) := \lim_{\varepsilon \rightarrow +0} \lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr} ((\lambda + i\varepsilon) I_N - X)^{-1}, \quad G_X^A(\lambda) := \lim_{\varepsilon \rightarrow +0} \lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr} ((\lambda - i\varepsilon) I_N - X)^{-1}, \quad (2.45)$$

$G_X^R(\lambda)$ と $G_X^A(\lambda)$ はそれぞれ物理学において、遅延 Green 関数 (retarded Green function) と先進 Green 関数 (advanced Green function) と呼ばれており、分配関数や自由エネルギーのように扱う系の情報をすべて含んだ母関数の役割を果たす重要な関数である [157]。特に定義から

$$\text{Re} G_X^R(\lambda) = \text{Re} G_X^A(\lambda), \quad \text{Im} G_X^R(\lambda) = -\text{Im} G_X^A(\lambda), \quad (2.46)$$

が成り立つことが知られている⁴。さらに両 Green 関数は変数 $\lambda \in \mathbf{C}$ としてとき $G_X^R(\lambda)$ は上半面で、 $G_X^A(\lambda)$ は下半面で解析的であるが、実軸で解析性を失うことに注意 [103]。加えてこの Green 関数は [202] で議論されている Stieltjes 変換にも密接に関係する。ここで次の関数を用意する。

$$\phi_X(\lambda) := \lim_{N \rightarrow \infty} \frac{1}{N} \log Z_X(\lambda), \quad (2.47)$$

$$Z_X(\lambda) := \det(\lambda I_N - X)^{-\frac{1}{2}} = \int_{-\infty}^{\infty} \frac{d\vec{S} e^{-\frac{1}{2} \vec{S}^T (\lambda I_N - X) \vec{S}}}{(2\pi)^{\frac{N}{2}}}, \quad (2.48)$$

このとき $G_X^R(\lambda) = -2 \frac{\partial \phi_X(\lambda)}{\partial \lambda}$ となる。ただし式 (2.47) と式 (2.48) において議論を簡単にするために引数に含まれる $+i\varepsilon$ は省略してある。これによりランダム行列 X によって定義される固有値分布は

$$\rho_X(\lambda) = -\frac{1}{\pi} \text{Im} G_X^R(\lambda), \quad (2.49)$$

で表すことができる。この方法を用いると典型的な固有値分布は

$$\rho(\lambda) = \langle \rho_X(\lambda) \rangle_X = \frac{2}{\pi} \text{Im} \left\langle \frac{\partial \phi_X(\lambda)}{\partial \lambda} \right\rangle_X \quad (2.50)$$

で表すことができる。ただし記号 $\langle A_X \rangle_X$ はランダム行列 X の確率測度での A_X の平均値を表す。最終的に式 (2.50) の評価するために

$$\langle \phi_X(\lambda) \rangle_X = \lim_{N \rightarrow \infty} \frac{1}{N} \langle \log Z_X(\lambda) \rangle_X, \quad (2.51)$$

を解析する必要がある。ここで1点注意が必要である。これまで議論したように、ランダム行列 X に従う確率変数 $Z_X(\lambda)$ の対数平均の評価は、いくつかの特殊な場合を除いて、 $Z_X(\lambda)$ の冪関数の平均の評価に比べて難しいことが知られている。しかしながらレプリカ法を用いることでこの困難性は解消されると期待できる [138, 158]。また数理構造的に前章で説明した球形模型と対応するため、 N が十分大きい極限で $\langle Z_X^n(\lambda) \rangle_X \simeq \langle Z_X(\lambda) \rangle_X^n$ が成り立つことも直感的に理解できるのだが、これについては後ほど説明する。

⁴ $z, \varepsilon \in \mathbf{R}$ としたとき $\text{Re}(z - i\varepsilon)^{-1} = \text{Re}(z + i\varepsilon)^{-1}$ が成り立つ。

Wigner の半円則 これより先レプリカ法を用いて Wigner の半円則を導出する方法を説明する。まず任意の $n \in \mathbf{N}$ における $\log \langle Z_X^n(\lambda) \rangle_X$ を用いて次の関数を定義することができる。

$$\psi(\lambda, n) := \lim_{N \rightarrow \infty} \frac{1}{N} \log \langle Z_X^n(\lambda) \rangle_X = \lim_{N \rightarrow \infty} \frac{1}{N} \log \int_{-\infty}^{\infty} \prod_{a=1}^n \left(\frac{d\vec{S}_a e^{-\frac{\lambda}{2} \vec{S}_a^T \vec{S}_a}}{(2\pi)^{\frac{N}{2}}} \right) \times \left\langle e^{\frac{1}{2} \text{Tr} X (\sum_{a=1}^n \vec{S}_a \vec{S}_a^T)} \right\rangle_X \quad (2.52)$$

この学位論文では S. F. Edwards and R. C. Jones [52] の仕事に従い、式 (2.52) の解析的評価を行う。議論を簡単にするためにランダム行列を $X := \left\{ \frac{x_{ij}}{\sqrt{N}} \right\} \in \mathcal{M}_{N \times N}$ で表し、対角成分が $x_{ii} = 0$ で、非対角成分 $x_{ij} = x_{ji}$ が各々独立に平均 0 で分散 1 の正規分布にそれぞれ従うことを仮定する。このとき式 (2.52) のランダム行列 X の確率測度に対する平均は、 $O(N)$ だけに着目すると

$$\log \left\langle \exp \left[\frac{1}{2} \text{Tr} X \left(\sum_{a=1}^n \vec{S}_a \vec{S}_a^T \right) \right] \right\rangle_X \simeq \frac{N}{4} \text{Tr} Q^2, \quad (2.53)$$

となる。ただしここでは各成分が秩序変数で特徴付けられる n 次対称行列 $Q := \{q_{ab}\} \in \mathcal{M}_{n \times n}$ をすでに用いた。このときの秩序変数は

$$q_{ab} = \frac{1}{N} \sum_{i=1}^N S_{ia} S_{ib}. \quad (2.54)$$

と一意に表される。大自由度複雑系のシナリオでは多くの場合レプリカ対称 (RS) 解を仮定するが、ここでの解析においては RS 解を仮定せずとも結果的に RS 解が安定であることが以下の説明より理解できる。さて式 (2.52) は

$$\psi(\lambda, n) = \text{Extr}_{\tilde{Q}, \tilde{Q}} \left\{ -\frac{\lambda}{2} \text{Tr} \tilde{Q} + \frac{1}{4} \text{Tr} \tilde{Q}^2 - \frac{1}{2} \log \det \tilde{Q} + \frac{1}{2} \text{Tr} \tilde{Q} \tilde{Q} \right\} \quad (2.55)$$

となる。ただし Legendre 変換で共役な秩序変数 $\tilde{q}_{ab}$ を用いて $\tilde{Q} := \{\tilde{q}_{ab}\} \in \mathcal{M}_{n \times n}$ とおいた。式 (2.55) は簡単に極値評価することができ、

$$Q = \tilde{Q}^{-1} = \frac{\lambda \pm \sqrt{\lambda^2 - 2^2}}{2} I_n, \quad (2.56)$$

となる。ただし $I_n \in \mathcal{M}_{n \times n}$ は単位行列を表す。これより式 (2.52)、式 (2.55) と式 (2.56) から

$$\left\langle \frac{\partial \phi_X(\lambda)}{\partial \lambda} \right\rangle_X = \lim_{n \rightarrow 0} \left\langle \frac{\partial^2 \psi(\lambda, n)}{\partial \lambda \partial n} \right\rangle = \frac{-\lambda \pm \sqrt{\lambda^2 - 2^2}}{4} \quad (2.57)$$

が得られる。ちなみに $G_X^R(\lambda)$ の定義式 (2.45) と式 (2.50) から十分大きな λ に対して $G_X^R(\lambda) \sim \frac{1}{\lambda}$ が成り立つため、自然に式 (2.55) の 2 つの解のうち正しい解を選択することができる。これより式 (2.44) の半径が 2 の Wigner の半円則が導出されたことが分かる。ここで 4 点注意を行う。

1. 式 (2.56) からわかるように式 (2.55) の解は非対角成分を持たない。これにより十分大きな N に対して $\langle Z_X(\lambda) \rangle_X^n \simeq \langle Z_X^n(\lambda) \rangle_X$ を満たすことが分かる。さらに $\langle \phi_X(\lambda) \rangle_X = \psi(\lambda, 1)$ であることから次のことも簡単に示される。

$$\rho(\lambda) = -\frac{1}{\pi} \text{Im} \left\langle \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N S_k^2 \right\rangle_{\lambda, X} = -\frac{1}{\pi} \text{Im} q, \quad (2.58)$$

ただし $n=1$ のとき Q を q として置き換え、最適解は $q = \frac{\lambda - i\sqrt{2^2 - \lambda^2}}{2}$ である。ここで $P_{\lambda, X}(\vec{S}) \propto e^{-\frac{1}{2} \vec{S}^T (\lambda I_N - X) \vec{S}}$ と $\langle A(\vec{S}) \rangle_{\lambda, X} := \int d\vec{S} \langle P_{\lambda, X}(\vec{S}) \rangle_X A(\vec{S})$ を用いた。ところで $\langle Z_X(\lambda) \rangle_X^n \simeq \langle Z_X^n(\lambda) \rangle_X$ の関係は新しい知見を提供するが、それは以下の項目で説明する。

2. もし確率変数 $Z_X(\lambda)$ の確率測度が N が十分大きい極限でデルタ関数と見なすことができれば, $\langle Z_X(\lambda) \rangle_X^n \simeq \langle Z_X^n(\lambda) \rangle_X$ は自明である. つまり自己平均性が成り立っていれば $\langle Z_X(\lambda) \rangle_X^n \simeq \langle Z_X^n(\lambda) \rangle_X$ が得られる. また逆も成り立つ. すなわち $\langle Z_X(\lambda) \rangle_X^n \simeq \langle Z_X^n(\lambda) \rangle_X$ が成り立っていれば, 確率変数 $Z_X(\lambda)$ の累積分布関数のタイトな上界評価で定義される「誤り指数」により, $N \rightarrow \infty$ で確率測度がデルタ関数に漸近することが数学的に保証されているため, 結論としてこの模型に関しては, 正しい解を採用できれば, レプリカ法で求まる解は厳密な解を導出することができることが裏付けられる.
3. 一般的に, 有効なスケーリングを施したランダム行列の各成分 x_{ij} が, 各々独立で1次モーメントと2次モーメントがそれぞれ0と1で, さらに $\sqrt{N}x_{ij}$ の高次のモーメントが有界である任意のランダム行列のアンサンブルに対しても, 式 (2.53) の評価において N が大きい極限で高次モーメントの効果が漸近的に弱まることから, 上記の議論が成り立ち, $N \rightarrow \infty$ の極限で半径2のWignerの半円則を導出することができる. つまりこれはランダム行列理論の「中心極限定理」と考えることもできる性質である [202].
4. 式 (2.55) を満たす極値を求めるためには次の鞍点方程式 $Q = (\lambda I_n - Q)^{-1}$ を解かなければならない. Q の対角化を用いれば簡単化されるが, 2次方程式の解のように2つのうちのどちらかを最終的に選ばなければならない. 幸いなことに任意のランダム行列において $|\lambda|$ が十分大きいところで $G_X^R(\lambda) \sim \frac{1}{\lambda}$ になるため, Q のすべての固有値が一致し, 式 (2.56) が得られる. さらにもし解析的な方法で鞍点方程式を解くことができない場合, 今回の方程式を用いて説明すれば, 2種類の反復法 $Q_{t+1} = (\lambda I_n - Q_t)^{-1}$ と $Q_{t+1} = \lambda I_n - Q_t^{-1}$ のどちらかを用いて数値的にとかなければならない. ただし Q_t は反復回数 t の Q を表す. つまり一般的な鞍点方程式 $x = f(x)$ が与えられたとき, $x^{t+1} = f(x^t)$ で解くのか, その逆関数の $x^{t+1} = f^{-1}(x^t)$ を用いて解くのかで, 両方の反復の引き込み領域が λ の値, もしくは $f(x)$ の関数形によって入れ替わることが分かる. 所望な解を求める際には反復法の設計にも注意が必要である.

Marčenko-Pastur 則 $p \times N$ のランダム長方形行列 $X := \left\{ \frac{x_{\mu k}}{\sqrt{N}} \right\} \in \mathcal{M}_{p \times N}$ に対しても上記のような固有値分布の議論をすることができる. つまり自己相関行列 $X^T X \in \mathcal{M}_{N \times N}$ の固有値分布を

$$\rho_{X^T X}(\lambda) = \lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr} \delta(\lambda I_N - X^T X) \quad (2.59)$$

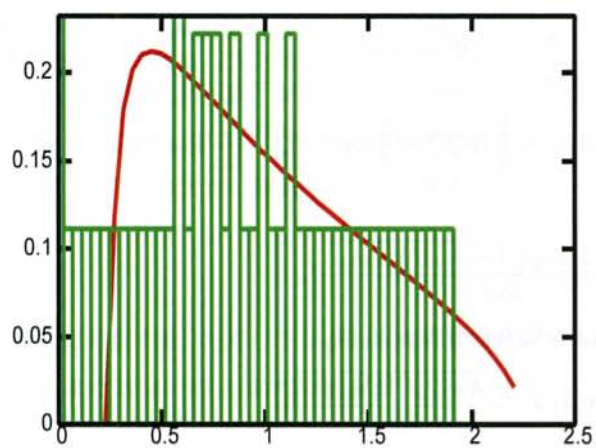
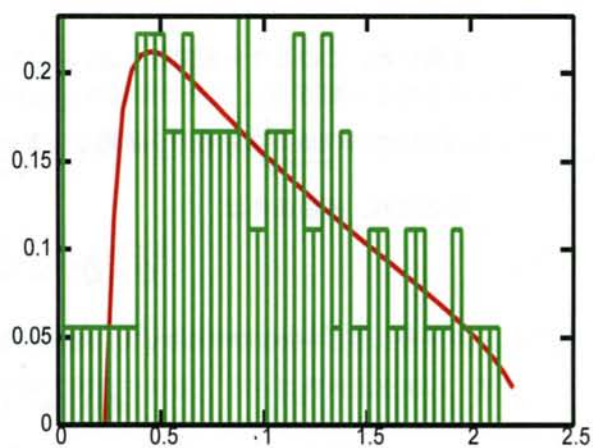
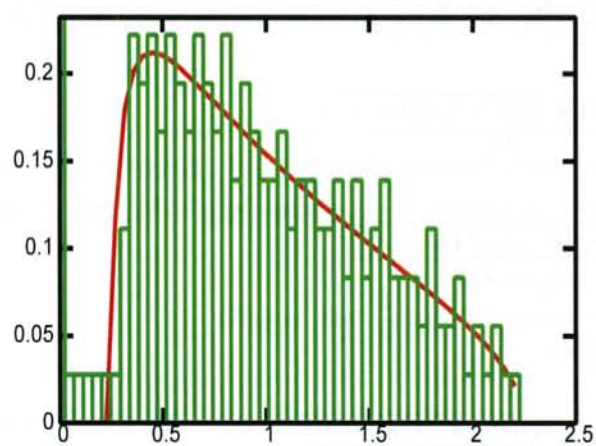
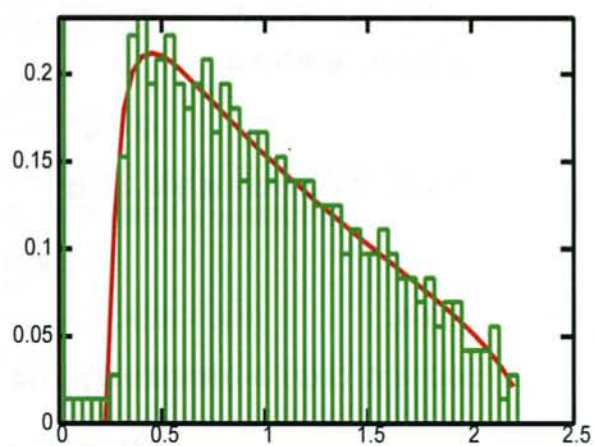
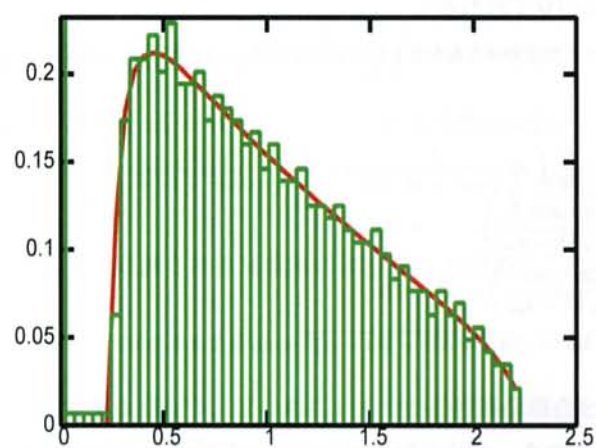
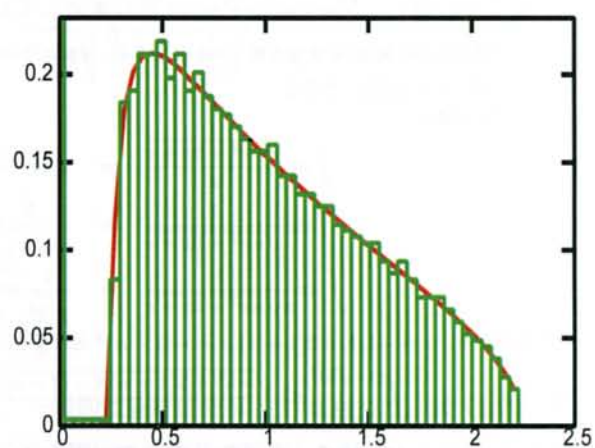
と定義できる. これより $\psi(\lambda, n) = \lim_{N \rightarrow \infty} \frac{1}{N} \log \langle Z_{X^T X}^n(\lambda) \rangle_X$ と $Z_{X^T X}(\lambda) = \int_{-\infty}^{\infty} \frac{d\vec{s}}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2} \vec{s}^T (\lambda I_N - X^T X) \vec{s}}$ を用いることで固有値分布を導出することができる. 議論を簡単にするためにランダム行列の成分 $x_{\mu k}$ を各々独立に平均0で分散1の正規分布に従うとし, このランダム行列のアンサンブルに対する典型的な固有値分布についてここで議論してみよう. さてWignerの半円則と同様の手順で固有値分布を導出することができるが, 正方行列と違い, 長方形行列の縦横の比 $\alpha = p/N$ の大小で場合わけが必要である.

1. $p < N \rightarrow \alpha < 1$ のとき, $X^T X$ のランクが小さいほうの p に依存するため, N 個の固有値の中で $N - p$ 個は固有値0である. つまり固有値分布は

$$\begin{aligned} \rho_{X^T X}(\lambda) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \delta(\lambda - \lambda_k) = \lim_{N \rightarrow \infty} \frac{1}{N} \left\{ \sum_{k=1+p}^N \delta(\lambda - 0) + \sum_{\mu=1}^p \delta(\lambda - \lambda_{\mu}) \right\} \\ &= (1 - \alpha) \delta(\lambda) + \alpha \lim_{p \rightarrow \infty} \frac{1}{p} \sum_{\mu=1}^p \delta(\lambda - \lambda_{\mu}) = (1 - \alpha) \delta(\lambda) + \alpha \tilde{\rho}(\lambda) \end{aligned} \quad (2.60)$$

となる. ただし

$$\tilde{\rho}(\lambda) := \lim_{p \rightarrow \infty} \frac{1}{p} \text{Tr} \delta(\lambda - X X^T) \quad (2.61)$$

図 2.9: $N = 200$ 図 2.10: $N = 400$ 図 2.11: $N = 800$ 図 2.12: $N = 1600$ 図 2.13: $N = 3200$ 図 2.14: $N = 6400$

を用いた。このとき分配関数を $Z_{XX^T}(\lambda) := \int_{-\infty}^{\infty} \frac{d\vec{S}}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2}\vec{S}^T(\lambda - XX^T)\vec{S}}$ とおくと、 $\tilde{\rho}(\lambda)$ の生成母関数は

$$\psi(\lambda, n) := \lim_{p \rightarrow \infty} \frac{1}{p} \log \langle Z_{XX^T}^n(\lambda) \rangle_X = \text{Extr}_{Q, \tilde{Q}} \left\{ -\frac{\lambda}{2} \text{Tr} Q + \frac{1}{2} \text{Tr} Q \tilde{Q} - \frac{1}{2} \log \det \tilde{Q} - \frac{1}{2\alpha} \log \det(I_n - \alpha Q) \right\} \quad (2.62)$$

と表され、極値条件は

$$Q = \tilde{Q}^{-1} = \frac{\lambda + \alpha - 1 \pm \sqrt{(\lambda - \lambda_-)(\lambda - \lambda_+)}}{2\alpha\lambda} I_n \quad (2.63)$$

となる。ただし $\lambda_{\pm} := (1 \pm \sqrt{\alpha})^2$ を用いた。これより固有値分布 $\rho(\lambda)$ は

$$\rho(\lambda) = (1 - \alpha)\delta(\lambda) + \frac{\sqrt{[\lambda - \lambda_-]^+[\lambda_+ - \lambda]^+}}{2\pi\lambda} \quad (2.64)$$

となる。ちなみに記号 $[x]^+ := \lim_{L \rightarrow \infty} \frac{1}{L} \log(1 + e^{Lx})$ を用いた。

2. $p > N \rightarrow \alpha > 1$ のとき、 $X^T X$ のランクは小さいほうの N に依存するため、固有値 0 を考える必要がない。そこで $\rho(\lambda)$ の生成母関数は

$$\psi(\lambda, n) := \lim_{N \rightarrow \infty} \frac{1}{N} \log \langle Z_{X^T X}^n(\lambda) \rangle = \text{Extr}_{Q, \tilde{Q}} \left\{ -\frac{\lambda}{2} \text{Tr} Q + \frac{1}{2} \text{Tr} Q \tilde{Q} - \frac{1}{2} \log \det \tilde{Q} - \frac{\alpha}{2} \log \det(I_n - Q) \right\} \quad (2.65)$$

と表され、極値条件は

$$Q = \tilde{Q}^{-1} = \frac{\lambda - \alpha + 1 \pm \sqrt{(\lambda - \lambda_-)(\lambda - \lambda_+)}}{2\lambda} I_n \quad (2.66)$$

となる。これより固有値分布 $\rho(\lambda)$ は

$$\rho(\lambda) = \frac{\sqrt{[\lambda - \lambda_-]^+[\lambda_+ - \lambda]^+}}{2\pi\lambda} \quad (2.67)$$

となる。

式 (2.64) と式 (2.67) から固有値分布 $\rho(\lambda)$ は

$$\rho(\lambda) = [1 - \alpha]^+ \delta(\lambda) + \frac{\sqrt{[\lambda - \lambda_-]^+[\lambda_+ - \lambda]^+}}{2\pi\lambda} \quad (2.68)$$

となる⁵。これは Marčenko-Pastur 則と呼ばれる分布である。この分布の普遍性や解の性質については上記の Wigner の半円則の注意事項と同様であるためここでは省略する。最後に図 2.9 から図 2.14 はそれぞれ $N = 200$ から $N = 6400$ の $\alpha = 1/4$ のランダム長方形行列の固有値分布を実数値的に導き、Marčenko-Pastur 則と比較して見た。 N が大きくなるに従い、Marčenko-Pastur 則に漸近していることが見て取れる⁶。

⁵ 図 2.9 から図 2.14 までを見て分かるように、右のプロードなところで固有値分布が最大になる固有値は $\lambda_o = \frac{(1-\alpha)^2}{1+\alpha} = 1 + \alpha - \frac{4\alpha}{1+\alpha}$ であり、 $\rho(\lambda_o) = \frac{\sqrt{\alpha}}{\pi|1-\alpha|}$ となる。

⁶ 不定積分

$$\int dx \frac{\sqrt{ax^2 + bx + c}}{x} = \sqrt{ax^2 + bx + c} + \frac{b}{2} \int dx \frac{1}{\sqrt{ax^2 + bx + c}} + c \int dx \frac{1}{x\sqrt{ax^2 + bx + c}} \quad (2.69)$$

$$\int dx \frac{1}{\sqrt{ax^2 + bx + c}} = -\frac{1}{\sqrt{|a|}} \sin^{-1} \left(\frac{2ax + b}{\sqrt{b^2 - 4ac}} \right), \quad a < 0 \quad (2.70)$$

$$\int dx \frac{1}{x\sqrt{ax^2 + bx + c}} = \frac{1}{\sqrt{|c|}} \sin^{-1} \left(\frac{2c + bx}{x\sqrt{b^2 - 4ac}} \right), \quad c < 0 \quad (2.71)$$

を用いる。 $a = -1, b = 2(1 + \alpha), c = -(1 - \alpha)^2$ とすると、 $\sqrt{b^2 - 4ac} = \sqrt{4(1 + \alpha)^2 - 4(1 - \alpha)^2} = 4\sqrt{\alpha}$ であるので、

$$\begin{aligned} & \int_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} d\lambda \frac{\sqrt{-\lambda^2 + 2(1+\alpha)\lambda - (1-\alpha)^2}}{2\pi\lambda} \\ &= \frac{(1+\alpha)}{2\pi} \left[\sin^{-1} \left(\frac{2(1+\alpha) - 2\lambda}{4\sqrt{\alpha}} \right) \right]_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} - \frac{|1-\alpha|}{2\pi} \left[\sin^{-1} \left(\frac{2(1+\alpha)\lambda - 2(1-\alpha)^2}{4\sqrt{\alpha}\lambda} \right) \right]_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} = \frac{1+\alpha - |1-\alpha|}{2} \\ &= \min(1, \alpha) \end{aligned} \quad (2.72)$$

が得られる。

2.3.2 ランダム正方行列の積分公式

N 次元ベクトル $\vec{w} \in \mathbf{R}^N$ が固定され, 「対称な正方行列 $X \in \mathcal{M}_{N \times N}$ の固有値分布 $\rho(\lambda)$ が同一となるランダム行列のアンサンブル」で $e^{\frac{1}{2}\vec{w}^T X \vec{w}}$ の平均評価することを考える. ここでは議論を簡単にするためにランダム正方行列 X の次元 N を十分大きい極限で $O(1)$ になる

$$G(x) := \lim_{N \rightarrow \infty} \frac{1}{N} \log \sum_X P(X) e^{\frac{1}{2}\vec{w}^T X \vec{w}} \quad (2.73)$$

を具体的に考察する. ただしランダム行列のアンサンブル平均を確率測度 $P(X)$ での平均で表すことにし, $Nx = \vec{w}^T \vec{w}$ とする. また対称行列 X なので固有値 λ_k , ($k = 1, 2, \dots, N$) はすべて実数である. さらに対角化を用い, X は

$$X := O D O^T, \quad (2.74)$$

と分解できる. ただし対角行列と回転行列をそれぞれ $D := \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_N\} \in \mathcal{M}_{N \times N}$ と $O \in \mathcal{M}_{N \times N}$ で表す (自明な置換を除いて一意的に定まる). 固有値分布が一定であるアンサンブルで平均化するということは, Haar 測度 (ランダム直交行列のアンサンブル) に従うランダム直交行列 O によって生成される $X = O D O^T$ により平均化することと等価である. さて新しい座標系を $\vec{t} = O^T \vec{w}$ とすると次の関係式

$$\vec{w}^T \vec{w} = \vec{t}^T \vec{t} = Nx \quad (2.75)$$

を満たすことが分かる. つまりランダム行列のアンサンブルで平均を行うことは, N の大きい極限において $\vec{t}^T \vec{t} = Nx$ を満たす $\vec{t}$ で $e^{\frac{1}{2}\vec{t}^T D \vec{t}}$ を積分評価することに相当する. これより式 (2.73) は新しい座標系では

$$G(x) = \lim_{N \rightarrow \infty} \frac{1}{N} \log \frac{\int_{-\infty}^{\infty} d\vec{t} \delta(\vec{t}^T \vec{t} - Nx) e^{\frac{1}{2}\vec{t}^T D \vec{t}}}{\int_{-\infty}^{\infty} d\vec{t} \delta(\vec{t}^T \vec{t} - Nx)} = \mathcal{A}(D) - \mathcal{A}(0) \quad (2.76)$$

と表すことができる. ここで

$$\mathcal{A}(D) := \text{Extr}_{\Gamma} \lim_{N \rightarrow \infty} \frac{1}{N} \log \int_{-\infty}^{\infty} \frac{d\vec{t}}{(2\pi)^{\frac{N}{2}}} e^{-\frac{\Gamma}{2}(\vec{t}^T \vec{t} - Nx) + \frac{1}{2}\vec{t}^T D \vec{t}} = \text{Extr}_{\Gamma} \left\{ \frac{x\Gamma}{2} - \frac{1}{2} \int d\lambda \rho(\lambda) \log(\Gamma - \lambda) \right\} \quad (2.77)$$

となり, また式 (2.76) の被積分関数の分母を, 固有値分布が $\rho(\lambda) = \delta(\lambda)$ のときの $\mathcal{A}(D=0)$ に対応すると考えれば, $\mathcal{A}(0) = \text{Extr}_{\Gamma} \left\{ \frac{x\Gamma}{2} - \frac{1}{2} \log \Gamma \right\} = \frac{1}{2} + \frac{1}{2} \log x$ が求まる. これより $G(x)$ は

$$G(x) = \text{Extr}_{\Gamma} \left\{ \frac{x\Gamma - 1}{2} - \frac{1}{2} \int d\lambda \rho(\lambda) \log(x\Gamma - x\lambda) \right\} = \text{Extr}_{\Lambda} \left\{ \frac{\Lambda - 1}{2} - \frac{1}{2} \int d\lambda \rho(\lambda) \log(\Lambda - x\lambda) \right\} \quad (2.78)$$

となる. ただし $x\Gamma = \Lambda$ を用いた. これより極値条件,

$$\int \frac{d\lambda \rho(\lambda)}{\Lambda - x\lambda} = 1, \quad (2.79)$$

もしくは次に定義される反復法によって Λ を決めることができる.

$$\Lambda = 1 + x \int \frac{d\lambda \rho(\lambda) \lambda}{\Lambda - x\lambda} \quad (2.80)$$

ここで注意を3点行う. はじめにいくつかの特殊な場合を除いて, 一般的に固有値分布 $\rho(\lambda)$ に対して, 式 (2.79) を用いて解析的に Λ を求めることは容易ではない. そこで式 (2.80) を用いて反復法で収束した結果として Λ を決定する方法が考えられるが, ここでも更新則によってアトラクタや安定・不安定多様体が変わることを指摘しておく. 次に $x=0$ のとき式 (2.79) や式 (2.80) から任意の固有値分布 $\rho(\lambda)$ に対して $\Lambda=1$ となり, $G(0)=0$ が得られる.

なぜなら $G(x)$ の定義から $x = 0$ になる $\vec{w}, \vec{r}$ は原点しかないため、任意の固有値分布に対して $e^{\frac{1}{2}\vec{w}^T X \vec{w}} = 1$ となり $G(0) = 0$ は明らかである。最後に一般的に次の関係が成り立つ。

$$\frac{\partial G(x)}{\partial x} = \frac{1}{2} \int d\lambda \frac{\rho(\lambda)\lambda}{\Lambda - x\lambda} = \frac{\Lambda - 1}{2x} \quad (2.81)$$

ここでの Λ は式 (2.80) を満たすとして、最後の等式は式 (2.80) を用いれば明らかである [195]。

独立同分布の場合 $X = \left\{ \frac{x_{ij}}{\sqrt{N}} \right\} \in \mathcal{M}_{N \times N}$ の成分が各々独立に $x_{ij} \sim N(0, 1)$ に従うとして、 $x = \frac{\vec{w}^T \vec{w}}{N}$ のときの $G(x)$ をここで求める。式 (2.79) に対して $\Lambda - x\lambda = y$ とおくと、

$$\frac{1}{2\pi} \int_{-2}^2 d\lambda \frac{\sqrt{2^2 - \lambda^2}}{\Lambda - x\lambda} = \int_{\Lambda-2x}^{\Lambda+2x} dy \frac{\sqrt{2^2 x^2 - (\Lambda - y)^2}}{2\pi y} = \int_{\lambda_-}^{\lambda_+} dy \frac{\sqrt{(y - \lambda_-)(\lambda_+ - y)}}{2\pi y} = 1 \quad (2.82)$$

ただし $\lambda_{\pm} = (1 \pm \sqrt{\alpha})^2 = \Lambda \pm 2x$ とおき、 $x = \sqrt{\alpha}$ と $\Lambda = 1 + \alpha = 1 + x^2$ が得られる。式 (2.81) から

$$\frac{\partial G(x)}{\partial x} = \frac{x}{2}, \quad \Rightarrow \quad G(x) = \frac{x^2}{4} \quad (2.83)$$

と求まる。しかしながらこれは上の議論を用いなくても簡単に

$$\begin{aligned} G(x) &= \lim_{N \rightarrow \infty} \frac{1}{N} \log \sum_X P(X) e^{\frac{1}{2}\vec{w}^T X \vec{w}} = \lim_{N \rightarrow \infty} \frac{1}{N} \log \sum_X P(X) \exp \left[\frac{1}{\sqrt{N}} \sum_{i>j} x_{ij} w_i w_j \right] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \left(\frac{1}{2N} \sum_{i>j} w_i^2 w_j^2 \right) = \lim_{N \rightarrow \infty} \frac{1}{4} \left(\frac{1}{N} \sum_{i=1}^N w_i^2 \right)^2 = \frac{x^2}{4} \end{aligned} \quad (2.84)$$

となる。これより半径 2 の Wigner の半円則のときの $G(x)$ が導かれた。

2.3.3 ランダム長方形列の積分公式

p 次元ベクトルと N 次元ベクトルがそれぞれ $\vec{u} \in \mathbf{R}^p$ と $\vec{w} \in \mathbf{R}^N$ に固定されており、 $p \times N$ のランダム長方形列 $X \in \mathcal{M}_{p \times N}$ で定義される $X^T X \in \mathcal{M}_{N \times N}$ の「固有値分布 $\rho(\lambda)$ が同一のランダム長方形列のアンサンブル」で $e^{-i\vec{u}^T X \vec{w}}$ のアンサンブル平均を行うことを考える。具体的には次に定義される

$$F(x, y) = \lim_{N \rightarrow \infty} \frac{1}{N} \log \sum_X P(X) e^{-i\vec{u}^T X \vec{w}} \quad (2.85)$$

を考える。ただし $Nx = \vec{w}^T \vec{w}$, $py = \vec{u}^T \vec{u}$ とおく。このアンサンブルの特徴付けとして次の点に着目する。まず特異値分解を用いてランダム長方形列 X を次のように分解することができる（自明な置換を除いて一意的に決まる）。

$$X = U D V^T \quad (2.86)$$

ただし対角行列 $D := \text{diag}\{d_k\} \in \mathcal{M}_{p \times N}$ は特異値 d_k を持ち、非対角成分が 0 の長方形列である。このアンサンブルの特徴付けとして、「 U と V はそれぞれ $p \times p$ と $N \times N$ の Haar 測度から一様に生成されている」という点だけを採用する。さて新しい座標系を $\vec{t} = V^T \vec{w}$ と $\vec{\tau} = U^T \vec{u}$ で表すと、次の拘束条件

$$\vec{w}^T \vec{w} = \vec{t}^T \vec{t} = Nx, \quad \vec{u}^T \vec{u} = \vec{\tau}^T \vec{\tau} = py \quad (2.87)$$

を満たすことが分かる。これより式 (2.85) は新しい座標系では

$$F(x, y) = \lim_{N \rightarrow \infty} \frac{1}{N} \log \frac{\int_{-\infty}^{\infty} d\vec{\tau} d\vec{t} \delta(\vec{\tau}^T \vec{\tau} - py) \delta(\vec{t}^T \vec{t} - Nx) e^{-i\vec{\tau}^T D \vec{t}}}{\int_{-\infty}^{\infty} d\vec{\tau} d\vec{t} \delta(\vec{\tau}^T \vec{\tau} - py) \delta(\vec{t}^T \vec{t} - Nx)} = \mathcal{A}(D) - \mathcal{A}(0) \quad (2.88)$$

と表すことができる。ここで

$$\begin{aligned}
 \mathcal{A}(D) &:= \text{Extr}_{\Gamma_x, \Gamma_y} \lim_{N \rightarrow \infty} \frac{1}{N} \log \int_{-\infty}^{\infty} \frac{d\vec{\tau} d\vec{t}}{(2\pi)^{\frac{N+p}{2}}} e^{-\frac{\Gamma_x}{2}(\vec{t}^T \vec{t} - Nx) - \frac{\Gamma_y}{2}(\vec{\tau}^T \vec{\tau} - py) - i\vec{\tau}^T D \vec{t}} \\
 &= \text{Extr}_{\Gamma_x, \Gamma_y} \lim_{N \rightarrow \infty} \frac{1}{N} \left(\frac{N}{2} x \Gamma_x + \frac{p}{2} y \Gamma_y - \frac{1}{2} \log \det \begin{vmatrix} \Gamma_y & iD \\ iD^T & \Gamma_x \end{vmatrix} \right) \\
 &= \text{Extr}_{\Gamma_x, \Gamma_y} \left\{ \frac{x \Gamma_x}{2} + \frac{\alpha y \Gamma_y}{2} - \frac{\alpha - 1}{2} \log \Gamma_y - \frac{1}{2} \int d\lambda \rho(\lambda) \log(\Gamma_x \Gamma_y + \lambda) \right\} \quad (2.89)
 \end{aligned}$$

となり、また式 (2.88) の被積分関数の分母は、固有値分布が $\rho(\lambda) = \delta(\lambda)$ のときの $\mathcal{A}(D = 0)$ に対応し、 $\mathcal{A}(0) = \text{Extr}_{\Gamma_x, \Gamma_y} \left\{ \frac{x \Gamma_x}{2} + \frac{\alpha y \Gamma_y}{2} - \frac{\alpha}{2} \log \Gamma_y - \frac{1}{2} \log \Gamma_x \right\} = \frac{1}{2} + \frac{\alpha}{2} + \frac{\alpha}{2} \log y + \frac{1}{2} \log x$ が得られる。ただし $\alpha := p/N \sim O(1)$ とおいた。これより

$$F(x, y) = \text{Extr}_{\Lambda_x, \Lambda_y} \left\{ \frac{\Lambda_x - 1}{2} + \alpha \frac{\Lambda_y - 1}{2} - \frac{\alpha - 1}{2} \log \Lambda_y - \frac{1}{2} \int d\lambda \rho(\lambda) \log(\Lambda_x \Lambda_y + \lambda xy) \right\} \quad (2.90)$$

が求まる。ちなみに $x \Gamma_x = \Lambda_x$ と $y \Gamma_y = \Lambda_y$ を用いた。これより極値条件は以下の条件によって求まる。

$$\frac{1}{\Lambda_y} = \int d\lambda \frac{\rho(\lambda)}{\Lambda_x \Lambda_y + \lambda xy} \quad (2.91)$$

$$\Lambda_x - 1 = \alpha (\Lambda_y - 1) \quad (2.92)$$

ここで3点注意する。はじめに式 (2.90) より $F(x, y) =: \mathcal{F}(xy)$ のように1変数関数と見なすことができる。なぜなら $x \frac{\partial F(x, y)}{\partial x} = y \frac{\partial F(x, y)}{\partial y} = xy \frac{\partial \mathcal{F}(u)}{\partial u}$ を満たすためである。次に $xy = 0$ のとき、 $\Lambda_x = \Lambda_y = 1$ と $\mathcal{F}(0) = 0$ が得られる。これは上記のランダム正方行列の理由と同じである。3番目に一般的に次の関係が成り立つ。

$$x \frac{\partial F(x, y)}{\partial x} = -\frac{1}{2} \int d\lambda \rho(\lambda) \frac{\lambda xy + \Lambda_x \Lambda_y - \Lambda_x \Lambda_y}{\Lambda_x \Lambda_y + \lambda xy} = -\frac{1 - \Lambda_x}{2} \implies \begin{cases} \frac{\partial F(x, y)}{\partial x} = \frac{\Lambda_x - 1}{2x} \\ \frac{\partial F(x, y)}{\partial y} = \alpha \frac{\Lambda_y - 1}{2y} \end{cases} \quad (2.93)$$

ここでの Λ_x, Λ_y は極値条件を満たしているとし、式 (2.91) を用いれば2番目の等号が成り立つことが分かる。また1番目の指摘と式 (2.92) より最後の関係が得られる [110]。

最後にもう1点。上記のランダム正方行列の積分公式 $G(x)$ とランダム長方形行列の積分公式 $F(x, y)$ の関係について少しだけコメントする。議論を簡単にするために $\alpha = p/N = 1$ のときのランダム正方行列とランダム長方形行列の積分公式、 $G(x)$ と $F(x, y) = \mathcal{F}(xy)$ の関係について考察する。このときランダム正方行列 X_G の固有値分布を $\rho_G(\lambda)$ とし、ランダム長方形行列 X_F の自己相関行列 $X_F^T X_F$ の固有値分布を $\rho_F(\lambda)$ で表すことにする。混乱を避けるためにランダム行列と固有値分布に G, F の添え字をつけた。ここで

$$\rho_G(\lambda) = \rho_G(-\lambda) \quad (2.94)$$

を満たす固有値分布に限定する。さらに次の仮定

$$\rho_F(\lambda) = \int d\lambda' \rho_G(\lambda') \delta(\lambda - \lambda'^2) \quad (2.95)$$

が成り立つとき、

$$\mathcal{F}(xy) = 2G(\sqrt{-xy}) \quad (2.96)$$

の関係が得られる。なぜなら $\alpha = 1$ のとき式 (2.92) から $\Lambda_x = \Lambda_y$ であることがわかり、それらを Λ とおくと、記号 Extr を省略して $\mathcal{F}(xy)$ は次のように展開できる。

$$\begin{aligned}
 \mathcal{F}(xy) &= \frac{\Lambda - 1}{2} + \frac{\Lambda - 1}{2} - \frac{1}{2} \int d\lambda \int d\lambda' \rho_G(\lambda') \delta(\lambda - \lambda'^2) \log(\Lambda^2 + xy\lambda) \\
 &= \left(\frac{\Lambda - 1}{2} - \frac{1}{2} \int d\lambda' \rho_G(\lambda') \log(\Lambda - \lambda' \sqrt{-xy}) \right) + \left(\frac{\Lambda - 1}{2} - \frac{1}{2} \int d\lambda' \rho_G(\lambda') (\Lambda + \lambda' \sqrt{-xy}) \right) \quad (2.97)
 \end{aligned}$$

これより式 (2.94) の下で式 (2.96) が得られる。

独立同分布の場合 $X = \left\{ \frac{x_{\mu k}}{\sqrt{N}} \right\} \in \mathcal{M}_{p \times N}$ の成分が各々独立に $x_{\mu k} \sim N(0, 1)$ に従うとして, $x = \frac{\vec{w}^T \vec{w}}{N}$ と $y = \frac{\vec{u}^T \vec{u}}{p}$ が与えられた下での $F(x, y)$ を求める. まず不定積分

$$\int dx \frac{\sqrt{ax^2 + bx + c}}{x(x + \beta)} = \frac{1}{\beta} \int dx \sqrt{ax^2 + bx + c} \left(\frac{1}{x} - \frac{1}{x + \beta} \right) \quad (2.98)$$

を計算する. 各項は

$$\begin{aligned} \int \frac{dx \sqrt{ax^2 + bx + c}}{x} &= \sqrt{ax^2 + bx + c} + \frac{b}{2} \int \frac{dx}{\sqrt{ax^2 + bx + c}} + c \int \frac{dx}{x \sqrt{ax^2 + bx + c}} \\ \int \frac{dx \sqrt{ax^2 + bx + c}}{x + \beta} &= \sqrt{ax^2 + bx + c} + \frac{b - 2a\beta}{2} \int \frac{dx}{\sqrt{ax^2 + bx + c}} + (c - b\beta + a\beta^2) \int \frac{dx}{(x + \beta) \sqrt{ax^2 + bx + c}} \\ \int \frac{dx}{\sqrt{ax^2 + bx + c}} &= -\frac{1}{\sqrt{|a|}} \sin^{-1} \left(\frac{2ax + b}{\sqrt{b^2 - 4ac}} \right), \quad a < 0 \\ \int \frac{dx}{x \sqrt{ax^2 + bx + c}} &= \frac{1}{\sqrt{|c|}} \sin^{-1} \left(\frac{2c + bx}{x \sqrt{b^2 - 4ac}} \right), \quad c < 0 \\ \int \frac{dx}{(x + \beta) \sqrt{ax^2 + bx + c}} &= \frac{1}{\sqrt{|c - b\beta + a\beta^2|}} \begin{cases} \log \left| \frac{x + \beta}{bx + 2c - \beta(b + 2ax) + 2\sqrt{(c - b\beta + a\beta^2)(ax^2 + bx + c)}} \right| & c - b\beta + a\beta^2 > 0 \\ \sin^{-1} \left(\frac{bx + 2c - \beta(2ax + b)}{(x + \beta) \sqrt{b^2 - 4ac}} \right) & c - b\beta + a\beta^2 < 0 \end{cases} \end{aligned}$$

を用いる. $a = -1, b = 2(1 + \alpha), c = -(1 - \alpha)^2$ のとき $\sqrt{b^2 - 4ac} = 4\sqrt{\alpha}$ と $c - b\beta + a\beta^2 = 4\alpha - (\beta + 1 + \alpha)^2 < 0$ が得られる. ここで $\beta = \frac{\Lambda_x \Lambda_y}{xy}$ とおき, 式 (2.91) から⁷

$$\frac{1}{\Lambda_y} = -\frac{1}{2xy} + \frac{1 - \alpha}{2\Lambda_x \Lambda_y} - \frac{\sqrt{\left(\frac{\Lambda_x \Lambda_y}{xy} + 1 + \alpha \right)^2 - 4\alpha}}{2\Lambda_x \Lambda_y} \quad (2.99)$$

となる. そこで $\Lambda_x = 1 - \delta \alpha xy + \varepsilon, \Lambda_y = 1 - \delta xy + \frac{\varepsilon}{\alpha}$ と仮定すると $\delta = 1, \varepsilon = 0$ が得られ, $\Lambda_x = 1 - \alpha xy$ と $\Lambda_y = 1 - xy$ となる. また式 (2.93) から

$$\frac{\partial F(x, y)}{\partial x} = -\frac{\alpha y}{2}, \quad \frac{\partial F(x, y)}{\partial y} = -\frac{\alpha x}{2}, \quad \Rightarrow \quad F(x, y) = -\frac{\alpha xy}{2} \quad (2.100)$$

となる. しかしながら上の議論を用いる必要はなく, 簡単に

$$\begin{aligned} F(x, y) &= \lim_{N \rightarrow \infty} \frac{1}{N} \log \sum_X P(X) e^{-i \vec{u}^T X \vec{w}} = \lim_{N \rightarrow \infty} \frac{1}{N} \log \sum_X P(X) \exp \left[-\frac{i}{\sqrt{N}} \sum_{\mu=1}^p \sum_{k=1}^N x_{\mu k} u_{\mu} w_k \right] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \left(-\frac{1}{2N} \sum_{\mu=1}^p \sum_{k=1}^N u_{\mu}^2 w_k^2 \right) = -\frac{1}{2} \times \lim_{N \rightarrow \infty} \frac{p}{N} \left(\frac{1}{p} \sum_{\mu=1}^p u_{\mu}^2 \right) \left(\frac{1}{N} \sum_{k=1}^N w_k^2 \right) = -\frac{\alpha xy}{2} \quad (2.101) \end{aligned}$$

⁷式 (2.99) の計算は

$$\begin{aligned} \frac{1}{\Lambda_y} &= \frac{1}{2\pi xy} \int_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} d\lambda \frac{\sqrt{(\lambda - (1 - \sqrt{\alpha})^2)((1 + \sqrt{\alpha})^2 - \lambda)}}{\lambda(\lambda + \beta)} + \frac{[1 - \alpha]^+}{\Lambda_x \Lambda_y} \\ &= \frac{1}{2\pi xy \beta} \left\{ a\beta \int_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} \frac{dx}{\sqrt{(\lambda - (1 - \sqrt{\alpha})^2)((1 + \sqrt{\alpha})^2 - \lambda)}} + c \int_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} \frac{dx}{x \sqrt{(\lambda - (1 - \sqrt{\alpha})^2)((1 + \sqrt{\alpha})^2 - \lambda)}} \right. \\ &\quad \left. - (c - b\beta + a\beta^2) \int_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} \frac{dx}{(x + \beta) \sqrt{(\lambda - (1 - \sqrt{\alpha})^2)((1 + \sqrt{\alpha})^2 - \lambda)}} \right\} + \frac{[1 - \alpha]^+}{\Lambda_x \Lambda_y} \\ &= \frac{1}{2\pi xy \beta} \left\{ \beta \left[\sin^{-1} \left(\frac{2(1 + \alpha) - 2\lambda}{4\sqrt{\alpha}} \right) \right]_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} - |1 - \alpha| \left[\sin^{-1} \left(\frac{2(1 + \alpha)\lambda - 2(1 - \alpha)^2}{4\sqrt{\alpha}\lambda} \right) \right]_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} \right. \\ &\quad \left. + \sqrt{(\beta + \alpha + 1)^2 - 4\alpha} \left[\sin^{-1} \left(\frac{2(1 + \alpha)\lambda - 2(1 - \alpha)^2 - \beta(2(1 + \alpha) - 2\lambda)}{4\sqrt{\alpha}(\lambda + \beta)} \right) \right]_{(1-\sqrt{\alpha})^2}^{(1+\sqrt{\alpha})^2} \right\} + \frac{[1 - \alpha]^+}{\Lambda_x \Lambda_y} \\ &= -\frac{1}{2xy} - \frac{|1 - \alpha|}{2\Lambda_x \Lambda_y} - \frac{\sqrt{(\beta + 1 + \alpha)^2 - 4\alpha}}{2\Lambda_x \Lambda_y} + \frac{[1 - \alpha]^+}{\Lambda_x \Lambda_y} = -\frac{1}{2xy} + \frac{1 - \alpha}{2\Lambda_x \Lambda_y} - \frac{\sqrt{\left(\frac{\Lambda_x \Lambda_y}{xy} + 1 + \alpha \right)^2 - 4\alpha}}{2\Lambda_x \Lambda_y} \end{aligned}$$

となる.

と確認することができる。これより Marčenko-Pastur 則の時の $F(x, y)$ が導かれた。

2.3.4 まとめ

本節ではランダム行列について、特に本学位論文と関連する部分について詳説した。ここで議論したのは以下の通りである。

1. ランダム行列の固有値分布について議論した。ランダム対称行列の各成分が独立同分布に従うときの固有値分布は、半径 2 の Wigner の半円則に確率 1 で一致することを示し、また同様にランダム長方形行列の場合は Marčenko-Pastur 則に確率 1 で一致することを明らかにした。
2. 2 種類の「ランダム行列の固有値分布」で定義される関数 $G(x), F(x, y)$ を提案した。また変数間の関係についても言及した。この積分公式は本学位論文の主要な部分に関連する。
3. 独立同分布の場合において実際に $G(x)$ と $F(x, y)$ を求めた。

2.4 まとめ

本章では本研究に関連のある予備知識である、パーセプトロン学習、スピングラス理論、ランダム行列について説明した。

1. パーセプトロン学習を Bayes 推定の枠組みを用いて定式化し、その事後確率を Boltzmann 分布であるとみなし、統計力学的枠組みで解析できることを説明した。
2. スピングラス理論で開発された 2 つの解析方法であるレプリカ法と TAP 法を SK 模型を用いて説明した。
3. ランダム行列のアンサンブルにより定義される積分公式を導出した。

さらにここで紹介した手法のいくつかは以下で我々の提案する手法の拡張の手助けになることを付け加えておく。

第3章 先行研究

この章ではパーセプトロン学習の先行研究で得られた結果の中で、特に我々の研究と関係する、入力が各々独立な例題に対する「記憶容量問題」と「教師推定問題」について解説する。上述したようにパーセプトロンを含め、学習器の典型的な性能を評価するには、与える参照例題で平均化（配位平均）を行う必要がある。しかしながらこの操作はシステムサイズ N, p が大きくなるにつれて、計算量的に実行困難となるため厳密な解析は難しいことが知られており、我々はシステムサイズが無限大の極限で成り立つ枠組み（統計力学的な手法）を用いて学習器の性能評価をする方法（レプリカ法）を本章で説明する [138, 158]。さらにシステムサイズが十分大きいが有限の場合に個別の参照例題に対する各結合加重のモーメントを算出する方法（平均場近似法）を説明する [208]。この章の構成は以下の通りである。はじめにパーセプトロン学習を統計力学的方法で定式化する。次にレプリカ法を用いた解析方法を紹介し、先行研究で得られた結果のいくつかを説明する。第 3.3 節では実際に与えられた参照例題 ξ^p に基づいた事後確率 $P(\vec{w}|\xi^p)$ を用いた各結合加重 w_k の推定量 m_{wk} を効率的に求めることのできる平均場近似の方法を導出する。最後にまとめを述べる。

3.1 パーセプトロン学習の統計力学的アプローチ

p 個の入出力関係を $\xi^p = \{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_p, y_p)\}$ で表し、式 (1.21) で定義されたパーセプトロンを用いて学習する問題を統計力学的手法を用いて議論する。2.1.5 小節と 2.1.6 小節の議論から統計力学の枠組みに習い、コスト関数もしくはハミルトニアンを

$$\mathcal{H}(\vec{w}|\xi^p) : = \sum_{\mu=1}^p \Theta \left(-\frac{y_\mu}{\sqrt{N}} \vec{x}_\mu^T \vec{w} \right) \quad (3.1)$$

と定義する。これより Boltzmann 分布は

$$P(\vec{w}|\xi^p) = \frac{P(\vec{w}) e^{-\beta \mathcal{H}(\vec{w}|\xi^p)}}{Z(\xi^p)} \quad (3.2)$$

となる。ただし分配関数

$$\begin{aligned} Z(\xi^p) : &= \sum_{\vec{w}} P(\vec{w}) e^{-\beta \mathcal{H}(\vec{w}|\xi^p)} = \sum_{\vec{w}} P(\vec{w}) \prod_{\mu=1}^p \left(e^{-\beta} + (1 - e^{-\beta}) \Theta \left(\frac{y_\mu}{\sqrt{N}} \vec{x}_\mu^T \vec{w} \right) \right) \\ &= \sum_{\vec{w}} P(\vec{w}) \prod_{\mu=1}^p \left\{ \int_{-\infty}^{\infty} d\Delta_\mu g(y_\mu, \Delta_\mu) \delta \left(\Delta_\mu - \frac{\vec{x}_\mu^T \vec{w}}{\sqrt{N}} \right) \right\} \end{aligned} \quad (3.3)$$

を用いた。また便宜上 Boltzmann 因子をまとめて

$$g(y_\mu, \Delta_\mu) : = e^{-\beta \Theta(-y_\mu \Delta_\mu)} = e^{-\beta} + (1 - e^{-\beta}) \Theta(y_\mu \Delta_\mu) \quad (3.4)$$

で表した。また参照例題 ξ^p が与えられた下での条件付きエントロピー $S(\xi^p)$ は

$$-\log P(\vec{w}|\xi^p) = -\sum_{\mu=1}^p \log g \left(y_\mu, \frac{\vec{x}_\mu^T \vec{w}}{\sqrt{N}} \right) - \log P(\vec{w}) + \log Z(\xi^p) = \beta \sum_{\mu=1}^p \Theta \left(-\frac{y_\mu \vec{x}_\mu^T \vec{w}}{\sqrt{N}} \right) - \log P(\vec{w}) + \log Z(\xi^p) \quad (3.5)$$

を用いて

$$\begin{aligned} S(\xi^p) : &= -\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\vec{w}} P(\vec{w}|\xi^p) \log P(\vec{w}|\xi^p) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \left\{ \beta \sum_{\vec{w}} P(\vec{w}|\xi^p) \mathcal{H}(\vec{w}|\xi^p) - \sum_{\vec{w}} P(\vec{w}|\xi^p) \log P(\vec{w}) + \log Z(\xi^p) \right\} \end{aligned} \quad (3.6)$$

と表すことができる。前章で議論したように、特にここでは参照する出力が外的な要因でノイズ反転していない状況を考えているので、最終的には $\beta \rightarrow \infty$ をとる必要がある。これより絶対零度のとき、Boltzmann 分布の定義から $\sum_{\vec{w}} P(\vec{w}|\xi^p) H(\vec{w}|\xi^p) = 0$ となることが分かる。

ここでエントロピーを導入した理由を述べる。直感的な説明になるが記憶容量問題において、結合加重 $\vec{w}$ のすべての可能性のどれをとっても与えられた参照例題 ξ^p を記憶することができない場合、式 (3.3) の定義から $Z(\xi^p) \geq e^{-p\beta}$ となり、また $\sum_{\mu=1}^p \Theta\left(-\frac{y_\mu \vec{x}_\mu^T \vec{w}}{\sqrt{N}}\right) \leq p$ から、絶対零度でエントロピーは負になる。言い換えれば、エントロピーが 0 になるパターン比 α_C を境に、ちょうど記憶可能・不可能の切り替わりが起こることがわかる。これは図 2.1 で説明した記憶できる確率 P の切り替わりに対応する。一方教師推定問題において、教師ベクトルの $\vec{w}_0$ 以外のすべての可能性が消えたとき、つまり

$$P(\vec{w}|\xi^p) = \begin{cases} 1 & \vec{w} = \vec{w}_0 \\ 0 & \text{otherwise} \end{cases} \quad (3.7)$$

のとき、参照例題 ξ^p のエントロピー $S(\xi^p)$ は 0 になることが分かる。つまりどちらの問題においても典型的なエントロピーを評価し、その値が 0 になるパターン比を評価することが必要である。さらに配位平均（参照例題 ξ^p で平均化すること）を $P(\xi^p) := P(X)P(\vec{y}|X)$ を用いて

$$[\mathcal{A}(\xi^p)]_{\xi^p} := \sum_{\xi^p} P(X)P(\vec{y}|X)\mathcal{A}(\xi^p) \quad (3.8)$$

で表す。ただし入力行列の分布を $P(X)$ とし、入力行列 $X = \frac{1}{\sqrt{N}}\{\vec{x}_1, \dots, \vec{x}_p\} \in \mathcal{M}_{p \times N}$ が与えられた下での出力 $\vec{y} := \{y_1, y_2, \dots, y_p\}$ の条件付き確率 $P(\vec{y}|X)$ を扱う学習問題に合わせて、

$$P(\vec{y}|X) \propto \begin{cases} P(\vec{y}) & \text{記憶容量問題} \\ Z(\xi^p)P(\vec{y}) & \text{教師推定問題} \end{cases} \quad (3.9)$$

で表す。記憶容量問題では出力 $\vec{y}$ はまったくランダムに選ばれており、入力 X に依存しないため $P(\vec{y}) = 2^{-p}$ であり、教師推定問題では教師ネットワークの結合加重 $\vec{w}_0$ に応じて出力の確率が変わるため $Z(\xi^p)P(\vec{y})$ に比例した形で表すことができる。これより $S(\xi^p)$ を参照例題 ξ^p で平均化を行ったエントロピー S は

$$S := [S(\xi^p)]_{\xi^p} = - \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\vec{w}} [P(\vec{w}|\xi^p)]_{\xi^p} \log P(\vec{w}) + \lim_{N \rightarrow \infty} \frac{1}{N} [\log Z(\xi^p)]_{\xi^p} = S_0 + \phi \quad (3.10)$$

で与えられる。ただしキウムラント母関数は

$$\begin{aligned} \phi &= \lim_{N \rightarrow \infty} \frac{1}{N} [\log Z(\xi^p)]_{\xi^p} = \lim_{N \rightarrow \infty} \frac{1}{N} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \log [Z^n(\xi^p)]_{\xi^p} \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \begin{cases} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \log \sum_{\xi^p} P(X)P(\vec{y})Z^n(\xi^p) & \text{記憶容量問題} \\ \lim_{n \rightarrow 1} \frac{\partial}{\partial n} \log \sum_{\xi^p} P(X)P(\vec{y})Z^n(\xi^p) & \text{教師推定問題} \end{cases} \end{aligned} \quad (3.11)$$

であり、 $n \in \mathbb{N}$ のとき

$$Z^n(\xi^p) = \sum_{\vec{w}_1, \dots, \vec{w}_n} \prod_{a=1}^n P(\vec{w}_a) \int_{-\infty}^{\infty} \prod_{a=1}^n \prod_{\mu=1}^p \frac{d\Delta_{\mu a} du_{\mu a}}{(2\pi)^{pn}} g(y_\mu, \Delta_{\mu a}) \exp \left[i \sum_{a=1}^n \sum_{\mu=1}^p u_{\mu a} \left(\Delta_{\mu a} - \frac{\vec{x}_\mu^T \vec{w}_a}{\sqrt{N}} \right) \right] \quad (3.12)$$

と表せる。また

$$S_0 := - \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\vec{w}} P(\vec{w}) \log P(\vec{w}) \quad (3.13)$$

を用いた。さらに別の表現として式 (3.10) は $P(\vec{w}|\xi^p)P(\xi^p) = P(\vec{w}, \xi^p)$ を用いて

$$S = - \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\vec{w}, \xi^p} P(\vec{w}, \xi^p) \log \left(\frac{P(\vec{w}, \xi^p)P(\vec{w})}{P(\xi^p)P(\vec{w})} \right) = S_0 - \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\vec{w}, \xi^p} P(\vec{w}, \xi^p) \log \frac{P(\vec{w}, \xi^p)}{P(\xi^p)P(\vec{w})} \quad (3.14)$$

となるので、式 (3.10) から

$$\phi = - \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\vec{w}, \xi^p} P(\vec{w}, \xi^p) \log \frac{P(\vec{w}, \xi^p)}{P(\xi^p)P(\vec{w})} \leq 0 \quad (3.15)$$

とおくことで、 ϕ と KL 情報量を関係付けられることが分かる [185, 186].

3.2 レプリカ解析と既存結果

代表的なパーセプトロン学習である「記憶容量問題」と「教師推定問題」の解析のためには、エントロピーを評価する必要があり、エントロピーの評価 (式 (3.10)) には式 (3.11) の評価が必要になるため、これより先、分配関数の配位平均を議論する。

3.2.1 レプリカ解析

具体的に分配関数の n 次モーメントは

$$\begin{aligned} [Z^n(\xi^p)]_{\xi^p} &= \sum_{\vec{y}} P(\vec{y}) \sum_{\vec{w}_1, \dots, \vec{w}_n} \prod_{a=1}^n P(\vec{w}_a) \int_{-\infty}^{\infty} \prod_{a=1}^n \prod_{\mu=1}^p \frac{d\Delta_{\mu a} du_{\mu a}}{(2\pi)^{pn}} g(y_\mu, \Delta_{\mu a}) \\ &\quad \times \sum_X P(X) \exp \left[i \sum_{a=1}^n \sum_{\mu=1}^p u_{\mu a} \left(\Delta_{\mu a} - \frac{\vec{x}_\mu^T \vec{w}_a}{\sqrt{N}} \right) \right] \end{aligned} \quad (3.16)$$

と表される。ここで独立パターンとして、各入力成分 $x_{\mu k}$ が各々独立で、平均 0、分散 1 の正規分布に従うとして、式 (3.16) の入力 $x_{\mu k}$ に対する平均は式 (2.101) を参考にして¹

$$\begin{aligned} \sum_X P(X) \exp \left[-\frac{i}{\sqrt{N}} \sum_{\mu=1}^p \sum_{k=1}^N x_{\mu k} \left(\sum_{a=1}^n u_{\mu a} w_{ka} \right) \right] &= \exp \left[-\frac{1}{2N} \sum_{\mu=1}^p \sum_{k=1}^N \left(\sum_{a=1}^n u_{\mu a} w_{ka} \right)^2 \right] \\ &= \exp \left[-\frac{1}{2} \sum_{\mu=1}^p \vec{u}_\mu^T Q \vec{u}_\mu \right] \end{aligned} \quad (3.18)$$

と計算できる。ただし 2 種類の n 次正方行列 $Q := \{q_{ab}\}$, $\tilde{Q} := \{\tilde{q}_{ab}\} \in \mathcal{M}_{n \times n}$ と p 個の n 次ベクトル $\vec{u}_\mu := \{u_{\mu 1}, u_{\mu 2}, \dots, u_{\mu n}\}^T \in \mathbf{R}^n, \mu (= 1, 2, \dots, p)$ を用いた。また秩序変数を

$$q_{ab} = \frac{1}{N} \sum_{k=1}^N w_{ka} w_{kb} \quad (3.19)$$

とおいた。これより $\psi(n) := \lim_{N \rightarrow \infty} \frac{1}{N} \log [Z^n(\xi^p)]_{\xi^p}$ とおき、次の結果

$$\psi(n) = \text{Extr}_{Q, \tilde{Q}} \left\{ \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{u} d\vec{\Delta}}{(2\pi)^n} \prod_{a=1}^n g(y, \Delta_a) e^{-\frac{1}{2} \vec{u}^T Q \vec{u} + i \vec{u}^T \vec{\Delta}} + \log \sum_{\vec{w}} P(\vec{w}) e^{\frac{1}{2} \text{Tr} \tilde{Q} (\vec{w} \vec{w}^T - Q)} \right\} \quad (3.20)$$

が解析的に得られる。ここで 1 点コメントする。実は上記の計算で Boltzmann 因子を $g(y, \Delta)$ で表すことで、パーセプトロン学習の評価に限定したものではなく、より広いクラスについても同様の解析を行うことができることを示唆している。つまり学習問題などの個別のシナリオや問題構造は、統計学で広く知られる尤度関数 $g(y, \Delta)$ に組み込まれており、この節で説明する解析方法の多くは、Bayes 推論の枠組みを必要とする一般の問題にも援用することが可能である [110].

¹式 (3.18) では

$$\int_{-\infty}^{\infty} Dx e^{ax} = e^{\frac{a^2}{2}} \quad (3.17)$$

を用いた。

3.2.2 レプリカ対称解

上記の表現において、いくつかの特別な場合を除いて式 (3.20) を解析的に評価することが困難である。しかしながら解表現を何らかの形に限定することでその困難を回避できることが知られている。そこで我々は次の条件を満たす RS 解を仮定する。

$$q_{ab} = \begin{cases} 1 & a = b \\ q & \text{otherwise} \end{cases} \quad \tilde{q}_{ab} = \begin{cases} \tilde{\chi} & a = b \\ \tilde{q} & \text{otherwise} \end{cases} \quad (3.21)$$

これより式 (3.20) は

$$\begin{aligned} \psi(n) = & \text{Exrt}_{q, \tilde{\chi}, \tilde{q}} \left\{ \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, \sqrt{1-q}x + \sqrt{q}z) \right)^n - \frac{n\tilde{\chi}}{2} - \frac{n(n-1)q\tilde{q}}{2} \right. \\ & \left. + \log \int_{-\infty}^{\infty} Dy \left(\sum_w P(w) e^{\frac{\tilde{\chi}-\tilde{q}}{2}w^2 + yw\sqrt{\tilde{q}}} \right)^n \right\} \end{aligned} \quad (3.22)$$

となる。式 (3.22) に対して、自然数 n を解析接続できると仮定し、エントロピーは

$$S = S_0 + \begin{cases} \lim_{n \rightarrow 0} \frac{\partial \psi(n)}{\partial n} & \text{記憶容量問題} \\ \lim_{n \rightarrow 1} \frac{\partial \psi(n)}{\partial n} & \text{教師推定問題} \end{cases} \quad (3.23)$$

と表せる。以下では式 (3.22) を結合加重が離散値の場合と連続値の場合の各々に対して、式変形を行う。その解析結果を受けて記憶容量問題と教師推定問題に対する既存結果を紹介する。

3.2.3 結合加重が離散値の場合

結合加重が離散値 ($w_k = \pm 1$) で等確率の場合を考える。ここで式 (3.22) は

$$\begin{aligned} \psi(n) = & \text{Extr}_{q, \tilde{\chi}, \tilde{q}} \left\{ \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n - \frac{n\tilde{\chi}}{2} - \frac{n(n-1)}{2}q\tilde{q} \right. \\ & \left. + \log \int_{-\infty}^{\infty} Dy \left(\sum_{w=\pm 1} \frac{1}{2} e^{\frac{\tilde{\chi}-\tilde{q}}{2}w^2 + yw\sqrt{\tilde{q}}} \right)^n \right\} \\ = & \text{Extr}_{q, \tilde{q}} \left\{ \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n - \frac{n\tilde{q}}{2} - \frac{n(n-1)}{2}q\tilde{q} \right. \\ & \left. + \log \int_{-\infty}^{\infty} Dy \cosh^n(y\sqrt{\tilde{q}}) \right\} \end{aligned} \quad (3.24)$$

となる。ただし $\tilde{\chi}$ はキャンセルできるため、Extr から外した。ここで極値条件は容易に

$$\tilde{q} = \alpha \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n \left(\frac{\int_{-\infty}^{\infty} Dxg'(y, x\sqrt{1-q} + z\sqrt{q})}{\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q})} \right)^2}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n} \quad (3.25)$$

$$q = \frac{\int_{-\infty}^{\infty} Dz \cosh^n(z\sqrt{\tilde{q}}) \tanh^2(z\sqrt{\tilde{q}})}{\int_{-\infty}^{\infty} Dz \cosh^n(z\sqrt{\tilde{q}})} \quad (3.26)$$

と導出できる。ただし $g'(y, \omega) := \frac{\partial g(y, \omega)}{\partial \omega}$ を用いた。また式 (3.22) から

$$\begin{aligned} \frac{\partial \psi(n)}{\partial n} = \text{Extr}_{q, \tilde{q}} \left\{ \alpha \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n \log \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n} \right. \\ \left. - \frac{\tilde{q}}{2} - \frac{2n-1}{2} q\tilde{q} + \frac{\int_{-\infty}^{\infty} Dy \cosh^n(y\sqrt{\tilde{q}}) \log \cosh(y\sqrt{\tilde{q}})}{\int_{-\infty}^{\infty} Dy \cosh^n(y\sqrt{\tilde{q}})} \right\} \end{aligned} \quad (3.27)$$

が得られる。

パーセプトロン学習の場合 式 (3.4) で定義された $g(y, \omega) = e^{-\beta} + (1 - e^{-\beta})\Theta(y\omega)$ を用いて議論を進める。これより

$$\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) = \tilde{H}\left(z\sqrt{\frac{q}{1-q}}\right) \quad (3.28)$$

$$\int_{-\infty}^{\infty} Dxg'(y, x\sqrt{1-q} + z\sqrt{q}) = -\frac{(1 - e^{-\beta})}{\sqrt{2\pi(1-q)}} e^{-\frac{qz^2}{2(1-q)}} \quad (3.29)$$

が得られる。ただし $\tilde{H}(u) := e^{-\beta} + (1 - e^{-\beta})H(u)$ と $H(u) := \int_{-\infty}^u Dx$ を用いた。また $\int_{-\infty}^{\infty} Dxf(x) = \int_{-\infty}^{\infty} Dxf(-x)$ の関係を用いて y の平均を省略することができる。これより式 (3.25) は

$$\tilde{q} = \frac{\alpha(1 - e^{-\beta})^2}{2\pi(1-q)} \frac{\int_{-\infty}^{\infty} Dz \tilde{H}^n\left(z\sqrt{\frac{q}{1-q}}\right) \left(\frac{e^{-\frac{qz^2}{2(1-q)}}}{\tilde{H}\left(z\sqrt{\frac{q}{1-q}}\right)} \right)^2}{\int_{-\infty}^{\infty} Dz \tilde{H}^n\left(z\sqrt{\frac{q}{1-q}}\right)} \quad (3.30)$$

となる。ちなみに式 (3.26) は $g(y, w)$ を含んでいないため、変わらないことに注意。

3.2.4 結合加重が連続値の場合

結合加重が $\sum_{k=1}^N w_k^2 = N$ の拘束条件を満たす場合を考える。ここで

$$\begin{aligned} \psi(n) &= \text{Extr}_{q, \tilde{q}, \tilde{\chi}, \tilde{\Lambda}} \left\{ \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n - \frac{n\tilde{\chi}}{2} - \frac{n(n-1)}{2} q\tilde{q} \right. \\ &\quad \left. + \log \int_{-\infty}^{\infty} Dy \left(\int_{-\infty}^{\infty} \frac{dw}{\sqrt{2\pi}} e^{-\frac{\Lambda}{2}(w^2-1) + \frac{\tilde{\chi}-\tilde{q}}{2}w^2 + yw\sqrt{\tilde{q}}} \right)^n - \frac{n}{2} \right\} \\ &= \text{Extr}_{q, \tilde{q}, \tilde{\chi}, \tilde{\Lambda}} \left\{ \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n - \frac{n\tilde{\chi}}{2} - \frac{n(n-1)}{2} q\tilde{q} \right. \\ &\quad \left. + \frac{n\Lambda}{2} - \frac{n-1}{2} \log(\Lambda - \tilde{\chi} + \tilde{q}) - \frac{1}{2} \log(\Lambda - \tilde{\chi} + \tilde{q} - n\tilde{q}) - \frac{n}{2} \right\} \end{aligned} \quad (3.31)$$

となる。最後の項 $-\frac{n}{2}$ は $\tilde{\chi}, \tilde{q} = 0$ のときに定義から w の平均の対数部分が 0 になるために付け加えた。また $\Lambda - \tilde{\chi} + \tilde{q} = A$ と $\Lambda - \tilde{\chi} + \tilde{q} - n\tilde{q} = B$ とおくと、 $\tilde{\chi}, \tilde{q}$ に関係している部分は

$$\begin{aligned} &\text{Extr}_{\tilde{\chi}, \tilde{q}, \tilde{\Lambda}} \left\{ -\frac{n\tilde{\chi}}{2} - \frac{n(n-1)}{2} q\tilde{q} + \frac{n\Lambda}{2} - \frac{n-1}{2} \log(\Lambda - \tilde{\chi} + \tilde{q}) - \frac{1}{2} \log(\Lambda - \tilde{\chi} + \tilde{q} - n\tilde{q}) - \frac{n}{2} \right\} \\ &= \text{Extr}_{A, B} \left\{ \frac{n-1}{2} (A(1-q) - 1) - \frac{n-1}{2} \log A + \frac{1}{2} (B(1-q + nq) - 1) - \frac{1}{2} \log B \right\} \\ &= \frac{n-1}{2} \log(1-q) + \frac{1}{2} \log(1-q + nq) \end{aligned} \quad (3.32)$$

となる。これより

$$\begin{aligned} \psi(n) = & \text{Extr}_q \left\{ \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n \right. \\ & \left. + \frac{n-1}{2} \log(1-q) + \frac{1}{2} \log(1-q+nq) \right\} \end{aligned} \quad (3.33)$$

が得られる。極値条件を満たす q は式 (3.31) から²

$$\begin{aligned} & \frac{q}{(1-q)(1-q+nq)} \\ & \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n \left(\frac{\int_{-\infty}^{\infty} Dxg'(y, x\sqrt{1-q} + z\sqrt{q})}{\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q})} \right)^2}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n} \end{aligned} \quad (3.35)$$

を満たす。ただし $g'(y, \omega) = \frac{\partial g(y, \omega)}{\partial \omega}$ を用いた。また n の微分より

$$\begin{aligned} \frac{\partial \psi(n)}{\partial n} = & \text{Extr}_q \left\{ \alpha \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n \log \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n} \right. \\ & \left. + \frac{1}{2} \log(1-q) + \frac{q}{2(1-q+nq)} \right\} \end{aligned} \quad (3.36)$$

が得られる。ここで注意を1つ。離散値の場合と違って $\tilde{q}$ があらわに出てこない。なぜなら式 (3.32) の極値評価の際に、うまく変数を取り直すことができたため省略できるからだ。もしくは式 (3.25) に対応して

$$\tilde{q} = \frac{q}{(1-q)(1-q+nq)} \quad (3.37)$$

と表すことができる。

パーセプトロン学習の場合 式 (3.4) で定義された $g(y, \omega)$ を用いて議論を進める。式 (3.32) と同様に極値条件を満たす q に関する式として、

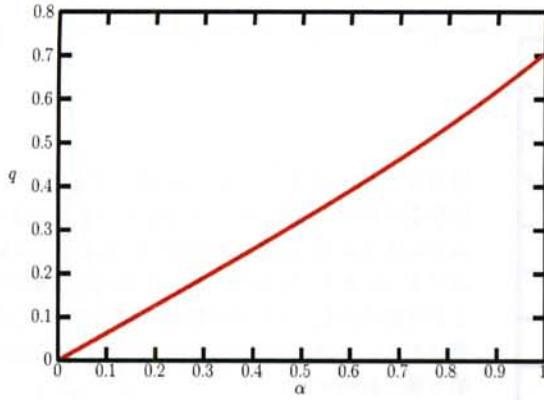
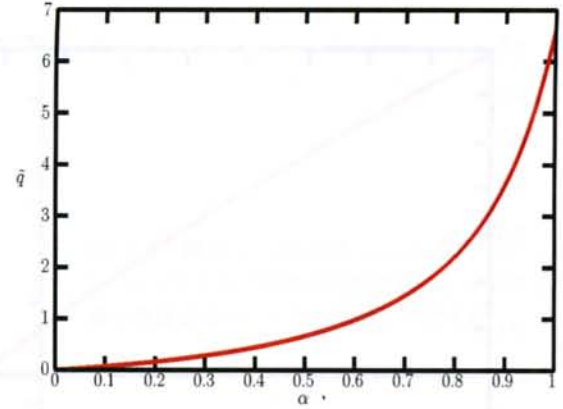
$$\frac{q}{(1-q)(1-q+nq)} = \frac{\alpha(1-e^{-\beta})^2}{2\pi(1-q)} \frac{\int_{-\infty}^{\infty} Dz \tilde{H}^n \left(z \sqrt{\frac{q}{1-q}} \right) \left(\frac{e^{-\frac{qz^2}{2(1-q)}}}{\tilde{H} \left(z \sqrt{\frac{q}{1-q}} \right)} \right)^2}{\int_{-\infty}^{\infty} Dz \tilde{H}^n \left(z \sqrt{\frac{q}{1-q}} \right)} \quad (3.38)$$

が得られる。

²式 (3.31) に式 (3.32) を代入して極値を取ると、秩序変数 q は

$$0 = \frac{\alpha n(n-1)}{2} \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n \left(\frac{\int_{-\infty}^{\infty} Dxg'(y, x\sqrt{1-q} + z\sqrt{q})}{\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q})} \right)^2}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q} + z\sqrt{q}) \right)^n} - \frac{n(n-1)q}{2(1-q)(1-q+nq)} \quad (3.34)$$

を満たさなければならない。

図 3.1: パターン比 α と秩序変数 q の図.図 3.2: パターン比 α と共役変数 $\tilde{q}$ の図.

3.2.5 記憶容量問題

記憶容量問題の場合は, $n = 0$ のときの解析を行えばよい.

離散値の場合 式 (3.27) から

$$\lim_{n \rightarrow 0} \frac{\partial \psi(n)}{\partial n} = \text{Extr}_{q, \tilde{q}} \left\{ \alpha \int_{-\infty}^{\infty} Dz \log \tilde{H} \left(z \sqrt{\frac{q}{1-q}} \right) - \frac{1-q}{2} \tilde{q} + \int_{-\infty}^{\infty} Dy \log \cosh \left(y \sqrt{\tilde{q}} \right) \right\} \quad (3.39)$$

となり, $S_0 = -\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\vec{w}} P(\vec{w}) \log P(\vec{w}) = \log 2$ を用いてエントロピーは

$$S = \log 2 + \text{Extr}_{q, \tilde{q}} \left\{ \alpha \int_{-\infty}^{\infty} Dz \log \tilde{H} \left(z \sqrt{\frac{q}{1-q}} \right) - \frac{1-q}{2} \tilde{q} + \int_{-\infty}^{\infty} Dy \log \cosh \left(y \sqrt{\tilde{q}} \right) \right\} \quad (3.40)$$

となる. ここで $\beta \rightarrow \infty$ で式 (3.26) と式 (3.30) から次の極値条件

$$\tilde{q} = \frac{\alpha}{2\pi(1-q)} \int_{-\infty}^{\infty} Dz \left(\frac{e^{-\frac{qz^2}{2(1-q)}}}{H \left(z \sqrt{\frac{q}{1-q}} \right)} \right)^2, \quad q = \int_{-\infty}^{\infty} Dz \tanh^2 \left(z \sqrt{\tilde{q}} \right) \quad (3.41)$$

が得られる. これを α を固定した下で, $(q, \tilde{q})$ の固定点を求め, 収束解を式 (3.40) に代入することでエントロピーを評価することができる. 図 3.1 と図 3.2 は式 (3.41) を用いて解析した q と $\tilde{q}$ であり, それらを用いてエントロピーを評価した結果が図 3.3 である. 上記は独立同分布に従うパターンを用いた結合加重が離散の場合の記憶容量問題の解析結果であり, W. Krauth and M. Mézard (1989) によって解析された [124].

連続値の場合 $n = 0$ のとき $\beta \rightarrow \infty$ で議論する. 式 (3.38) より

$$\frac{q}{1-q} = \frac{\alpha}{2\pi} \int_{-\infty}^{\infty} Dy \left(\frac{e^{-\frac{qy^2}{2(1-q)}}}{H \left(y \sqrt{\frac{q}{1-q}} \right)} \right)^2 \quad (3.42)$$

が得られる. 大雑把に言えば, ある α に対して式 (3.42) を満たす q は

$$\delta := \frac{1}{2\pi} \int_{-\infty}^{\infty} Dy \left(\frac{e^{-\frac{qy^2}{2(1-q)}}}{H \left(y \sqrt{\frac{q}{1-q}} \right)} \right)^2 \quad (3.43)$$

とおくと, $q = 1 - \frac{1}{1+\alpha\delta}$ となり, α の変化に比べて δ の変化は少ないと仮定すると, q は α に対する単調非減少関数とみなすことができる. また q はレプリカ間の内積として定義したため, $|q| \leq 1$ である. これより極限として $q \rightarrow 1$

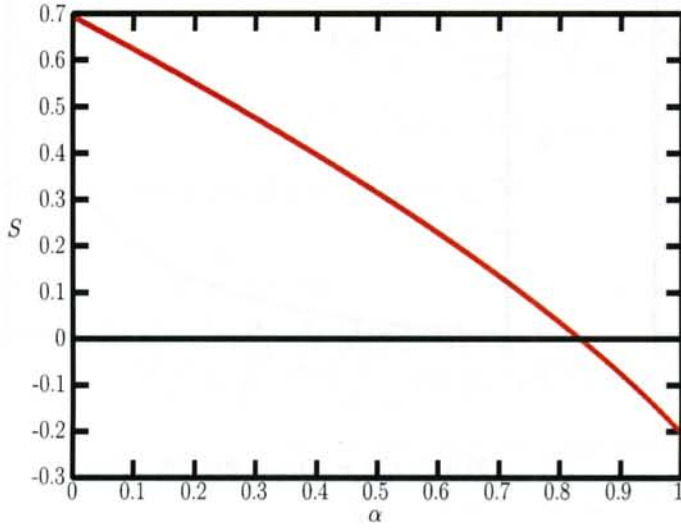


図 3.3: パターン比 α とエントロピー S の図. $S \rightarrow 0$ になるパターン比は $\alpha_C \simeq 0.833$ となる. つまりシステムサイズ N に対して典型的に $p \simeq 0.833N$ 個のパターンまでパーセプトロンを用いて記憶することが可能である. また別の数値実験を用いてこの臨界点が正しいかどうかの検証を行った. その結果は第5章で説明する.

となる α を解析することで学習器の性能を特徴づけることができる. $q \rightarrow 1$ のとき $\sqrt{\frac{q}{1-q}} \rightarrow \infty$ となるので被積分関数は $u = y\sqrt{\frac{q}{1-q}}$ とおくと, y の正負で,

$$\frac{e^{-\frac{u^2}{2}}}{H(u)} \simeq \begin{cases} \sqrt{2\pi}u & u \rightarrow \infty \\ 0 & u \rightarrow -\infty \end{cases} \quad (3.44)$$

が得られる (付録 A). これより $q \rightarrow 1$ になるパターン比 α_C は

$$\frac{q}{(1-q)} - \frac{\alpha_C}{2\pi} \left\{ \int_0^\infty Dy (\sqrt{2\pi}y)^2 \times \frac{q}{1-q} + \int_{-\infty}^0 Dy \times 0 \right\} = 0 \quad \Rightarrow \quad \alpha_C = 2 \quad (3.45)$$

となり, 記憶容量として $\alpha_C = 2$ が得られる. この結果は E. Gardner (1988) [71] によって得られた結果であり, T. Cover (1965) [39] が組み合わせ論によって求めた結果と一致する. さらに $\beta \rightarrow \infty$ で $g(y, \omega) = \Theta(y\omega - \kappa)$ とするとき, つまりマージン κ が存在するとき上と同様の議論ができ, 臨界値 $\alpha_C(\kappa)$ は

$$\alpha_C(\kappa) = \left(\int_{-\kappa}^\infty Dy (y + \kappa)^2 \right)^{-1} = \frac{1}{(1 + \kappa^2)H(-\kappa) + \frac{\kappa}{\sqrt{2\pi}}e^{-\frac{\kappa^2}{2}}} \quad (3.46)$$

となる. 図 3.4 から $\alpha_C(\kappa)$ は $\kappa \geq 0$ の単調非増加関数であることが分かる.

3.2.6 教師推定問題

教師推定問題の場合は, $n = 1$ のときの解析を行えばよい.

離散値の場合 式 (3.27) から

$$\lim_{n \rightarrow 1} \frac{\partial \psi(n)}{\partial n} = \text{Extr}_{q, \tilde{q}} \left\{ \frac{2\alpha}{1 + e^{-\beta}} \int_{-\infty}^\infty Dz \tilde{H} \left(z \sqrt{\frac{q}{1-q}} \right) \log \tilde{H} \left(z \sqrt{\frac{q}{1-q}} \right) - \frac{1+q}{2} \tilde{q} + \int_{-\infty}^\infty Dy \log \cosh \left(\tilde{q} + y \sqrt{\tilde{q}} \right) \right\}$$

となる. ただし $\int_{-\infty}^\infty Dz \tilde{H} \left(z \sqrt{\frac{q}{1-q}} \right) = \frac{1 + e^{-\beta}}{2}$ と $\int_{-\infty}^\infty Dx H(ax) = 0.5$ を用いた. これよりエントロピーは

$$S = \log 2 + \lim_{n \rightarrow 1} \frac{\partial \psi(n)}{\partial n} \quad (3.48)$$

となる. ここで $\beta \rightarrow \infty$ で式 (3.26) と式 (3.30) から次の極値条件

$$\tilde{q} = \frac{\alpha}{\pi(1-q)} \int_{-\infty}^\infty Dz \frac{e^{-\frac{qz^2}{1-q}}}{H \left(z \sqrt{\frac{q}{1-q}} \right)} = \frac{\alpha}{\pi\sqrt{1-q}} \int_{-\infty}^\infty Dz \frac{e^{-\frac{qz^2}{2}}}{H(z\sqrt{q})}, \quad q = \int_{-\infty}^\infty Dz \tanh^2 \left(\tilde{q} + z \sqrt{\tilde{q}} \right) \quad (3.49)$$

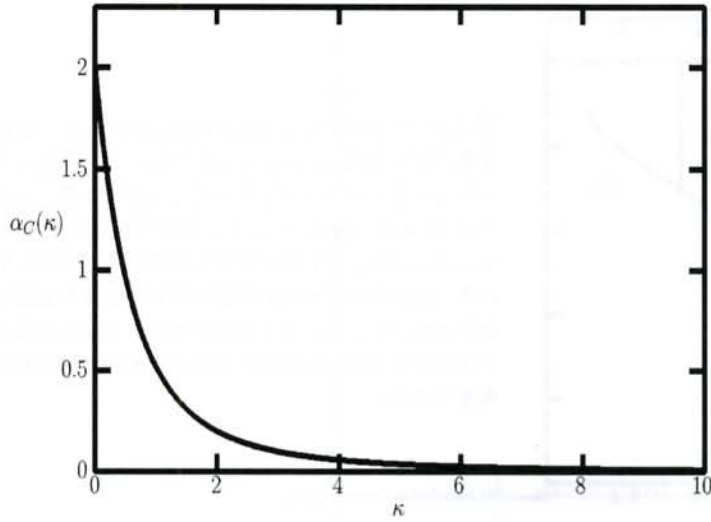


図 3.4: 閾値 κ と臨界値 $\alpha_C(\kappa)$ の図. $\kappa = 0$ で $\alpha_C = 2$ となる. 「模型の対称性から $\kappa = 0$ のときに最も性能がよい」と言う結果と一致する.

が得られる. 図 3.5 は式 (3.48) と式 (3.49) を用いて解析したエントロピーの結果であり, G. Györgyi(1990) によって得られたものである [80]. またより詳細の解析を行うと, 図 3.6 が示すとおり, α_{GD} の前後で q の不連続に変化している. これはエントロピーが $\alpha < \alpha_{GD}$ において $q \neq 1$ の解が安定であり, $\alpha > \alpha_{GD}$ で $q = 1$ の解が安定になるためである. これは相転移と呼ばれる現象であり, 今回は秩序変数 q に不連続性, 正確には非解析性が含まれるため, 1 次相転移である. さらに RS 解の $\alpha > \alpha_{GD}$ での局所安定性を求めてみると, $\alpha < \alpha_{AT} \simeq 1.493$ の範囲で式 (3.49) の解が安定であることが保証される. 詳しくは第 5 章で議論する.

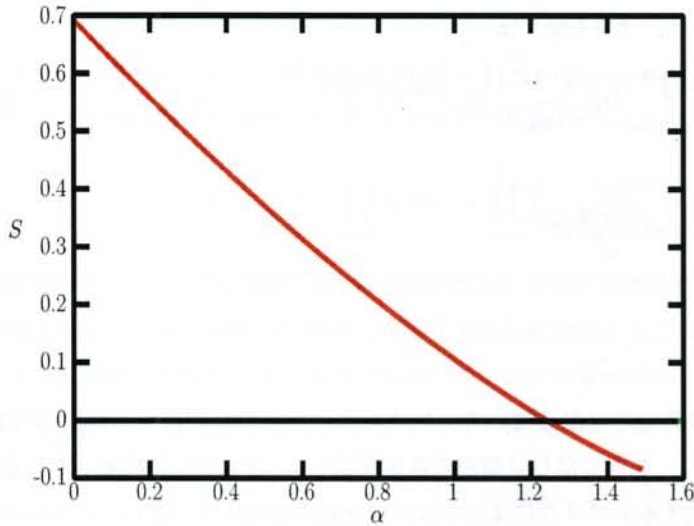


図 3.5: パターン比 α とエントロピー S の図. $S \rightarrow 0$ になるパターン比は $\alpha_{GD} \simeq 1.245$ となる. また $\alpha > \alpha_{GD}$ においては同定できたので, $S = 0$ の解が選ばれる. この結果の意味として, システムサイズ N に対して典型的に $p \simeq 1.245N$ 個のパターンで, パーセプトロンを用いれば教師ベクトルを同定することが可能であることを示唆している. また別の数値実験を用いてこの臨界点が正しいかどうかの検証を行った. その結果は第 5 章で説明する.

連続値の場合 $n = 1$ のとき式 (3.38) は

$$\frac{q}{1-q} = \frac{2\alpha(1-e^{-\beta})^2}{2\pi(1-q)(1+e^{-\beta})} \int_{-\infty}^{\infty} Dz \frac{e^{-\frac{qz^2}{1-q}}}{\tilde{H}\left(z\sqrt{\frac{q}{1-q}}\right)} = \frac{\alpha(1-e^{-\beta})^2}{\pi\sqrt{1-q}(1+e^{-\beta})} \int_{-\infty}^{\infty} Dy \frac{e^{-\frac{qy^2}{2}}}{\tilde{H}(y\sqrt{q})} \quad (3.50)$$

となる. ただし $\frac{z}{\sqrt{1-q}} =: y$ を用いた. $\beta \rightarrow \infty$ に限定し, 記憶容量問題と同様に教師推定問題の解析も行う. 記憶容量問題と決定的に違う点は教師ベクトル $\vec{w}_0$ が結合加重の空間に含まれているため (記憶容量問題の場合, 含まれて

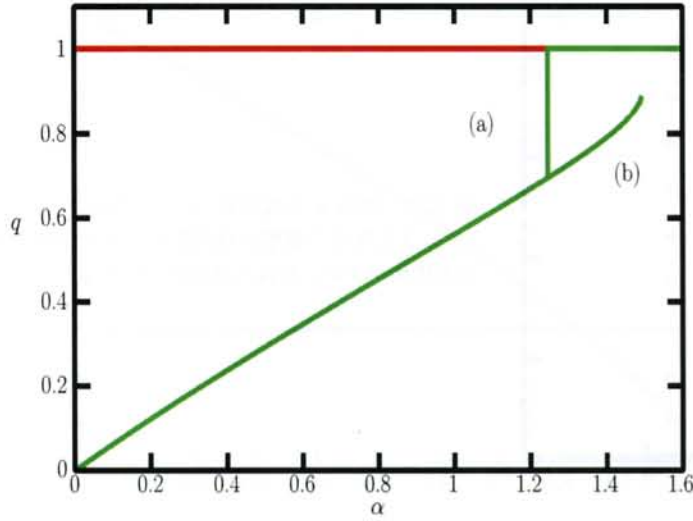


図 3.6: パターン比 α と秩序変数 q の図. $S \rightarrow 0$ になるパターン比 $\alpha_{GD} \simeq 1.245$ において, (a) $q \neq 1$ から $q = 1$ にジャンプしていることが分かる. つまり $\alpha < \alpha_{GD}$ において $q \neq 1$ の解が安定であるが, $\alpha > \alpha_{GD}$ で $q = 1$ の解が安定であることを示している. (b) さらに RS 解の解の局所安定性を議論したところ, $\alpha < \alpha_{AT} \simeq 1.493$ において安定であることが保証される. 物理的には 1 次相転移と呼ばれる現象である.

いることは保証されていない), 完全に $q = 1$ となることが予測されるが, α を大きくしても完全に一致することが難しい³. そこで α の大きいところで $q = 1$ からのズレ $\varepsilon = 1 - q$ の漸近形を議論する. 式 (3.50) において $q \rightarrow 1$ のとき $c := \int_{-\infty}^{\infty} Dy \frac{e^{-\frac{y^2}{2}}}{H(y)} \simeq 2.263$ とおき, 式 (3.50) の両辺を 2 乗し $q^2 = \frac{\alpha^2 c^2}{\pi^2} (1 - q)$ となるので, ε は

$$\varepsilon \simeq \frac{\pi^2}{c^2 \alpha^2} \simeq \frac{1.926}{\alpha^2} \quad (3.51)$$

という関数形に漸近することが分かる⁴. ただし $q^2 \simeq 1$ とした. この結果は G. Györgyi and N. Tishby(1990) により導出されたものである [81]. その結果を図 3.7 に両対数でプロットしてある.

またマージン付きの場合 ($g(y, \Delta) = \Theta(y\Delta - \kappa)$), 式 (3.50) は

$$\frac{q}{\sqrt{1-q}} = \frac{\alpha}{\pi} \int_{-\infty}^{\infty} Dy \frac{e^{-\frac{q}{2}y^2}}{H(y\sqrt{q})} \times e^{-\frac{\kappa^2}{2} + \sqrt{\frac{1-q}{q}}y\kappa} \quad (3.53)$$

となり, $\alpha \gg 1$ で

$$\varepsilon \simeq \frac{1.926}{\alpha^2} e^{\kappa^2} \quad (3.54)$$

が得られる.

3.3 平均場近似法

レプリカ法は N が大きい極限でのシステムの典型的な振る舞いを少数の変数に注目して評価する枠組みである. しかしながら現実的な状況では, N は有限であり, またしばしば個別の参照例題 ξ^p に対する性能評価や期待値評価が必要な場合も生じる. すべての可能性で平均化するレプリカ法はこのような評価には向いていない. こうした場合への対応策として, ξ^p の背後にある統計法則を考慮しながら精度の高い近似法を構成することのできる「平均場近似法 (mean field approach)」, もしくは確率推論では「確率伝搬法 (belief propagation)」が知られている. この節ではその平均場近似法を説明する.

³ 離散値と連続値の場合で, ある点の近傍に含まれる結合加重の量の振る舞いが異なるため, 離散値のように教師を完全に同定することが難しい.

⁴ ちなみに有限の β においては

$$\varepsilon \simeq \frac{\gamma}{\alpha^2}, \quad \gamma := \frac{\pi^2}{c^2(\beta)} \left(\frac{1 + e^{-\beta}}{1 - e^{-\beta}} \right)^4, \quad c(\beta) := \int_{-\infty}^{\infty} Dy \frac{(1 + e^{-\beta})e^{-\frac{y^2}{2}}}{\tilde{H}(y)} \quad (3.52)$$

という漸近形が得られる.

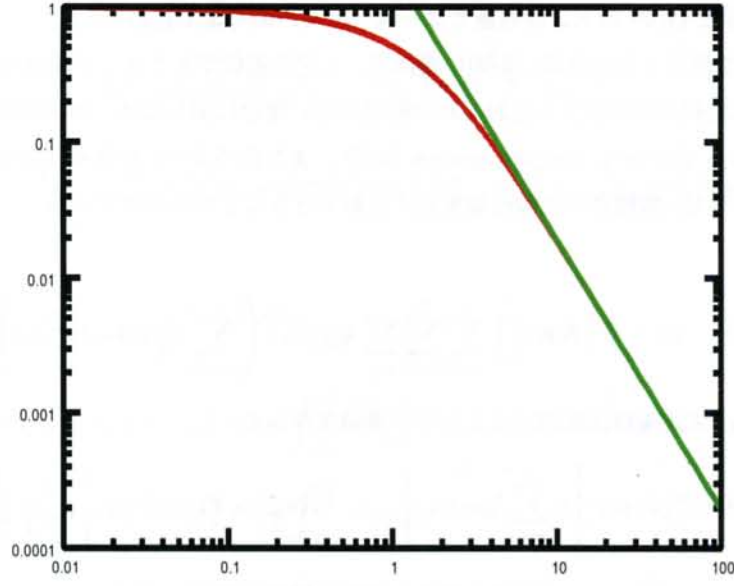


図 3.7: 式 (3.50) を用いて求めた α と ε の図. $\alpha \gg 1$ で漸近形 $\varepsilon \simeq \frac{1.926}{\alpha^2}$ が得られる.

3.3.1 Bethe 自由エネルギー

一般的に, ある確率モデルの族を仮定し, その中で事後確率 $P(\vec{w}|\xi^p)$ に KL 情報量の意味で最も近い近似分布 $Q(\vec{w})$ を求めることによって平均場近似が構成される. ここで事後確率は

$$P(\vec{w}|\xi^p) \propto P(\vec{w}) \prod_{\mu=1}^p P_{\mu}(\vec{w}) = P^{1-p}(\vec{w}) \prod_{\mu=1}^p (P_{\mu}(\vec{w})P(\vec{w})) \quad (3.55)$$

のように表される. ただし事前確率 $P(\vec{w}) = \prod_{k=1}^N P_k(w_k)$ と条件付き確率 $P_{\mu}(\vec{w}) = P(y_{\mu}|\vec{x}_{\mu}, \vec{w}) \propto g\left(y_{\mu}, \frac{\vec{x}_{\mu}^T \vec{w}}{\sqrt{N}}\right)$ を用いた. それに対して近似分布 $Q(\vec{w})$ をビリーフ $b_k(w_k), b_{\mu}(\vec{w}), (k=1, 2, \dots, N; \mu=1, 2, \dots, p)$ を用いて

$$Q(\vec{w}) \propto \prod_{k=1}^N b_k(w_k) \times \prod_{\mu=1}^p \left(\frac{b_{\mu}(\vec{w})}{\prod_{k=1}^N b_k(w_k)} \right) = \prod_{k=1}^N b_k^{1-p}(w_k) \times \prod_{\mu=1}^p b_{\mu}(\vec{w}) \quad (3.56)$$

で定義される. 一般的に確率変数間の因果関係を表現する確率モデルが⁴, 木構造のグラフィカルモデルで表されている場合, 式 (3.56) を用いて厳密な周辺化尤度を評価できることが知られている [208]. その方針を局所的には木構造と見なせるが大域的には木構造ではないグラフィカルモデルに対しても適用することで, 実効的なアルゴリズムが構成できることが平均場近似でのアイデアである. ただしビリーフは

$$b_k(w_k) = \sum_{\vec{w} \setminus w_k} b_{\mu}(\vec{w}) \quad (3.57)$$

の条件を満足することを仮定する. ただし記号 $\vec{w} \setminus w_k$ は変数 w_k を固定して, それ以外の変数で和を実行することを表す. これより KL 情報量は

$$NKL = \sum_{\mu=1}^p \sum_{\vec{w}} b_{\mu}(\vec{w}) \log \frac{b_{\mu}(\vec{w})}{P_{\mu}(\vec{w})P(\vec{w})} + (1-p) \sum_{k=1}^N \sum_{w_k} b_k(w_k) \log \frac{b_k(w_k)}{P_k(w_k)} \quad (3.58)$$

と定義される. 式 (3.58) は Bethe 自由エネルギーとも呼ばれる. この KL 情報量は各 μ ごとの事後確率 $P_{\mu}(\vec{w})P(\vec{w})$ を近似する $b_{\mu}(\vec{w})$ と各 k ごとの事前分布 $P_k(w_k)$ を近似する $b_k(w_k)$ の KL 情報量の重ね合わせとして定義される. 一

一般的にこれは $P(\vec{w}|\xi^p)$ と $Q(\vec{w})$ の KL 情報量とは一致しなく、あくまで近似である。しかしながら疎結合でループ（グラフ上を一筆書きで戻るときの経路）が短い場合は、上記の近似でもうまく事後確率で求まる統計量を算出することができる。さらに密結合の場合でもそれぞれの相関が弱い場合に関しては、事後確率で求まる統計量を効率的に求めることが可能である（loopy belief propagation など）。また各パターンが各々独立の場合（および相関が強くない場合）、式（3.58）の KL 情報量で比較的精度良く近似できることが知られている。

さて

$$L := NKL + \sum_{\mu=1}^p \sum_{k=1}^N \sum_{w_k} \lambda_{\mu k}(w_k) \left(\sum_{\vec{w} \setminus w_k} b_{\mu}(\vec{w}) - b_k(w_k) \right) \quad (3.59)$$

で定義された Lagrange 未定乗数法を用いることで、極値条件から

$$b_{\mu}(\vec{w}) \propto P_{\mu}(\vec{w})P(\vec{w}) \exp \left[- \sum_{k=1}^N \lambda_{\mu k}(w_k) \right], \quad b_k(w_k) \propto P_k(w_k) \exp \left[\frac{1}{1-p} \sum_{\mu=1}^p \lambda_{\mu k}(w_k) \right] \quad (3.60)$$

が得られる。また

$$\tilde{\lambda}_{\mu k}(w_k) := \frac{1}{1-p} \sum_{\mu=1}^p \lambda_{\mu k}(w_k) + \lambda_{\mu k}(w_k) \quad (3.61)$$

で新しい関数 $\tilde{\lambda}_{\mu k}(w_k)$ を定義すると、 $\frac{1}{1-p} \sum_{\mu=1}^p \lambda_{\mu k}(w_k) = \sum_{\mu=1}^p \tilde{\lambda}_{\mu k}(w_k)$ と $\lambda_{\mu k}(w_k) = - \sum_{\nu(\neq \mu)} \tilde{\lambda}_{\nu k}(w_k)$ が成り立ち⁵、式（3.60）は

$$b_{\mu}(\vec{w}) \propto P_{\mu}(\vec{w})P(\vec{w}) \exp \left[\sum_{k=1}^N \sum_{\nu(\neq \mu)} \tilde{\lambda}_{\nu k}(w_k) \right], \quad b_k(w_k) \propto P_k(w_k) \exp \left[\sum_{\mu=1}^p \tilde{\lambda}_{\mu k}(w_k) \right] \quad (3.62)$$

と簡単化される。式（3.62）から分かるように関数 $\tilde{\lambda}_{\nu k}(w_k)$ によって、もとの $P(\vec{w}|\xi^p)$ の他の変数からの効果を繰り込んだ形になっている。以下では結合加重が離散値の場合と連続値の場合の2種類に対して平均場近似法のアルゴリズムを導出する。

3.3.2 離散値の場合

$w_k = \pm 1$ で $P_k(w_k) = 1/2$ のときを考える。このとき $\tilde{\lambda}_{\mu k}(w_k)$ を次のように定義できる。

$$\tilde{\lambda}_{\mu k}(w_k) := w_k \tilde{h}_{\mu k} \quad (3.63)$$

定数項は式（3.62）の分母と分子でキャンセルするため省略できる。同様に $w_k^2 = 1$ より2次以上は考えなくてよい。ここでは統計量を求めるためにそれぞれの分布の特性関数を用いる方法を採用する。特性関数は次のように定義され、

$$\zeta_k(\theta_k) := \log \sum_{w_k=\pm 1} b_k(w_k) e^{\theta_k w_k} = \log \cosh(h_{wk} + \theta_k) + \text{const.} \quad h_{wk} := \sum_{\mu=1}^p \tilde{h}_{\mu k} \quad (3.64)$$

となる。ただし $\theta_k = 0$ のとき $\zeta_k(\theta_k = 0) = 0$ となるように θ_k に依存しない const. が決められる。さて w_k を平均 m_{wk} 、分散 χ_{wk} に従うと仮定すると、 h_{wk} を用いて

$$m_{wk} = \left. \frac{\partial \zeta_k(\theta_k)}{\partial \theta_k} \right|_{\theta_k \rightarrow 0} = \tanh(h_{wk}) \quad (3.65)$$

$$\chi_{wk} = \left. \frac{\partial^2 \zeta_k(\theta_k)}{\partial \theta_k^2} \right|_{\theta_k \rightarrow 0} = 1 - \tanh^2(h_{wk}) \quad (3.66)$$

⁵ただし記号 $\sum_{\nu(\neq \mu)} \alpha_{\nu} := \sum_{\nu=1}^p \alpha_{\nu} - \alpha_{\mu}$ とする。

という関係が成り立つことが分かる。次にデルタ関数のフーリエ変換を用いると、 $b_\mu(\vec{w})$ の評価は

$$\begin{aligned} b_\mu(\vec{w}) &\propto \int_{-\infty}^{\infty} d\Delta_\mu g(y_\mu, \Delta_\mu) \delta\left(\Delta_\mu - \frac{\vec{x}_\mu^T \vec{w}}{\sqrt{N}}\right) \exp\left[\sum_{k=1}^N w_k \sum_{\nu(\neq \mu)} \tilde{h}_{\nu k}\right] \\ &= \int_{-\infty}^{\infty} \frac{d\Delta_\mu du_\mu}{2\pi} g(y_\mu, \Delta_\mu) \exp\left[iu_\mu \left(\Delta_\mu - \frac{\vec{x}_\mu^T \vec{w}}{\sqrt{N}}\right) + \sum_{k=1}^N w_k (h_{wk} - \tilde{h}_{\mu k})\right] \end{aligned} \quad (3.67)$$

となり、特性関数は

$$\begin{aligned} \zeta_\mu(\vec{\theta}) &:= \log \sum_{\vec{w}} b_\mu(\vec{w}) e^{\vec{\theta}^T \vec{w}} = \log \int_{-\infty}^{\infty} \frac{d\Delta_\mu du_\mu}{2\pi} g(y_\mu, \Delta_\mu) e^{iu_\mu \Delta_\mu} \prod_{l=1}^N \cosh\left(\theta_l + h_{wl} - \tilde{h}_{\mu l} - \frac{iu_\mu x_{\mu l}}{\sqrt{N}}\right) + const. \\ &= \sum_{k=1}^N \log \cosh(\theta_k + h_{wk} - \tilde{h}_{\mu k}) + \log \int_{-\infty}^{\infty} \frac{d\Delta_\mu du_\mu}{2\pi} g(y_\mu, \Delta_\mu) \exp\left[iu_\mu (\Delta_\mu - h_{u\mu}) - \frac{u_\mu^2}{2} \tilde{\chi}_{u\mu}\right] + const. \\ &= \sum_{k=1}^N \log \cosh(\theta_k + h_{wk} - \tilde{h}_{\mu k}) + \log \int_{-\infty}^{\infty} Dx g(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu}) + const. \end{aligned} \quad (3.68)$$

となる。式 (3.68) の第 1 行目から第 2 行目は被積分関数の指数の肩の $\log \cosh\left(\theta_l + h_{wl} - \tilde{h}_{\mu l} - \frac{iu_\mu x_{\mu l}}{\sqrt{N}}\right)$ を $\theta_l + h_{wl} - \tilde{h}_{\mu l}$ の周りでテーラー展開し、 N の大きい極限で u_μ の 3 次以上は無視できることを用いた。またはすでに

$$\tilde{\chi}_{u\mu} := \frac{1}{N} \sum_{l=1}^N x_{\mu l}^2 (1 - \tanh^2(\theta_l + h_{wl} - \tilde{h}_{\mu l})), \quad h_{u\mu} := \frac{1}{\sqrt{N}} \sum_{l=1}^N x_{\mu l} \tanh(\theta_l + h_{wl} - \tilde{h}_{\mu l}) \quad (3.69)$$

と新しい変数に置き換えている。これより

$$m_{wk} = \left. \frac{\partial \zeta_\mu(\vec{\theta})}{\partial \theta_k} \right|_{\vec{\theta} \rightarrow 0} = \tanh(h_{wk} - \tilde{h}_{\mu k}) + \frac{\partial h_{u\mu}}{\partial \theta_k} m_{u\mu} + \frac{1}{2} \frac{\partial \tilde{\chi}_{u\mu}}{\partial \theta_k} \delta_{u\mu} \quad (3.70)$$

$$\begin{aligned} \chi_{wk} &= \left. \frac{\partial^2 \zeta_\mu(\vec{\theta})}{\partial \theta_k^2} \right|_{\vec{\theta} \rightarrow 0} = 1 - \tanh^2(h_{wk} - \tilde{h}_{\mu k}) + \frac{\partial^2 h_{u\mu}}{\partial \theta_k^2} m_{u\mu} + \frac{1}{2} \frac{\partial^2 \tilde{\chi}_{u\mu}}{\partial \theta_k^2} \delta_{u\mu} \\ &\quad + \left(\frac{\partial h_{u\mu}}{\partial \theta_k}\right)^2 (\delta_{u\mu} - m_{u\mu}^2) + O\left(\frac{1}{N^{\frac{3}{2}}}\right) \end{aligned} \quad (3.71)$$

が得られる。ただし

$$m_{u\mu} := \frac{\int_{-\infty}^{\infty} Dx g'(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}{\int_{-\infty}^{\infty} Dx g(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}, \quad \delta_{u\mu} := \frac{\int_{-\infty}^{\infty} Dx g''(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}{\int_{-\infty}^{\infty} Dx g(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})} \quad (3.72)$$

を用いた⁶。ここで $\tilde{m}_{\mu k} := \tanh(\tilde{h}_{\mu k}) \simeq \tilde{h}_{\mu k}$ とおき、式 (3.64) から $h_{wk} \sim O(1)$ であるためには、十分大きな N に対して $\tilde{m}_{\mu k} = \frac{A}{\sqrt{N}} + \frac{B}{N}$ であることが期待される。また式 (3.70) と式 (3.71) において⁷

$$\tanh(h_{wk} - \tilde{h}_{\mu k}) \simeq m_{wk} - (1 - m_{wk}^2) \tilde{m}_{\mu k} - m_{wk} (1 - m_{wk}^2) \tilde{m}_{\mu k}^2 \quad (3.74)$$

$$1 - \tanh^2(h_{wk} - \tilde{h}_{\mu k}) \simeq (1 - m_{wk}^2) (1 + 2m_{wk} \tilde{m}_{\mu k} - (1 - 3m_{wk}^2) \tilde{m}_{\mu k}^2) \quad (3.75)$$

⁶ 式 (3.70) と式 (3.71) において

$$\begin{aligned} \frac{\partial h_{u\mu}}{\partial \theta_k} &= \frac{x_{\mu k}}{\sqrt{N}} (1 - \tanh^2(h_{wk} - \tilde{h}_{\mu k})), \quad \frac{\partial^2 h_{u\mu}}{\partial \theta_k^2} = -\frac{2x_{\mu k}}{\sqrt{N}} \tanh(h_{wk} - \tilde{h}_{\mu k}) (1 - \tanh^2(h_{wk} - \tilde{h}_{\mu k})), \\ \frac{\partial \tilde{\chi}_{u\mu}}{\partial \theta_k} &= -2 \frac{x_{\mu k}^2}{N} \tanh(h_{wk} - \tilde{h}_{\mu k}) (1 - \tanh^2(h_{wk} - \tilde{h}_{\mu k})), \quad \frac{\partial^2 \tilde{\chi}_{u\mu}}{\partial \theta_k^2} = -\frac{2x_{\mu k}}{N} (1 - \tanh^2(h_{wk} - \tilde{h}_{\mu k})) (1 - 3 \tanh^2(h_{wk} - \tilde{h}_{\mu k})) \end{aligned}$$

が得られる。

⁷ 一般的に $\tilde{m}_{\mu k}$ の 2 次までの展開として

$$\tanh^n(h_{wk} - \tilde{h}_{\mu k}) \simeq m_{wk}^n - n m_{wk}^{n-1} (1 - m_{wk}^2) \tilde{m}_{\mu k} + \frac{\tilde{m}_{\mu k}^2}{2} ((n^2 - n) m_{wk}^{n-2} - 2n^2 m_{wk}^n + (n^2 + n) m_{wk}^{n+2}) \quad (3.73)$$

が得られる。

と展開することができ、注意深くオーダーを比較することで、 $A = x_{\mu k} m_{u\mu}$ と $B = m_{wk} x_{\mu k}^2 (m_{u\mu}^2 - \delta_{u\mu})$ が得られる。これより

$$h_{wk} = \frac{1}{\sqrt{N}} \sum_{\mu=1}^p x_{\mu k} m_{u\mu} + \tilde{\chi}_{wk} m_{wk}, \quad \tilde{\chi}_{wk} = \frac{1}{N} \sum_{\mu=1}^p x_{\mu k}^2 \chi_{u\mu}, \quad \chi_{u\mu} = m_{u\mu}^2 - \delta_{u\mu} \quad (3.76)$$

が得られる。また $\tilde{\chi}_{u\mu} = \frac{1}{N} \sum_{k=1}^N x_{\mu k}^2 (1 - m_{wk}^2) = \frac{1}{N} \sum_{k=1}^N x_{\mu k}^2 \chi_{wk}$ により

$$h_{u\mu} = \frac{1}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} (m_{wk} - (1 - m_{wk}^2) \tilde{m}_{\mu k}) = \frac{1}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} m_{wk} - \tilde{\chi}_{u\mu} m_{u\mu} \quad (3.77)$$

となり、Onsager の反跳場 $\tilde{\chi}_{wk}, \tilde{\chi}_{u\mu}$ が導出される。以上のアルゴリズムをまとめたのが表 3.1 である。

さらに式 (3.58) の NKL は

$$\begin{aligned} NKL &= \sum_{\mu=1}^p \sum_{\vec{w}} b_{\mu}(\vec{w}) \log \frac{P_{\mu}(\vec{w}) P(\vec{w}) \exp \left(\sum_{k=1}^N w_k (h_{wk} - \tilde{h}_{\mu k}) \right)}{P_{\mu}(\vec{w}) P(\vec{w}) \sum_{\vec{w}} P_{\mu}(\vec{w}) P(\vec{w}) \exp \left(\sum_{k=1}^N w_k (h_{wk} - \tilde{h}_{\mu k}) \right)} \\ &\quad + (1-p) \sum_{k=1}^N \sum_{w_k=\pm 1} b_k(w_k) \log \frac{b_k(w_k)}{P(w_k)} \\ &= \sum_{\mu=1}^p \sum_{\vec{w}} b_{\mu}(\vec{w}) \sum_{k=1}^N w_k (h_{wk} - \tilde{h}_{\mu k}) - \sum_{\mu=1}^p \log \sum_{\vec{w}} P_{\mu}(\vec{w}) P(\vec{w}) \exp \left(\sum_{k=1}^N w_k (h_{wk} - \tilde{h}_{\mu k}) \right) \\ &\quad + N(1-p) \log 2 + \log 2(p-1) \sum_{k=1}^N \mathcal{H} \left(\frac{1+m_{wk}}{2} \right) \\ &= (p-1) \sum_{k=1}^N m_{wk} h_{wk} - \sum_{\mu=1}^p \log \sum_{\vec{w}} g \left(y_{\mu}, \frac{\vec{x}_{\mu}^T \vec{w}}{\sqrt{N}} \right) \prod_{k=1}^N \left(\frac{e^{w_k (h_{wk} - \tilde{h}_{\mu k})}}{2 \cosh(h_{wk} - \tilde{h}_{\mu k})} \right) \\ &\quad + N(1-p) \log 2 - \sum_{\mu=1}^p \sum_{k=1}^N \log \cosh(h_{wk} - \tilde{h}_{\mu k}) + \log 2(p-1) \sum_{k=1}^N \mathcal{H} \left(\frac{1+m_{wk}}{2} \right) \end{aligned} \quad (3.78)$$

となる。ただし Shannon の 2 値エントロピー関数 $\mathcal{H}(u) = -u \log_2 u - (1-u) \log_2 (1-u)$ を用いた。各項は

$$\sum_{\vec{w}} g \left(y_{\mu}, \frac{\vec{x}_{\mu}^T \vec{w}}{\sqrt{N}} \right) \prod_{k=1}^N \left(\frac{e^{w_k (h_{wk} - \tilde{h}_{\mu k})}}{2 \cosh(h_{wk} - \tilde{h}_{\mu k})} \right) = \int_{-\infty}^{\infty} Dxg \left(y_{\mu}, x \sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu} \right) \quad (3.79)$$

$$\sum_{\mu=1}^p \log \cosh(h_{wk} - \tilde{h}_{\mu k}) = p \log \cosh(h_{wk}) - h_{wk} m_{wk} + \frac{1}{2} \sum_{\mu=1}^p \frac{x_{\mu k}^2 m_{u\mu}^2}{N} (1 - m_{wk}^2) \quad (3.80)$$

とまとまる。これより

$$\begin{aligned} NKL &= p \sum_{k=1}^N m_{wk} h_{wk} - \sum_{\mu=1}^p \log \int_{-\infty}^{\infty} Dxg \left(y_{\mu}, x \sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu} \right) + N(1-p) \log 2 - \frac{1}{2} \sum_{\mu=1}^p m_{u\mu}^2 \tilde{\chi}_{u\mu} \\ &\quad - p \sum_{\mu=1}^N \log \cosh(h_{wk}) + \log 2(p-1) \sum_{k=1}^N \mathcal{H} \left(\frac{1+m_{wk}}{2} \right) \\ &= N \log 2 - \sum_{\mu=1}^p \log \int_{-\infty}^{\infty} Dxg \left(y_{\mu}, x \sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu} \right) - \frac{1}{2} \sum_{\mu=1}^p m_{u\mu}^2 \tilde{\chi}_{u\mu} - \log 2 \sum_{k=1}^N \mathcal{H} \left(\frac{1+m_{wk}}{2} \right) \end{aligned} \quad (3.81)$$

となり、エントロピーは $S = S_0 - KL$ を用いて

$$S = \frac{1}{N} \sum_{\mu=1}^p \log \int_{-\infty}^{\infty} Dxg \left(y_{\mu}, x \sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu} \right) + \frac{1}{2N} \sum_{\mu=1}^p m_{u\mu}^2 \tilde{\chi}_{u\mu} + \frac{\log 2}{N} \sum_{k=1}^N \mathcal{H} \left(\frac{1+m_{wk}}{2} \right) \quad (3.82)$$

が得られる [31].

— 平均場近似アルゴリズム (離散値の場合) —

結合加重 w_k の更新

$$\tilde{\chi}_{wk} = \frac{1}{N} \sum_{\mu=1}^p x_{\mu k}^2 \chi_{u\mu} \quad (3.83)$$

$$h_{wk} = \frac{1}{\sqrt{N}} \sum_{\mu=1}^p x_{\mu k} m_{u\mu} + \tilde{\chi}_{wk} m_{wk} \quad (3.84)$$

$$m_{wk} = \tanh(h_{wk}) \quad (3.85)$$

$$\chi_{wk} = 1 - m_{wk}^2 \quad (3.86)$$

結合加重 w_k と拘束変数 u_μ の平均値 m_{wk} と $m_{u\mu}$ とそれらの共役変数の固定点を求めることで、近似的に事後確率 $P(\vec{w}|\xi^p)$ で平均化された m_{wk} を求めることができる。

拘束変数 u_μ の更新

$$\tilde{\chi}_{u\mu} = \frac{1}{N} \sum_{k=1}^N x_{\mu k}^2 \chi_{wk} \quad (3.87)$$

$$h_{u\mu} = \frac{1}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} m_{wk} - \tilde{\chi}_{u\mu} m_{u\mu} \quad (3.88)$$

$$m_{u\mu} = \frac{\int_{-\infty}^{\infty} Dx g'(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}{\int_{-\infty}^{\infty} Dx g(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})} \quad (3.89)$$

$$\delta_{u\mu} = \frac{\int_{-\infty}^{\infty} Dx g''(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}{\int_{-\infty}^{\infty} Dx g(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})} \quad (3.90)$$

$$\chi_{u\mu} = m_{u\mu}^2 - \delta_{u\mu} \quad (3.91)$$

表 3.1: 一般に Onsager 項 $\tilde{\chi}_{wk}, \tilde{\chi}_{u\mu}$ は他の変数に比べてゆっくり更新したほうが早く安定な解に収束することが知られている。また SOR 法を用いることでアトラクタの振動する効果が抑えられ、安定な解軌道になる。

3.3.3 連続値の場合

$\sum_{k=1}^N w_k^2 = N$ に従う球形模型の場合を考える。これより事前確率は $P_k(w_k) = \delta(w_k^2 - c_k)$ と仮定する。ただし $\sum_{k=1}^N c_k = N, (c_k \geq 0)$ である。離散値の場合と同様に定数項は無視でき、 $\tilde{\lambda}_{\mu k}(w_k)$ は

$$\tilde{\lambda}_{\mu k}(w_k) = -\frac{w_k^2}{2} \beta_{\mu k} + \tilde{h}_{\mu k} w_k \quad (3.92)$$

と表される。これより $\tilde{\chi}_{wk} := \sum_{\mu=1}^p \beta_{\mu k}$ と $h_{wk} := \sum_{\mu=1}^p \tilde{h}_{\mu k}$ を用いて

$$b_k(w_k) \propto \delta(w_k^2 - c_k) \exp\left[-\frac{w_k^2}{2} \tilde{\chi}_{wk} + w_k h_{wk}\right] = \delta(w_k^2 - c_k) \exp\left[-\frac{c_k}{2} \tilde{\chi}_{wk} + w_k h_{wk}\right] \quad (3.93)$$

となる。ここでも特性関数を用いた方法を採用する。特性関数を次のように定義される。

$$\begin{aligned} \zeta_k(\theta_k) &:= \log \sum_{w_k} b_k(w_k) e^{\theta_k w_k} = \text{Extr}_{\Lambda_k} \log \int_{-\infty}^{\infty} \frac{dw_k}{\sqrt{2\pi}} \exp\left[-\frac{\Lambda_k}{2} (w_k^2 - c_k) - \frac{c_k}{2} \tilde{\chi}_{wk} + w_k h_{wk} + \theta_k w_k\right] + \text{const.} \\ &= \text{Extr}_{\Lambda_k} \left\{ \frac{c_k \Lambda_k}{2} - \frac{c_k \tilde{\chi}_{wk}}{2} - \frac{1}{2} \log \Lambda_k + \frac{1}{2} \frac{(\theta_k + h_{wk})^2}{\Lambda_k} \right\} + \text{const.} \end{aligned} \quad (3.94)$$

w_k は平均 m_{wk} , 分散 χ_{wk} に従うと仮定すると,

$$m_{wk} = \left. \frac{\partial \zeta_k(\theta_k)}{\partial \theta_k} \right|_{\theta_k \rightarrow 0} = \frac{h_{wk}}{\Lambda_k} = \chi_{wk} h_{wk}, \quad \chi_{wk} = \left. \frac{\partial^2 \zeta_k(\theta_k)}{\partial \theta_k^2} \right|_{\theta_k \rightarrow 0} = \frac{1}{\Lambda_k} \quad (3.95)$$

が得られる。また事前分布の共役な変数 Λ_k の拘束条件から,

$$\lim_{\theta_k \rightarrow 0} \frac{\partial}{\partial \Lambda_k} \left\{ \frac{c_k \Lambda_k}{2} - \frac{1}{2} \log \Lambda_k + \frac{1}{2} \frac{(\theta_k + h_{wk})^2}{\Lambda_k} \right\} = 0 \quad \Rightarrow \quad c_k = \chi_{wk} + m_{wk}^2 \quad (3.96)$$

となり変数間の関係 $\chi_{wk} = c_k - m_{wk}^2$ が得られる。次に $b_\mu(\vec{w})$ での特性関数の計算を考える。

$$b_\mu(\vec{w}) \propto \int_{-\infty}^{\infty} \frac{d\Delta_\mu du_\mu}{(2\pi)^{1+\frac{N}{2}}} g(y_\mu, \Delta_\mu) \exp \left[-\frac{1}{2} \sum_{k=1}^N \Lambda_k (w_k^2 - c_k) + iu_\mu \Delta_\mu - \frac{iu_\mu}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} w_k + \sum_{k=1}^N w_k (h_{wk} - \tilde{h}_{\mu k}) \right] \quad (3.9)$$

上記からも分かるように $\tilde{\lambda}_{\mu k}(w_k)$ の2次の項は事前分布 $\delta(w_k^2 - c_k)$ の下では定数になるため、これ以降は2次の項は省略する。さて $b_\mu(\vec{w})$ の特性関数を次のように計算される。

$$\begin{aligned} \zeta_\mu(\vec{\theta}) &:= \log \sum_{\vec{w}} b_\mu(\vec{w}) e^{\vec{\theta}^T \vec{w}} \\ &= \log \int_{-\infty}^{\infty} \frac{d\Delta_\mu du_\mu d\vec{w}}{(2\pi)^{1+\frac{N}{2}}} g(y_\mu, \Delta_\mu) \exp \left[-\frac{1}{2} \sum_{k=1}^N \Lambda_k (w_k^2 - c_k) + iu_\mu \Delta_\mu - \frac{iu_\mu}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} w_k \right. \\ &\quad \left. + \sum_{k=1}^N w_k (\theta_k + h_{wk} - \tilde{h}_{\mu k}) \right] + \text{const.} \\ &= \frac{1}{2} \sum_{k=1}^N c_k \Lambda_k - \frac{1}{2} \sum_{k=1}^N \log \Lambda_k + \log \int_{-\infty}^{\infty} \frac{d\Delta_\mu du_\mu}{2\pi} g(y_\mu, \Delta_\mu) \\ &\quad \times \exp \left[iu_\mu \Delta_\mu + \frac{1}{2} \sum_{k=1}^N \frac{1}{\Lambda_k} \left(\theta_k + h_{wk} - \tilde{h}_{\mu k} - \frac{iu_\mu x_{\mu k}}{\sqrt{N}} \right)^2 \right] + \text{const.} \\ &= \frac{1}{2} \sum_{k=1}^N c_k \Lambda_k - \frac{1}{2} \sum_{k=1}^N \log \Lambda_k + \frac{1}{2} \sum_{k=1}^N \frac{(\theta_k + h_{wk} - \tilde{h}_{\mu k})^2}{\Lambda_k} + \log \int_{-\infty}^{\infty} Dxg(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu}) + \text{const.} \quad (3.9) \end{aligned}$$

ただし

$$\tilde{\chi}_{u\mu} := \frac{1}{N} \sum_{k=1}^N \frac{x_{\mu k}^2}{\Lambda_k}, \quad h_{u\mu} := \frac{1}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} \frac{\theta_k + h_{wk} - \tilde{h}_{\mu k}}{\Lambda_k} \quad (3.99)$$

を用いた。これより構成的に

$$m_{wk} = \left. \frac{\partial \zeta_\mu(\vec{\theta})}{\partial \theta_k} \right|_{\vec{\theta} \rightarrow 0} = \frac{h_{wk} - \tilde{h}_{\mu k}}{\Lambda_k} + \frac{x_{\mu k} m_{u\mu}}{\sqrt{N} \Lambda_k} \quad (3.100)$$

$$\chi_{wk} = \left. \frac{\partial^2 \zeta_\mu(\vec{\theta})}{\partial \theta_k^2} \right|_{\vec{\theta} \rightarrow 0} = \frac{1}{\Lambda_k} + \frac{x_{\mu k}^2}{N \Lambda_k^2} (\delta_{u\mu} - m_{u\mu}^2) = \frac{1}{\Lambda_k} - \frac{x_{\mu k}^2}{N \Lambda_k^2} \chi_{u\mu} \quad (3.101)$$

が導出される。ここで

$$m_{u\mu} := \frac{\int_{-\infty}^{\infty} Dxg'(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}{\int_{-\infty}^{\infty} Dxg(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}, \quad \delta_{u\mu} := \frac{\int_{-\infty}^{\infty} Dxg''(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}{\int_{-\infty}^{\infty} Dxg(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}, \quad \chi_{u\mu} := m_{u\mu}^2 - \delta_{u\mu} \quad (3.102)$$

を用いた。続いて Λ_k の極値条件から、

$$\frac{c_k}{2} - \frac{1}{2\Lambda_k} - \frac{(h_{wk} - \tilde{h}_{\mu k})^2}{2\Lambda_k^2} + \frac{\delta_{u\mu}}{2} \frac{\partial \tilde{\chi}_{u\mu}}{\partial \Lambda_k} + m_{u\mu} \frac{\partial h_{u\mu}}{\partial \Lambda_k} = 0 \quad \Rightarrow \quad c_k = \chi_{wk} + m_{wk}^2 \quad (3.103)$$

が得られる。これは式 (3.96) と同じである。さて $h_{wk} \sim O(1)$ から、漸近的に $\tilde{h}_{\mu k} = \frac{A}{\sqrt{N}} + \frac{B}{N}$ であると仮定する。これより式 (3.100) と式 (3.101) から $A = \frac{x_{\mu k} m_{u\mu}}{\chi_{wk} \Lambda_k} \simeq x_{\mu k} m_{u\mu}$ と $B = \frac{h_{wk} x_{\mu k}^2}{\Lambda_k^2 \chi_{wk}} \chi_{u\mu} \simeq m_{wk} x_{\mu k}^2 \chi_{u\mu}$ が得られる。ただし $\chi_{wk} \Lambda_k = 1 - \frac{x_{\mu k}^2 \chi_{u\mu}}{N \Lambda_k} \simeq 1$ を用いた。これをまとめると

$$\tilde{\chi}_{u\mu} = \frac{1}{N} \sum_{k=1}^N x_{\mu k}^2 \chi_{wk}, \quad h_{u\mu} = \frac{1}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} m_{wk} - \tilde{\chi}_{u\mu} m_{u\mu} \quad (3.113)$$

$$\tilde{\chi}_{wk} = \frac{1}{N} \sum_{\mu=1}^p x_{\mu k}^2 \chi_{u\mu}, \quad h_{wk} = \frac{1}{\sqrt{N}} \sum_{\mu=1}^p x_{\mu k} m_{u\mu} + \tilde{\chi}_{wk} m_{wk} \quad (3.114)$$

となり、Onsager の反跳場 $\tilde{\chi}_{wk}, \tilde{\chi}_{u\mu}$ が導出される。以上のアルゴリズムをまとめたのが表 3.2 である。

平均場近似アルゴリズム (連続値の場合)

結合加重 w_k の更新

$$\tilde{\chi}_{wk} = \frac{1}{N} \sum_{\mu=1}^p x_{\mu k}^2 \chi_{u\mu} \quad (3.104)$$

$$h_{wk} = \frac{1}{\sqrt{N}} \sum_{\mu=1}^p x_{\mu k} m_{u\mu} + \tilde{\chi}_{wk} m_{wk} \quad (3.105)$$

$$m_{wk} = \chi_{wk} h_{wk} \quad (3.106)$$

$$\chi_{wk} = c_k - m_{wk}^2 \quad (3.107)$$

結合加重 w_k と拘束変数 u_μ の平均値 m_{wk} と $m_{u\mu}$ とそれらの共役変数の固定点を求めることで、近似的に事後確率 $P(\vec{w}|\xi^p)$ で平均化された m_{wk} を求めることができる。

拘束変数 u_μ の更新

$$\tilde{\chi}_{u\mu} = \frac{1}{N} \sum_{k=1}^N x_{\mu k}^2 \chi_{wk} \quad (3.108)$$

$$h_{u\mu} = \frac{1}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} m_{wk} - \tilde{\chi}_{u\mu} m_{u\mu} \quad (3.109)$$

$$m_{u\mu} = \frac{\int_{-\infty}^{\infty} Dx g'(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}{\int_{-\infty}^{\infty} Dx g(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})} \quad (3.110)$$

$$\delta_{u\mu} = \frac{\int_{-\infty}^{\infty} Dx g''(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}{\int_{-\infty}^{\infty} Dx g(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})} \quad (3.111)$$

$$\chi_{u\mu} = m_{u\mu}^2 - \delta_{u\mu} \quad (3.112)$$

表 3.2: 一般に Onsager 項 $\tilde{\chi}_{wk}, \tilde{\chi}_{u\mu}$ は他の変数に比べてゆっくり更新したほうが早く安定な解に収束することが知られている。また SOR 法を用いることでアトラクタの振動する効果が抑えられ、安定な解軌道になる。

3.3.4 アルゴリズム

技術的な問題として変数の自由度が大きい場合は、収束が遅くなることが知られている。そこで $\chi_{wk}, \tilde{\chi}_{wk}, \chi_{u\mu}, \tilde{\chi}_{u\mu}$ に対して、次のような巨視化を行う。

$$\chi_w = \frac{1}{N} \sum_{k=1}^N \chi_{wk}, \quad \tilde{\chi}_w = \frac{1}{N} \sum_{k=1}^N \tilde{\chi}_{wk}, \quad \chi_u = \frac{1}{p} \sum_{\mu=1}^p \chi_{u\mu}, \quad \tilde{\chi}_u = \frac{1}{p} \sum_{\mu=1}^p \tilde{\chi}_{u\mu} \quad (3.115)$$

これによりアルゴリズムの収束時間が短くなることが見て取れる。また実際 $N = 5000$ のとき、結合加重が離散値の場合の記憶容量問題に対して、式 (3.82) の平均場近似法で求めたエントロピー S と秩序変数 q 結果が図 3.8 である。マーカが平均場近似法の結果であり、実線はそれに対応するレプリカ解析の結果である。図 3.8 よりレプリカ解析の結果と整合性が取れていることが分かる。

3.4 まとめ

本章では i.i.d. パターンに対するパーセプトロン学習について議論し、以下のことが解明された。

1. パーセプトロン学習を統計力学的手法で定式化を行い、自然な要請として、パーセプトロンの学習能力を評価するためには、エントロピーの評価が必要であることを説明した。
2. パーセプトロン学習問題に対してのレプリカ法を導出し、結合加重が離散と連続の両方で得られている既存結果を紹介した。
3. システムサイズが有限であるが十分大きな場合に、与えられた参照例題の下での事後確率から求まる結合加重 w_k の推定量 m_{wk} を効率的に導出できる平均場近似法を導出した。

最後にこれまで説明してきた既存手法の扱ってきた入力統計性に関して、暗黙のうちに「独立同分布の仮定」をおいていることに注意しなければならない。しかしながらパターン認識に代表される知的情報処理のできる学習器を用いた「データ解析」において、扱う実データは多くの場合独立同分布に従っているとは言いがたい。上記の解析手

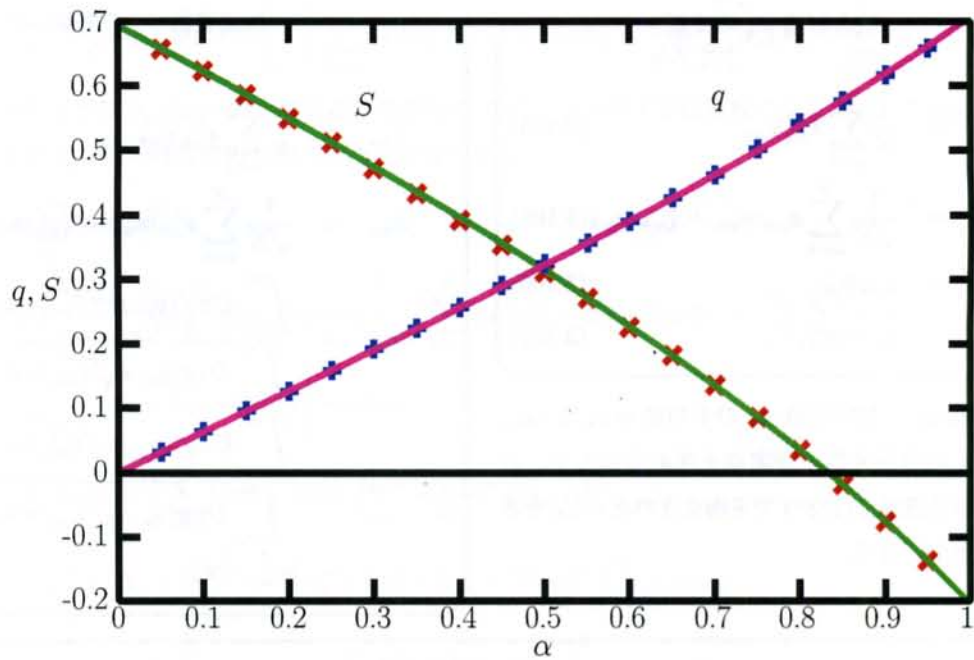


図 3.8: $N = 5000$ のときの記憶容量問題について解析を行った。パターン比 α と秩序変数 q とエントロピー S の図である。点線が平均場近似の数値結果を表す。この実験条件として 1 試行回数で式 (3.115) の巨視化を行ったアルゴリズムを用いた。比較として、レプリカ解析の結果（実線）と一緒にプロットしている。

法は独立同分布に特化した解析であるため、入力間に相関のあるような例題に対して既存手法を用いても正しくクラス識別能力を評価できるとは言い難い。すなわち既存研究で扱われてきた解析手法を含む形で「より広いクラス」のパターンに対しても統計力学的手法を拡張する必要がある。それに対する我々の提案手法を次章で説明する。

第4章 相関の特徴付けと定式化

この章では本研究の主要な解析方法についてまとめる。これまで議論したように、パーセプトロン学習に対する既存の理論研究では、与えられた入力に対して「独立同分布の仮定」を暗黙のうちにしていることが少なくない。その1つの理由として、入力が各々独立である課題は一般的に言って、入力に相関が含まれている課題に比べて解析的に扱いやすく、また確率伝搬法などの実効的なアルゴリズムが構成しやすいという利点が挙げられる。しかしながら実用の面で考えると、実データに対して多くの場合、入力に対する「独立同分布の仮定」の妥当性が保証されているとは言いがたい。そのため独立同分布の仮定のままで実データに適用した場合、パーセプトロンを用いて記憶容量問題や教師推定問題などを解析する際に、「系統的な誤り（相関が含まれるパターンに対して i.i.d. パターンだと仮定することによる、モデルミスマッチングに起因する予測誤差に含まれる誤りなど）」が発生する可能性が否定できない。このような状況では、i.i.d. パターンに対する学習器のクラス識別に対する知見やアルゴリズムなど、既存研究の方法では、クラス識別器の性能を十分に評価できるとは言い難い。さらにもう1つの例として、実データ解析において、入力が各々直交化する「前処理」をあらかじめデータに施している場合もある。扱う問題に依存するが、適切な「相関」を積極的に取り入れることで、クラス分類能力が上がるのが期待されるため、「前処理（もしくは入力間の相関）」が学習効率に与える影響を考察する必要がある。よって我々は入力間に相関が含まれるデータに対する、パーセプトロンのデータ解析能力の典型評価や実効的なアルゴリズムの開発などの理論的整備を以下でも説明する。この章の構成は以下の通りである。はじめに入力間に含まれる「相関」の特徴付けを行う。それを下にレプリカ法を用いてエントロピーを評価する方法を説明する。次に十分大きな有限のシステムサイズでレプリカ法の結果に漸近的に一致すると期待される、Thouless-Anderson-Palmer 法の、入力間に相関のある場合への拡張を行う。最後にまとめる述べる。

4.1 相関の特徴付け

はじめに相関のある例題集合の特徴付けについて議論しよう。与えられた参照例題 ξ^p に含まれる p 個の入力 $\vec{x}_\mu \in \mathbf{R}^N$, ($\mu = 1, 2, \dots, p$) を用いて、入力行列を $X = \frac{1}{\sqrt{N}} \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_p\}^T \in \mathcal{M}_{p \times N}$ のように定義する。単に入力を並べた行列を定義するだけでは問題の構造を変えたことにはならない。しかしながら、行列の性質を用いることで本質的に、入力間に含まれる相関を扱うことができるため、既存研究で扱われた手法の拡張が可能になる。

さて第3章で見たように、パーセプトロンの性能評価を行うためにはエントロピーの典型値を評価する必要がある。なぜなら第3章のエントロピーの議論は入力の統計性によらず、相関のある場合にもそのまま流用することができる。そこで再度ここで分配関数 $Z(\xi^p)$ とエントロピー S を

$$Z(\xi^p) = \sum_{\vec{w}} P(\vec{w}) e^{-\beta \mathcal{H}(\vec{w}|\xi^p)} = \sum_{\vec{w}} P(\vec{w}) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^p} \prod_{\mu=1}^p g(y_\mu, \Delta_\mu) \exp \left[i \vec{u}^T (\vec{\Delta} - X \vec{w}) \right] \quad (4.1)$$

$$S = S_0 + \phi \quad (4.2)$$

のように定義する。 ϕ は問題に依存して

$$\phi = \begin{cases} \lim_{n \rightarrow 0} \frac{\partial \psi(n)}{\partial n} & \text{記憶容量問題} \\ \lim_{n \rightarrow 1} \frac{\partial \psi(n)}{\partial n} & \text{教師推定問題} \end{cases} \quad (4.3)$$

と定義される。ただしキュムラント母関数

$$\psi(n) : = \lim_{N \rightarrow \infty} \frac{1}{N} \log [Z^n(\xi^p)]_{\xi^p} \quad (4.4)$$

を用いた。式 (4.3) の $n \rightarrow 0$ と $n \rightarrow 1$ の説明は 46 ページを参照されたい。式 (4.1) と式 (4.4) から相関のある例題集合に対する評価は、実質的には $e^{-i\vec{u}^T X \vec{w}}$ の入力行列 X でのアンサンブル平均に帰結される。そこで我々は入力行列であるランダム長方形列のアンサンブルの特徴付けとして次の仮定をおいて、既存手法の拡張を試みる。

相関の特徴付け 自明な置換を除いて、入力行列 X は

$$X = UDV^T \in \mathcal{M}_{p \times N} \quad (4.5)$$

のように特異値分解ができることが知られている。ただし D は対角成分に特異値 d_k , ($k = 1, 2, \dots, \min(p, N)$) を持ち、非対角成分が 0 の対角行列であり、 U と V はそれぞれ $p \times p$ と $N \times N$ の直交行列である。さらに説明すれば、 U は XX^T の固有ベクトルであり、 V は $X^T X$ の固有ベクトルである。

式 (4.5) を踏まえて我々は入力間に含まれる相関の特徴付けとして、以下の仮定に従う入力行列 X のアンサンブルに限定し、議論を展開することが可能である。

—— ランダム行列アンサンブルの特徴付け ——

—— 固有値分布の特徴付け ——

$N, p \rightarrow \infty$ で自己相関行列 $X^T X \in \mathcal{M}_{N \times N}$ の固有値分布

$$\rho_{X^T X}(\lambda) := \lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr} \delta(\lambda - X^T X) \quad (4.6)$$

が漸近固有値分布 $\rho(\lambda)$ に確率 1 で収束する。つまり X は固有値分布が一定のアンサンブルから生成されるランダム行列として特徴付けられる。

—— Haar 測度からの生成 ——

U と V が独立に Haar 測度から一様に生成される。

これは第2章で説明したランダム長方形列のときのアンサンブルに他ならない。さてこのとき $\vec{w} := \{w_1, w_2, \dots, w_N\}^T \in \mathbf{R}^N$ と $\vec{u} := \{u_1, u_2, \dots, u_p\}^T \in \mathbf{R}^p$ が与えられた下で、特に

$$\vec{w}^T \vec{w} = Nx, \quad \vec{u}^T \vec{u} = py \quad (4.7)$$

が成り立つとき、ランダム行列の積分公式

$$F(x, y) = \text{Extr}_{\Lambda_x, \Lambda_y} \left\{ \frac{\Lambda_x - 1}{2} + \alpha \frac{\Lambda_y - 1}{2} - \frac{\alpha - 1}{2} \log \Lambda_y - \frac{1}{2} \int d\lambda \rho(\lambda) \log(\Lambda_x \Lambda_y + xy\lambda) \right\} \quad (4.8)$$

が得られることをすでに説明した。この積分公式が相関例題に対するパーセプトロンの学習能力の性能評価の実効的な解決につながるものが以下で説明される。

4.2 レプリカ解析

n 個の N 次元ベクトルを $\vec{w}_a := \{w_{1a}, w_{2a}, \dots, w_{Na}\}^T \in \mathbf{R}^N$ で表し、2 種類の n 個の p 次元ベクトルをそれぞれ $\vec{\Delta}_a := \{\Delta_{1a}, \Delta_{2a}, \dots, \Delta_{pa}\}^T$, $\vec{u}_a := \{u_{1a}, u_{2a}, \dots, u_{pa}\}^T \in \mathbf{R}^p$, ($a = 1, 2, \dots, n$) で表す。これより導出は後述する

として、キウムラント母関数 $\psi(n) := \lim_{N \rightarrow \infty} \frac{1}{N} \log [Z^n(\xi^p)]$ は、

$$\begin{aligned}
 \psi(n) &= \lim_{N \rightarrow \infty} \frac{1}{N} \log \sum_{\vec{y}} P(\vec{y}) \sum_{\vec{w}_1, \dots, \vec{w}_n} \int_{-\infty}^{\infty} \prod_{a=1}^n P(\vec{w}_a) \frac{d\vec{\Delta}_a d\vec{u}_a}{(2\pi)^{pn}} \prod_{\mu=1}^p g(y_\mu, \Delta_{\mu a}) \\
 &\quad \times \sum_X P(X) \exp \left[i \sum_{a=1}^n \vec{u}_a^T (\vec{\Delta}_a - X \vec{w}_a) \right] \\
 &= \text{Extr}_{Q_w, \tilde{Q}_w, Q_u, \tilde{Q}_u} \left\{ \mathcal{F}(Q_w, Q_u) + \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \exp \left[i \vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u} \right] \prod_{a=1}^n g(y, \Delta_a) \right. \\
 &\quad \left. + \frac{\alpha}{2} \text{Tr} Q_u \tilde{Q}_u - \frac{1}{2} \text{Tr} Q_w \tilde{Q}_w + \log \sum_{\vec{w}} P(\vec{w}) e^{\frac{1}{2} \vec{w}^T \tilde{Q}_w \vec{w}} \right\} \quad (4.9)
 \end{aligned}$$

で表される。ただし2行目での $\vec{w}, \vec{u}, \vec{\Delta}$ は n 次元ベクトルを表し、さらに4つの n 次正方行列 $Q_w := \{q_{wab}\}, Q_u := \{q_{uab}\}, \tilde{Q}_w := \{\tilde{q}_{wab}\}, \tilde{Q}_u := \{\tilde{q}_{uab}\} \in \mathcal{M}_{n \times n}$ と

$$\begin{aligned}
 \mathcal{F}(Q_w, Q_u) : &= \text{Extr}_{\Lambda_w, \Lambda_u} \left\{ \frac{1}{2} \text{Tr} (\Lambda_w - I) + \frac{\alpha}{2} \text{Tr} (\Lambda_u - I) - \frac{\alpha-1}{2} \log \det \Lambda_u \right. \\
 &\quad \left. - \frac{1}{2} \int d\lambda \rho(\lambda) \log \det (\Lambda_w \Lambda_u + \lambda Q_w Q_u) \right\} \quad (4.10)
 \end{aligned}$$

を用いた。さらに2種類の秩序変数をそれぞれ $q_{wab} = \frac{\vec{w}_a^T \vec{w}_b}{N}$ と $q_{uab} = \frac{\vec{u}_a^T \vec{u}_b}{p}$ で表した。以下では式 (4.9) と式 (4.10) の導出について簡単に説明する。

4.2.1 $\mathcal{F}(Q_w, Q_u)$ の導出

$\vec{w}_a, \vec{u}_a$ が固定された下で $\exp \left[-i \sum_{a=1}^n \vec{u}_a^T X \vec{w}_a \right]$ のアンサンブル平均をとることを考えよう。ここでの議論から分かることだが、結果的にランダム長方形列の積分公式 $F(x, y)$ の引数が n 次正方行列に拡張されたものと解釈することができる。さて $\vec{w}_a, \vec{u}_a$ が固定されているため、アンサンブル平均をすることは、新たな変数 $\vec{t}_a = V^T \vec{w}_a$ と $\vec{\tau}_a = U^T \vec{u}_a$ を用いて

$$\vec{w}_a^T \vec{w}_b = \vec{t}_a^T \vec{t}_b = N q_{wab}, \quad \vec{u}_a^T \vec{u}_b = \vec{\tau}_a^T \vec{\tau}_b = p q_{uab} \quad (4.11)$$

の拘束条件の下で平均をとることと一致する。そこで

$$\begin{aligned}
 \mathcal{F}(Q_w, Q_u) : &= \lim_{N \rightarrow \infty} \frac{1}{N} \log \frac{\int_{-\infty}^{\infty} \prod_{a=1}^n d\vec{t}_a d\vec{\tau}_a \prod_{a>b} \delta(\vec{t}_a^T \vec{t}_b - N q_{wab}) \delta(\vec{\tau}_a^T \vec{\tau}_b - p q_{uab}) \exp \left[-i \sum_{a=1}^n \vec{\tau}_a^T D \vec{t}_a \right]}{\int_{-\infty}^{\infty} \prod_{a=1}^n d\vec{t}_a d\vec{\tau}_a \prod_{a>b} \delta(\vec{t}_a^T \vec{t}_b - N q_{wab}) \delta(\vec{\tau}_a^T \vec{\tau}_b - p q_{uab})} \\
 &= \mathcal{A}(D) - \mathcal{A}(0) \quad (4.12)
 \end{aligned}$$

とき、 $\mathcal{A}(D)$ を評価すると、

$$\begin{aligned}
\mathcal{A}(D) &:= \text{Extr}_{\Gamma_w, \Gamma_u} \lim_{N \rightarrow \infty} \frac{1}{N} \log \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{N+p}{2}n}} \prod_{a=1}^n d\vec{t}_a d\vec{\tau}_a \exp \left[-\frac{1}{2} \sum_{a,b} \Gamma_{wab} \left(\sum_{k=1}^N t_{ka} t_{kb} - N q_{wab} \right) \right. \\
&\quad \left. - \frac{1}{2} \sum_{a,b} \Gamma_{uab} \left(\sum_{\mu=1}^p \tau_{\mu a} \tau_{\mu b} - p q_{uab} \right) - i \sum_{a=1}^n \sum_{\mu=1}^p \tau_{\mu a} \left(\sum_{k=1}^N d_{\mu k} t_{ka} \right) \right] \\
&= \text{Extr}_{\Gamma_w, \Gamma_u} \left\{ \frac{1}{2} \text{Tr} Q_w \Gamma_w + \frac{\alpha}{2} \text{Tr} Q_u \Gamma_u - \frac{\alpha}{2} \log \det \Gamma_u + \lim_{N \rightarrow \infty} \frac{1}{N} \log \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{Nn}{2}}} \prod_{a=1}^n d\vec{t}_a \right. \\
&\quad \left. \times \exp \left[-\frac{1}{2} \sum_{\mu=1}^p \sum_{a,b} \sum_{k=1}^N d_{\mu k} t_{ka} \Gamma_{uab}^{-1} \sum_{l=1}^N d_{\mu l} t_{lb} - \frac{1}{2} \sum_{a,b} \Gamma_{wab} \sum_{k=1}^N t_{ka} t_{kb} \right] \right\} \\
&= \text{Extr}_{\Gamma_w, \Gamma_u} \left\{ \frac{1}{2} \text{Tr} Q_w \Gamma_w + \frac{\alpha}{2} \text{Tr} Q_u \Gamma_u - \frac{\alpha}{2} \log \det \Gamma_u - \frac{1}{2} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \log \det (\Gamma_w + \lambda_k \Gamma_u^{-1}) \right\} \quad (4.13)
\end{aligned}$$

となる。ただし $D^T D$ の成分

$$\sum_{\mu=1}^p d_{\mu k} d_{\mu l} = \lambda_k \delta_{kl} \quad (4.14)$$

を用いた。さらに固有値分布 $\rho(\lambda) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \delta(\lambda - \lambda_k)$ を用いれば、

$$\mathcal{A}(D) = \text{Extr}_{\Gamma_w, \Gamma_u} \left\{ \frac{1}{2} \text{Tr} Q_w \Gamma_w + \frac{\alpha}{2} \text{Tr} Q_u \Gamma_u - \frac{\alpha-1}{2} \log \det \Gamma_u - \frac{1}{2} \int d\lambda \rho(\lambda) \log \det (\Gamma_w \Gamma_u + \lambda) \right\} \quad (4.15)$$

となる。また $\mathcal{A}(0) = \frac{n}{2} + \frac{n\alpha}{2} + \frac{\alpha}{2} \log \det Q_u + \frac{1}{2} \log \det Q_w$ であるので、それぞれを式 (4.12) に代入することで

$$\begin{aligned}
\mathcal{F}(Q_w, Q_u) &= \text{Extr}_{\Lambda_w, \Lambda_u} \left\{ \frac{1}{2} \text{Tr} (\Lambda_w - I) + \frac{\alpha}{2} \text{Tr} (\Lambda_u - I) - \frac{\alpha-1}{2} \log \det \Lambda_u \right. \\
&\quad \left. - \frac{1}{2} \int d\lambda \rho(\lambda) \log \det (\Lambda_w \Lambda_u + \lambda Q_w Q_u) \right\} \quad (4.16)
\end{aligned}$$

が得られる。ただし $Q_w \Gamma_w = \Lambda_w$ と $\Gamma_u Q_u = \Lambda_u$ とおいた。

4.2.2 その他の項

N 個の n 次元ベクトル $\vec{w}_k, (k=1, 2, \dots, N)$ に関する平均は、

$$\begin{aligned}
&\text{Extr}_{\vec{Q}_w} \lim_{N \rightarrow \infty} \frac{1}{N} \log \sum_{\{\vec{w}_k\}} \prod_{k=1}^N P(\vec{w}_k) \exp \left[\frac{1}{2} \text{Tr} \tilde{Q}_w \left(\sum_{k=1}^N \vec{w}_k \vec{w}_k^T - N Q_w \right) \right] \\
&= \text{Extr}_{\vec{Q}_w} \left\{ -\frac{1}{2} \text{Tr} \tilde{Q}_w Q_w + \log \sum_{\vec{w}} P(\vec{w}) e^{\frac{1}{2} \vec{w}^T \tilde{Q}_w \vec{w}} \right\} \quad (4.17)
\end{aligned}$$

となる。また2種類の p 個の n 次元ベクトル $\vec{u}_\mu, \vec{\Delta}_\mu, (\mu=1, 2, \dots, p)$ に関する平均は

$$\begin{aligned}
&\text{Extr}_{\vec{Q}_u} \lim_{N \rightarrow \infty} \frac{1}{N} \log \sum_{\vec{y}} P(\vec{y}) \int_{-\infty}^{\infty} \prod_{\mu=1}^p \frac{d\vec{u}_\mu d\vec{\Delta}_\mu}{(2\pi)^{pn}} \prod_{a=1}^n g(y_\mu, \Delta_{\mu a}) \exp \left[i \sum_{\mu=1}^p \vec{u}_\mu^T \vec{\Delta}_\mu - \frac{1}{2} \text{Tr} \tilde{Q}_u \left(\sum_{\mu=1}^p \vec{u}_\mu \vec{u}_\mu^T - p Q_u \right) \right] \\
&= \text{Extr} \left\{ \frac{\alpha}{2} \text{Tr} \tilde{Q}_u Q_u + \alpha \log \sum_{\vec{y}} P(\vec{y}) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} e^{-\frac{1}{2} \vec{\Delta}^T \tilde{Q}_u^{-1} \vec{\Delta}}}{(2\pi)^{\frac{n}{2}} |\tilde{Q}_u|^{\frac{1}{2}}} \prod_{a=1}^n g(y, \Delta_a) \right\} \quad (4.18)
\end{aligned}$$

となる。式 (4.16)、式 (4.17) と式 (4.18) を用いて式 (4.9) が導出される。

4.2.3 レプリカ対称解

第3章同様に、式(4.9)の表現において、いくつかの特殊な場合を除いて解析的に評価することが困難である。しかしながら解表現を何らかの形に限定することで見通しがよくなり、その困難を回避することができる。そこで我々は議論を簡単にするために次のようなRS解を仮定する。

$$q_{wab} = \begin{cases} \chi_w + q_w & a = b \\ q_w & a \neq b \end{cases} \quad q_{uab} = \begin{cases} \chi_u + q_u & a = b \\ q_u & a \neq b \end{cases} \quad (4.19)$$

このとき Q_w の固有値は $n-1$ 個の χ_w と 1 個の $\chi_w + nq_w$ であり、 Q_u の固有値は $n-1$ 個の χ_u と 1 個の $\chi_u + nq_u$ である。ここで $z^n = 1$ である $z \in \mathbf{C}$ を用いて (ただし $z \neq 1$ とする)、 Q_w と Q_u の固有ベクトルは、それぞれに対応させて $n-1$ 個の $\vec{e}_a \propto \{z^{a+1}, z^{a+2}, z^{a+3}, \dots, z^{a+n-1}, z^a\}^T$, ($a = 1, 2, \dots, n-1$) と 1 個の $\vec{e}_n \propto \{1, 1, \dots, 1\}^T$ となる。よって Λ_w と Λ_u の固有ベクトルを $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$ と見なすと、相関の効果を組み込んだ $\mathcal{F}(Q_w, Q_u)$ は

$$\mathcal{F}(Q_w, Q_u) = (n-1)F(\chi_w, \chi_u) + F(\chi_w + nq_w, \chi_u + nq_u) \quad (4.20)$$

と簡単化される [195]。ただし Tr と $\det$ は固有ベクトル方向に向けると同時対角化でき、積分公式が分解できることを用いた。さらに

$$\tilde{q}_{wab} = \begin{cases} \tilde{\chi}_w & a = b \\ \tilde{q}_w & a \neq b \end{cases} \quad \tilde{q}_{uab} = \begin{cases} \tilde{\chi}_u & a = b \\ \tilde{q}_u & a \neq b \end{cases} \quad (4.21)$$

とすると、

$$\frac{\alpha}{2} \text{Tr} Q_w \tilde{Q}_w - \frac{1}{2} \text{Tr} Q_u \tilde{Q}_u = \frac{n\alpha}{2} \tilde{\chi}_u (\chi_u + q_u) + \frac{n(n-1)\alpha}{2} \tilde{q}_u q_u - \frac{n}{2} \tilde{\chi}_w (\chi_w + q_w) - \frac{n(n-1)}{2} \tilde{q}_w q_w \quad (4.22)$$

となる。また

$$\alpha \log \sum_y P(y) \int_{-\infty}^{\infty} d\vec{\Delta} P(\vec{\Delta}) \prod_{a=1}^n g(y, \Delta_a) = \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dx g\left(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}\right) \right)^n \quad (4.23)$$

$$\log \sum_{\vec{w}} P(\vec{w}) e^{\frac{1}{2} \vec{w}^T \tilde{Q}_w \vec{w}} = \log \int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n \quad (4.24)$$

と計算できる。これよりキュムラント母関数は

$$\begin{aligned} \psi(n) = & \text{Extr}_{\theta, \tilde{\theta}} \left\{ (n-1)F(\chi_w, \chi_u) + F(\chi_w + nq_w, \chi_u + nq_u) + \log \int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n \right. \\ & + \frac{n\alpha}{2} \tilde{\chi}_u (\chi_u + q_u) + \frac{n(n-1)\alpha}{2} \tilde{q}_u q_u - \frac{n}{2} \tilde{\chi}_w (\chi_w + q_w) - \frac{n(n-1)}{2} \tilde{q}_w q_w \\ & \left. + \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dx g\left(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}\right) \right)^n \right\} \quad (4.25) \end{aligned}$$

となる。ただし $\theta := \{\chi_w, q_w, \chi_u, q_u\}$, $\tilde{\theta} := \{\tilde{\chi}_w, \tilde{q}_w, \tilde{\chi}_u, \tilde{q}_u\}$ とした。また

$$\begin{aligned} \frac{\partial \psi(n)}{\partial n} = & \text{E}_{\tilde{\theta}}^{\text{extr}} \left\{ F(\chi_w, \chi_u) + q_w \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_w} + q_u \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_u} \right. \\ & + \frac{\int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n \log \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)}{\int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n} \\ & + \frac{\frac{\alpha}{2} \tilde{\chi}_u (\chi_u + q_u) + \frac{(2n-1)\alpha}{2} \tilde{q}_u q_u - \frac{1}{2} \tilde{\chi}_w (\chi_w + q_w) - \frac{(2n-1)}{2} \tilde{q}_w q_w}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)^n \log \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)} \\ & \left. + \alpha \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)^n}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)^n} \right\} \quad (4.26) \end{aligned}$$

となる。式 (4.25) と式 (4.26) の極値条件の導出は付録 B を参照するとして、結果的に

$$\chi_u + q_u = - \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)^n \left(\frac{\int_{-\infty}^{\infty} Dxg''(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})}{\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})} \right)}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)^n} \quad (4.27)$$

$$q_u = - \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)^n \left(\frac{\int_{-\infty}^{\infty} Dxg'(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})}{\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})} \right)^2}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)^n} \quad (4.28)$$

と

$$\chi_w + q_w = \frac{\int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n \left(\frac{\sum_w w^2 P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}}}{\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}}} \right)}{\int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n}, \quad (4.29)$$

$$q_w = \frac{\int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n \left(\frac{\sum_w w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}}}{\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}}} \right)^2}{\int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n} \quad (4.30)$$

が得られる。また共役変数 $\tilde{\chi}_w, \tilde{q}_w, \tilde{\chi}_u, \tilde{q}_u$ は

$$\begin{aligned} \tilde{\chi}_w &= 2 \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_w} - (n-1)\tilde{q}_w, & \tilde{q}_w &= \frac{2}{n} \left\{ \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_w} - \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} \right\}, \\ \tilde{\chi}_u &= -\frac{2}{\alpha} \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_u} - (n-1)\tilde{q}_u, & \tilde{q}_u &= -\frac{2}{n\alpha} \left\{ \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_u} - \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} \right\} \end{aligned} \quad (4.31)$$

となる。ただし積分公式の微分は

$$\begin{aligned} 2 \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} &= \frac{\Lambda_w - 1}{\chi_w}, & \frac{2}{\alpha} \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} &= \frac{\Lambda_u - 1}{\chi_u}, \\ 2 \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_w} &= \frac{\tilde{\Lambda}_w - 1}{\chi_w + nq_w}, & \frac{2}{\alpha} \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_u} &= \frac{\tilde{\Lambda}_u - 1}{\chi_u + nq_u} \end{aligned} \quad (4.32)$$

となる。さらに補助変数 $\Lambda_w, \Lambda_u, \tilde{\Lambda}_w, \tilde{\Lambda}_u$ は

$$\begin{aligned} \Lambda_w &= \alpha(\Lambda_u - 1) + 1, & \frac{1}{\Lambda_u} &= \int \frac{d\lambda \rho(\lambda)}{\Lambda_w \Lambda_u + \lambda \chi_w \chi_u}, \\ \tilde{\Lambda}_w &= \alpha(\tilde{\Lambda}_u - 1) + 1, & \frac{1}{\tilde{\Lambda}_u} &= \int \frac{d\lambda \rho(\lambda)}{\tilde{\Lambda}_w \tilde{\Lambda}_u + \lambda(\chi_w + nq_w)(\chi_u + nq_u)} \end{aligned} \quad (4.33)$$

を満たす。上記の表現を用いて、記憶容量問題の場合 ($n \rightarrow 0$) と教師推定問題の場合 ($n \rightarrow 1$) の解析ができるが、結合加重が離散値か連続値かで場合分けする必要がある。

4.2.4 結合加重が離散の場合

結合加重が離散値 ($w = \pm 1$) で等確率の場合を考える。結合加重の平均に関係する部分は

$$\begin{aligned} & -\frac{n}{2} \tilde{\chi}_w (\chi_w + q_w) - \frac{n(n-1)}{2} \tilde{q}_w q_w + \log \int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw \sqrt{\tilde{q}_w}} \right)^n \\ &= -\frac{n}{2} \tilde{\chi}_w (\chi_w + q_w - 1) - \frac{n(n-1)}{2} \tilde{q}_w q_w - \frac{n}{2} \tilde{q}_w + \log \int_{-\infty}^{\infty} Dz \cosh^n(z \sqrt{\tilde{q}_w}) \end{aligned} \quad (4.34)$$

となり、式 (4.29) と式 (4.30) から

$$\chi_w = 1 - q_w, \quad q_w = \frac{\int_{-\infty}^{\infty} Dz \cosh^n(z \sqrt{\tilde{q}_w}) \tanh^2(z \sqrt{\tilde{q}_w})}{\int_{-\infty}^{\infty} Dz \cosh^n(z \sqrt{\tilde{q}_w})} \quad (4.35)$$

が得られる。

4.2.5 結合加重が連続の場合

結合加重が連続値 ($\sum_{k=1}^N w_k^2 = N$) である場合を考える。結合加重の平均に関係する部分は

$$\begin{aligned} & -\frac{n}{2} \tilde{\chi}_w (\chi_w + q_w) - \frac{n(n-1)}{2} \tilde{q}_w q_w + \log \int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw \sqrt{\tilde{q}_w}} \right)^n \\ &= -\frac{n}{2} \tilde{\chi}_w (\chi_w + q_w) - \frac{n(n-1)}{2} \tilde{q}_w q_w + \frac{n\Lambda}{2} - \frac{n-1}{2} \log(\Lambda - \tilde{\chi}_w + \tilde{q}_w) - \frac{1}{2} \log(\Lambda - \tilde{\chi}_w + \tilde{q}_w - n\tilde{q}_w) - \frac{n}{2} \\ &= \frac{n-1}{2} \log \chi_w + \frac{1}{2} \log(\chi_w + nq_w) \end{aligned} \quad (4.36)$$

となる。ただし 3.2.4 小節の議論をそのまま用いた。また $\chi_w = 1 - q_w$, $\Lambda - \tilde{\chi}_w + \tilde{q}_w = \frac{1}{1-q_w}$, $\tilde{q}_w = \frac{q_w}{(1-q_w)(1-q_w+nq_w)}$ である。

4.2.6 既存研究との整合性

ここで提案手法が既存研究の解析も含むことを示す。第 2 章で議論したように、独立同分布に従う場合、つまり Marchenko-Pastur 則で定義されるランダム長方形列の積分公式は $F(x, y) = -\frac{\alpha xy}{2}$ である。これより式 (4.31) から

$$\begin{aligned} \tilde{\chi}_w &= -\alpha \chi_u + n\tilde{q}_w - (n-1)\tilde{q}_w, & \tilde{q}_w &= -\alpha q_u, \\ \tilde{\chi}_u &= \chi_w + nq_w - (n-1)\tilde{q}_u, & \tilde{q}_u &= q_w, \end{aligned} \quad \Rightarrow \quad \begin{cases} \tilde{\chi}_w &= -\alpha \chi_u + \tilde{q}_w \\ \tilde{\chi}_u &= \chi_w + q_w = 1 \end{cases} \quad (4.37)$$

が得られる。これよりキュムラント母関数は

$$\begin{aligned} \psi(n) = & \text{Extr}_{\theta, \tilde{\theta}} \left\{ -\frac{n\tilde{\chi}_w}{2} - \frac{n(n-1)}{2} q_w \tilde{q}_w + \log \int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n \right. \\ & \left. + \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{1-q_w} + z\sqrt{q_w}) \right)^n \right\} \end{aligned} \quad (4.38)$$

となり、式 (3.22) と一致することが見て取れる。つまり本学位論文の提案手法は既存の解析手法も含む形であることが分かる。

4.3 Thouless-Anderson-Palmer 法

上の節で紹介したレプリカ法を用いることでシステムの典型的な性質を調べることができる。ここではある個別の参照例題 $\xi^p = \{X, \vec{y}\}$ が与えられた下で、特に N が十分大きい、有限のときに成り立つもう1つの解析方法を説明する。それは2.2.2小節で紹介した Thouless-Anderson-Palmer 法 (TAP 法) である。またこの方法は、独立同分布の場合の平均場近似法 (54 ページ) にも対応するアルゴリズムを導出することができる。さらに N が十分大きい場合に、大数の法則に似た性質¹として「自己平均性」が保証されるため、レプリカ法の結果と漸近的に一致することが期待される。

導出の手続きは2.2.2小節と同じであるため、Legendre 変換や式 (2.33) などの計算操作の意味づけは省略し、ここでは計算結果だけに言及する。はじめに結合加重 $\vec{w}$ と補助変数 $\vec{u}$ の1次と2次のモーメントをそれぞれ $\langle \vec{w} \rangle = \vec{m}_w$, $\langle \vec{u} \rangle = i\vec{m}_u$, $\langle \vec{w}^T \vec{w} \rangle - \langle \vec{w} \rangle^T \langle \vec{w} \rangle = N\chi_w$, $\langle \vec{u}^T \vec{u} \rangle - \langle \vec{u} \rangle^T \langle \vec{u} \rangle = p\chi_u$ で表し、それらに対応した Legendre 変換の共役な変数を $\vec{h}_w, \vec{h}_u, \tilde{\chi}_w, \tilde{\chi}_u$ で表す。式 (2.33) の関係式により、次の2つの関数を定義する。

$$N\tilde{g}^\beta(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) = \text{Extr}_{\vec{h}_w, \vec{h}_u, \tilde{\chi}_w, \tilde{\chi}_u} \log \int_{-\infty}^{\infty} \frac{d\vec{w} d\vec{u} \tilde{\Delta} d\vec{u}}{(2\pi)^{p+\frac{N}{2}}} \prod_{\mu=1}^p g(y_\mu, \Delta_\mu) \exp \left[i\vec{u}^T (\tilde{\Delta} - \beta X \vec{w}) + A(\vec{w}, \vec{u}) \right] \quad (4.39)$$

$$Ng^\beta(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) = \text{Extr}_{\vec{h}_w, \vec{h}_u, \tilde{\chi}_w, \tilde{\chi}_u} \log \sum_{\vec{w}} P(\vec{w}) \int_{-\infty}^{\infty} \frac{d\vec{u} d\vec{u}}{(2\pi)^p} \prod_{\mu=1}^p g(y_\mu, \Delta_\mu) \exp \left[i\vec{u}^T (\tilde{\Delta} - \beta X \vec{w}) + A(\vec{w}, \vec{u}) \right] \quad (4.40)$$

となる。ただし

$$A(\vec{w}, \vec{u}) := \vec{h}_w^T (\vec{w} - \vec{m}_w) - i\vec{h}_u^T (\vec{u} - i\vec{m}_u) - \frac{\tilde{\chi}_w}{2} (\vec{w}^T \vec{w} - N\chi_w - \vec{m}_w^T \vec{m}_w) - \frac{\tilde{\chi}_u}{2} (\vec{u}^T \vec{u} - p\chi_u + \vec{m}_u^T \vec{m}_u) \quad (4.41)$$

を用いた。ここで2.2.2小節での制御変数 ε に対応するのがここでの β である。2.2.2小節の議論と同様に、 N が十分大きな場合に

$$Ng^1(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) - Ng^0(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) \simeq N\tilde{g}^1(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) - N\tilde{g}^0(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) \quad (4.42)$$

の条件が成り立つことが期待される。以下で各々の関数を評価していこう。

¹細かいが、大数の法則というよりもむしろ、大偏差統計の意味で、ある統計量 (ここでは自由エネルギーとか) の平均値からのズレ δf の示す確率 $P(\delta f)$ が、システムサイズ N に対して、指数関数的に0に漸近する、つまり $P(\delta f) \leq e^{-NE(\delta f)}$, $E(\delta f) \geq 0$ で上から押さえられるため、与えられた参照例題 ξ^p で評価した結果がレプリカ法の結果に一致する。

4.3.1 $N\tilde{g}^\beta(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u)$ の計算

以下では $N\tilde{g}^\beta(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u)$ の引数を省略して $N\tilde{g}^\beta$ で表すことにする。はじめに $\vec{w}$ と $\vec{u}$ はガウス積分なので、先に積分を実行すると、

$$\begin{aligned}
 N\tilde{g}^\beta &= \text{Extr}_{\vec{h}_w, \vec{h}_u, \tilde{\chi}_w, \tilde{\chi}_u} \log \int_{-\infty}^{\infty} \frac{d\vec{w} d\vec{\Delta} d\vec{u}}{(2\pi)^{p+\frac{N}{2}}} \prod_{\mu=1}^p g(y_\mu, \Delta_\mu) \exp \left[-\frac{1}{2} (\vec{w}^T, \vec{u}^T) \begin{pmatrix} \tilde{\chi}_w & i\beta X^T \\ i\beta X & \tilde{\chi}_u \end{pmatrix} \begin{pmatrix} \vec{w} \\ \vec{u} \end{pmatrix} \right. \\
 &\quad \left. + (\vec{w}^T, \vec{u}^T) \begin{pmatrix} \vec{h}_w \\ i\vec{\Delta} - i\vec{h}_u \end{pmatrix} - \vec{h}_w^T \vec{m}_w - \vec{h}_u^T \vec{m}_u + \frac{\tilde{\chi}_w}{2} (N\chi_w + \vec{m}_w^T \vec{m}_w) + \frac{\tilde{\chi}_u}{2} (p\chi_u - \vec{m}_u^T \vec{m}_u) \right] \\
 &= \text{Extr}_{\vec{h}_w, \vec{\theta}_u, \tilde{\chi}_w, \tilde{\chi}_u} \log \int_{-\infty}^{\infty} \frac{d\vec{\Delta}}{(2\pi)^{\frac{p}{2}}} \prod_{\mu=1}^p g(y_\mu, \Delta_\mu) \exp \left[\frac{1}{2} (\vec{h}_w^T, \vec{\theta}_u^T) \begin{pmatrix} \tilde{\chi}_w & i\beta X^T \\ i\beta X & \tilde{\chi}_u \end{pmatrix}^{-1} \begin{pmatrix} \vec{h}_w \\ \vec{\theta}_u \end{pmatrix} - \vec{m}_u^T \vec{\Delta} \right. \\
 &\quad \left. - \frac{1}{2} \log \det \begin{pmatrix} \tilde{\chi}_w & i\beta X^T \\ i\beta X & \tilde{\chi}_u \end{pmatrix} - (\vec{h}_w^T, \vec{\theta}_u^T) \begin{pmatrix} \vec{m}_w \\ i\vec{m}_u \end{pmatrix} + \frac{\tilde{\chi}_w}{2} (N\chi_w + \vec{m}_w^T \vec{m}_w) + \frac{\tilde{\chi}_u}{2} (p\chi_u - \vec{m}_u^T \vec{m}_u) \right] \quad (4.43)
 \end{aligned}$$

となる。ただし共役な変数を $\vec{\theta}_u = i\vec{\Delta} - i\vec{h}_u$ に書き換えた。続いて $\vec{h}_w, \vec{\theta}_u$ に対して極値を評価すると、

$$\text{Extr}_{\vec{h}_w, \vec{\theta}_u} \left\{ \frac{1}{2} (\vec{h}_w^T, \vec{\theta}_u^T) \begin{pmatrix} \tilde{\chi}_w & i\beta X^T \\ i\beta X & \tilde{\chi}_u \end{pmatrix}^{-1} \begin{pmatrix} \vec{h}_w \\ \vec{\theta}_u \end{pmatrix} - (\vec{h}_w^T, \vec{\theta}_u^T) \begin{pmatrix} \vec{m}_w \\ i\vec{m}_u \end{pmatrix} \right\} = \frac{\tilde{\chi}_u}{2} \vec{m}_u^T \vec{m}_u - \frac{\tilde{\chi}_w}{2} \vec{m}_w^T \vec{m}_w + \beta \vec{m}_u^T X \vec{m}_w$$

となり、式 (4.43) は

$$N\tilde{g}^\beta = \log \int_{-\infty}^{\infty} \frac{d\vec{\Delta} g(\vec{y}, \vec{\Delta})}{(2\pi)^{\frac{p}{2}}} e^{-\vec{m}_u^T \vec{\Delta} + \beta \vec{m}_u^T X \vec{m}_w} + \text{Extr}_{\tilde{\chi}_w, \tilde{\chi}_u} \left\{ \frac{N\chi_w \tilde{\chi}_w}{2} + \frac{p\chi_u \tilde{\chi}_u}{2} - \frac{p}{2} \log \tilde{\chi}_u - \frac{1}{2} \log \det \left(\tilde{\chi}_w + \frac{\beta^2}{\tilde{\chi}_u} X^T X \right) \right\}$$

と計算される。幸いなことに $\tilde{\chi}_w, \tilde{\chi}_u$ はスカラーなので $X^T X$ の固有値分布 $\rho(\lambda)$ を用いると N が十分大きな極限で、

$$-\frac{1}{2} \log \det \left(\tilde{\chi}_w + \frac{\beta^2}{\tilde{\chi}_u} X^T X \right) = -\frac{N}{2} \int d\lambda \rho(\lambda) \log \left(\tilde{\chi}_w + \frac{\beta^2 \lambda}{\tilde{\chi}_u} \right) \quad (4.44)$$

となり、結果的に

$$N\tilde{g}^1(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) - N\tilde{g}^0(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) = \vec{m}_u^T X \vec{m}_w + NF(\chi_w, \chi_u) \quad (4.45)$$

が導かれる。ここでは自然にランダム行列の積分公式を使う表式になっていることに注意。

4.3.2 $Ng^\beta(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u)$ の計算

一方 $\beta = 0$ のときの $Ng^0(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u)$ は容易に評価することができる。

$$\begin{aligned}
 Ng^0(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) &= \text{Extr}_{\vec{h}_w, \vec{h}_u, \tilde{\chi}_w, \tilde{\chi}_u} \left\{ -\vec{h}_w^T \vec{m}_w - \vec{h}_u^T \vec{m}_u + \frac{N\tilde{\chi}_w}{2} \left(\chi_w + \frac{\vec{m}_w^T \vec{m}_w}{N} \right) + \frac{p\tilde{\chi}_u}{2} \left(\chi_u - \frac{\vec{m}_u^T \vec{m}_u}{p} \right) \right. \\
 &\quad \left. + \sum_{k=1}^N \log \left(\sum_w P(w) e^{-\frac{\tilde{\chi}_w}{2} w^2 + h_{wk} w} \right) + \sum_{\mu=1}^p \log \left(\int_{-\infty}^{\infty} Dx g(y_\mu, x\sqrt{\tilde{\chi}_u} + h_{u\mu}) \right) \right\} \quad (4.46)
 \end{aligned}$$

これより式 (4.45) と合わせて

$$Ng^1(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) = \vec{m}_u^T X \vec{m}_w + NF(\chi_w, \chi_u) + Ng^0(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) \quad (4.47)$$

が求まり、 $Ng^1(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u)$ を、変数 $\vec{m}_w, \vec{m}_u, \chi_w, \chi_u$ に対して極値を評価することで

$$S = \lim_{N \rightarrow \infty} \frac{1}{N} \left\{ \text{Extr}_{\vec{m}_w, \vec{m}_u, \chi_w, \chi_u} Ng^1(\vec{m}_w, \vec{m}_u, \chi_w, \chi_u) \right\} \quad (4.48)$$

が得られる。

4.3.3 アルゴリズム

73 ページから、TAP 法のアルゴリズムと平均場近似法のアルゴリズムが多くの変数で対応していることが見て取れる。表 4.1 のアルゴリズムの中で平均場近似法 (表 4.2) と違うところは式 (4.55) から式 (4.58) であり、ランダム行列の積分公式の補助変数 Λ_w, Λ_u を用いることから、パターン間の相関の影響を取り込める形式になっていることが分かる。

4.4 まとめ

本章ではパーセプトロン学習の中で代表的な「記憶容量問題」と「教師推定問題」に対する、入力間に相関のある場合に対する既存手法の拡張を説明し、以下のことが解明された。

1. 入力行列 $X \in \mathcal{M}_{p \times N}$ の自己相関行列 $X^T X$ の固有値分布 $\rho(\lambda)$ で定義される、ランダム長方形行列の積分公式 $F(x, y)$ が相関のあるパターンを扱う際の重要な鍵になることを指摘した。
2. レプリカ解析では、エントロピーの評価の中で、入力行列によるアンサンブル平均が、ちょうどランダム長方形行列の積分公式 $F(x, y)$ に対応することを述べ、体系的に少数のマクロ変数間の満たす関係式を導出した。
3. 一方ミクロな変数間で満たす関係式から導出された Thouless-Anderson-Palmer 法は N が十分大きい極限でレプリカ解析の結果と一致することを説明した。

ここでの拡張手法を踏まえて、実際に相関例題に対してのパーセプトロン学習の性能解析ができるかどうか、次の章で確認してみよう。

TAP 法のアルゴリズム

モーメント

共役変数 $\vec{h}_w, \vec{h}_u, \tilde{\chi}_w, \tilde{\chi}_u$ が固定された下で、モーメント $\vec{m}_w, \vec{m}_u, \chi_w, \chi_u$ を求める。

$$m_{wk} = \frac{\sum_w w P(w) e^{-\frac{\tilde{\chi}_w}{2} w^2 + h_{wk} w}}{\sum_w P(w) e^{-\frac{\tilde{\chi}_w}{2} w^2 + h_{wk} w}} \quad (4.49)$$

$$m_{u\mu} = \frac{\int_{-\infty}^{\infty} Dxg'(y_\mu, x\sqrt{\tilde{\chi}_u} + h_{u\mu})}{\int_{-\infty}^{\infty} Dxg(y_\mu, x\sqrt{\tilde{\chi}_u} + h_{u\mu})} \quad (4.50)$$

$$c_{wk} = \frac{\sum_w w^2 P(w) e^{-\frac{\tilde{\chi}_w}{2} w^2 + h_{wk} w}}{\sum_w P(w) e^{-\frac{\tilde{\chi}_w}{2} w^2 + h_{wk} w}} \quad (4.51)$$

$$\delta_{u\mu} = \frac{\int_{-\infty}^{\infty} Dxg''(y_\mu, x\sqrt{\tilde{\chi}_u} + h_{u\mu})}{\int_{-\infty}^{\infty} Dxg(y_\mu, x\sqrt{\tilde{\chi}_u} + h_{u\mu})} \quad (4.52)$$

$$\chi_w = \frac{1}{N} \sum_{k=1}^N (c_{wk} - m_{wk}^2) \quad (4.53)$$

$$\chi_u = \frac{1}{p} \sum_{\mu=1}^p (m_{u\mu}^2 - \delta_{u\mu}) \quad (4.54)$$

共役変数

モーメント $\vec{m}_w, \vec{m}_u, \chi_w, \chi_u$ が固定された下で、共役変数 $\vec{h}_w, \vec{h}_u, \tilde{\chi}_w, \tilde{\chi}_u$ を求める。

$$\Lambda_w = \alpha (\Lambda_u - 1) + 1 \quad (4.55)$$

$$\frac{1}{\Lambda_u} = \int \frac{d\lambda \rho(\lambda)}{\Lambda_w \Lambda_u + \lambda \chi_w \chi_u} \quad (4.56)$$

$$\tilde{\chi}_w = -2 \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} = \frac{1 - \Lambda_w}{\chi_w} \quad (4.57)$$

$$\tilde{\chi}_u = -\frac{2}{\alpha} \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} = \frac{1 - \Lambda_u}{\chi_u} \quad (4.58)$$

$$\vec{h}_w = X^T \vec{m}_u + \tilde{\chi}_w \vec{m}_w \quad (4.59)$$

$$\vec{h}_u = X \vec{m}_w - \tilde{\chi}_u \vec{m}_u \quad (4.60)$$

表 4.1: モーメントと共役変数の両方が収束するまで反復的に計算する。また多くの変数で平均場近似法と対応していることが分かる。

平均場近似アルゴリズム (離散値・連続値の場合)

結合加重 w_k の更新

$$\tilde{\chi}_{wk} = \frac{1}{N} \sum_{\mu=1}^p x_{\mu k}^2 \chi_{u\mu} \quad (4.61)$$

$$h_{wk} = \frac{1}{\sqrt{N}} \sum_{\mu=1}^p x_{\mu k} m_{u\mu} + \tilde{\chi}_{wk} m_{wk} \quad (4.62)$$

$$m_{wk} = \begin{cases} \tanh(h_{wk}) & \text{離散値の場合} \\ \chi_{wk} h_{wk} & \text{連続値の場合} \end{cases} \quad (4.63)$$

$$\chi_{wk} = \begin{cases} 1 - m_{wk}^2 & \text{離散値の場合} \\ c_k - m_{wk}^2 & \text{連続値の場合} \end{cases} \quad (4.64)$$

表 4.2: 平均場近似アルゴリズム

拘束変数 u_μ の更新

$$\tilde{\chi}_{u\mu} = \frac{1}{N} \sum_{k=1}^N x_{\mu k}^2 \chi_{wk} \quad (4.65)$$

$$h_{u\mu} = \frac{1}{\sqrt{N}} \sum_{k=1}^N x_{\mu k} m_{wk} - \tilde{\chi}_{u\mu} m_{u\mu} \quad (4.66)$$

$$m_{u\mu} = \frac{\int_{-\infty}^{\infty} Dxg'(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}{\int_{-\infty}^{\infty} Dxg(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})} \quad (4.67)$$

$$\delta_{u\mu} = \frac{\int_{-\infty}^{\infty} Dxg''(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})}{\int_{-\infty}^{\infty} Dxg(y_\mu, x\sqrt{\tilde{\chi}_{u\mu}} + h_{u\mu})} \quad (4.68)$$

$$\chi_{u\mu} = m_{u\mu}^2 - \delta_{u\mu} \quad (4.69)$$

第5章 結果

第4章ではパーセプトロン学習での既存手法を，入力行列 X の自己相関行列 $X^T X$ の固有値分布 $\rho(\lambda)$ で特徴付けられるランダム長方形行列の積分公式 $F(x, y)$ を用いて拡張できることを見てきた．本章ではその有効性を検証するため，以下の「直交例題」と呼ばれる例題集合に対するパーセプトロン学習の学習能力を評価し，既存研究で扱われている「独立同分布」の例題に比べて，学習効率が上がるのかどうか議論する．本章の構成は以下の通りである．はじめに相関例題の代表例として「直交例題」の定式化を取り上げる．特にこの直交例題は，データ解析で広く用いられている「直交化」と呼ばれる「前処理」に対応するため，実際に独立同分布を仮定した（前処理を行っていない）サンプルセットと，前処理を施したサンプルセットを用いることで，前処理によるクラス識別の性能向上の定量的な評価が期待できる．次に具体的に記憶容量問題（5.2 節）と教師推定問題（5.3 節）に対する性能向上の有無を，レプリア解析と TAP 法を用いて議論する．さらに 5.4 節ではこの学習器の応用例として情報通信に用いた場合に，i.i.d. 例題と直交例題を比べて，通信効率が向上するかどうか議論する．最後にまとめを述べる．

5.1 直交例題

以下では，相関のあるパターンの例として各入力 $\vec{x}_\mu$ に次の条件を課した場合を考察する．

$$\frac{1}{N} \sum_{k=1}^N x_{\mu k} x_{\nu k} = \delta_{\mu\nu} \quad (p < N) \quad (5.1)$$

$$\frac{1}{p} \sum_{\mu=1}^p x_{\mu k} x_{\mu l} = \delta_{kl} \quad (p > N) \quad (5.2)$$

ただし δ_{ab} はクロネッカーのデルタを表す．この条件は $p < N$ に対しては，パターンごとの直交を， $p > N$ に対しては成分ごとの直交を表す． $p > N$ については通信理論における Welch 限界を達成する系列を特徴付ける条件と一致する [119, 191, 192, 196, 209]．このようなパターンセットは Gram-Schmidt の直交化アルゴリズムを用いることにより，実際的な時間で作ることができる．このとき固有値分布 $\rho(\lambda)$ はデルタ関数 $\delta(u)$ を用いて，

$$\rho(\lambda) := \begin{cases} \alpha\delta(\lambda - 1) + (1 - \alpha)\delta(\lambda) & \alpha < 1 \\ \delta(\lambda - \alpha) & \text{otherwise} \end{cases} \quad (5.3)$$

と表される．本学位論文ではこの例題を「直交例題」と呼ぶことにする．以下では既存研究で扱われた独立同分布に従う例題（以下 i.i.d. 例題）と比較することにより，学習器の性能を特徴付ける， α_C や α_{GD} などが入力間の相関によってどのような影響を受けるのかについて調べる．

5.2 結果 1：記憶容量問題

この節では記憶容量問題を取り上げる．

5.2.1 レプリカ解析

記憶容量問題は式 (4.26) に $n \rightarrow 0$ を行うことでレプリカ解析が得られる. Extr を省略して, ϕ は

$$\begin{aligned} \phi = & F(\chi_w, \chi_u) + q_w \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} + q_u \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} + \frac{\alpha}{2} \tilde{\chi}_u (\chi_u + q_u) - \frac{\alpha}{2} \tilde{q}_u q_u \\ & + \alpha \sum_y P(y) \int_{-\infty}^{\infty} Dz \log \left(\int_{-\infty}^{\infty} Dx g(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right) - \frac{1-q_w}{2} \tilde{q}_w - \frac{\tilde{\chi}_w}{2} (\chi_w + q_w - 1) \\ & + \begin{cases} \int_{-\infty}^{\infty} Dz \log \cosh(z\sqrt{\tilde{q}_w}) & \text{離散値の場合} \\ \frac{\Lambda}{2} - \frac{1}{2} \log \Lambda + \frac{\tilde{q}_w}{2\Lambda} - \frac{1}{2} & \text{連続値の場合} \end{cases} \end{aligned} \quad (5.4)$$

となる. また共役変数は

$$\begin{aligned} \tilde{\chi}_w &= \frac{\Lambda_w - 1}{\chi_w} + \tilde{q}_w, & \tilde{q}_w &= 2 \left(q_w \frac{\partial^2 F(\chi_w, \chi_u)}{\partial \chi_w^2} + q_u \frac{\partial^2 F(\chi_w, \chi_u)}{\partial \chi_w \partial \chi_u} \right) \\ \tilde{\chi}_u &= \frac{1 - \Lambda_u}{\chi_u} + \tilde{q}_u, & \tilde{q}_u &= -\frac{2}{\alpha} \left(q_w \frac{\partial^2 F(\chi_w, \chi_u)}{\partial \chi_u \partial \chi_w} + q_u \frac{\partial^2 F(\chi_w, \chi_u)}{\partial \chi_u^2} \right) \end{aligned} \quad (5.5)$$

の関係を満たす¹. そして式 (4.29) と式 (4.30) から

$$\chi_u + q_u = - \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\frac{\int_{-\infty}^{\infty} Dx g''(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})}{\int_{-\infty}^{\infty} Dx g(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})} \right) \quad (5.6)$$

$$q_u = - \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\frac{\int_{-\infty}^{\infty} Dx g'(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})}{\int_{-\infty}^{\infty} Dx g(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})} \right)^2 \quad (5.7)$$

が得られる. ここで $g(y, \omega) = e^{-\beta} + (1 - e^{-\beta})\Theta(y\omega)$ を用いることで,

$$\int_{-\infty}^{\infty} Dx g(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) = \tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u}} \right) \quad (5.8)$$

$$\int_{-\infty}^{\infty} Dx g'(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) = -\frac{1 - e^{-\beta}}{\sqrt{2\pi(\tilde{\chi}_u - \tilde{q}_u)}} \exp \left(-\frac{\tilde{q}_u}{2(\tilde{\chi}_u - \tilde{q}_u)} z^2 \right) \quad (5.9)$$

$$\int_{-\infty}^{\infty} Dx g''(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) = -\sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u}} \frac{1 - e^{-\beta}}{\sqrt{2\pi(\tilde{\chi}_u - \tilde{q}_u)}} z \exp \left(-\frac{\tilde{q}_u}{2(\tilde{\chi}_u - \tilde{q}_u)} z^2 \right) \quad (5.10)$$

となり,

$$\chi_u + q_u = -\frac{(1 - e^{-\beta})^2}{2\pi(\tilde{\chi}_u - \tilde{q}_u)} \frac{\tilde{q}_u}{\tilde{\chi}_u} \int_{-\infty}^{\infty} Dz \left(\frac{e^{-\frac{z^2 \tilde{q}_u}{2(\tilde{\chi}_u - \tilde{q}_u)}}}{\tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u}} \right)} \right)^2 \quad (5.11)$$

$$q_u = -\frac{(1 - e^{-\beta})^2}{2\pi(\tilde{\chi}_u - \tilde{q}_u)} \int_{-\infty}^{\infty} Dz \left(\frac{e^{-\frac{z^2 \tilde{q}_u}{2(\tilde{\chi}_u - \tilde{q}_u)}}}{\tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u}} \right)} \right)^2 \quad (5.12)$$

¹ $\Lambda_w = \alpha(\Lambda_u - 1) + 1$ と $\Lambda_u^{-1} = \int \frac{d\lambda \rho(\lambda)}{\Lambda_w \Lambda_u + \lambda \chi_w \chi_u}$ を満たす.

が得られる². これより

$$\begin{aligned} \phi = & F(\chi_w, \chi_u) + q_w \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} + q_u \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} + \frac{\alpha}{2} \tilde{\chi}_u (\chi_u + q_u) - \frac{\alpha}{2} \tilde{q}_u q_u \\ & + \alpha \int_{-\infty}^{\infty} Dz \log \tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u}} \right) + \begin{cases} -\frac{1-q_w}{2} \tilde{q}_w + \int_{-\infty}^{\infty} Dz \log \cosh \left(z \sqrt{\tilde{q}_w} \right) & \text{離散値の場合} \\ \frac{1}{2} \log(1-q_w) + \frac{q_w}{2\chi_w} & \text{連続値の場合} \end{cases} \end{aligned} \quad (5.14)$$

となる. 以下で離散値と連続値の場合で個別に議論していく.

5.2.2 離散値の場合

$\beta \rightarrow \infty$ のときを考える. 式 (5.14) から

$$q_w = \int_{-\infty}^{\infty} Dz \tanh^2 \left(z \sqrt{\tilde{q}_w} \right) \quad (5.15)$$

が得られる. また式 (5.3) から $\Lambda_w = \alpha(\Lambda_u - 1) + 1$ を用いて

$$\frac{1}{\Lambda_u} = \begin{cases} \frac{\alpha}{\Lambda_w \Lambda_u + \chi_w \chi_u} + \frac{1-\alpha}{\Lambda_w \Lambda_u} & \alpha < 1 \\ \frac{1}{\Lambda_w \Lambda_u + \alpha \chi_w \chi_u} & \text{otherwise} \end{cases} \Rightarrow \Lambda_u = \begin{cases} 1 - \frac{\chi_w \chi_u}{\Lambda_w} & \alpha < 1 \\ \frac{\alpha \chi_w \chi_u}{1 - \Lambda_w} & \text{otherwise} \end{cases} \quad (5.16)$$

が得られる. 式 (5.5), 式 (5.11), 式 (5.12) と式 (5.15) を用いて鞍点問題を解くことによりエントロピーを評価することができる. 図 5.1 は記憶容量問題に対するエントロピー S とパターン比 α の関係を表している [185]. 実線はレプリカ法の結果であり, マーカはそれに対応する TAP 法の数値結果である. TAP 法のデータは, $N = 100$ に対する 10^4 回の実験結果の平均値である. 誤差棒はマーカと比較して十分小さいため省略した. 数値結果より $0 < \alpha < 1$ の範囲でレプリカの結果と TAP 法の結果が高い精度で一致していることが分かる. 記憶容量 α_C はエントロピーが 0 になるパターン比として定義される. 図 5.1 より, 直交例題の記憶容量が $\alpha_C \simeq 0.940$ となり, i.i.d. 例題 ($\alpha_C \simeq 0.833$) に比べて多くのパターンを記憶できることがわかる.

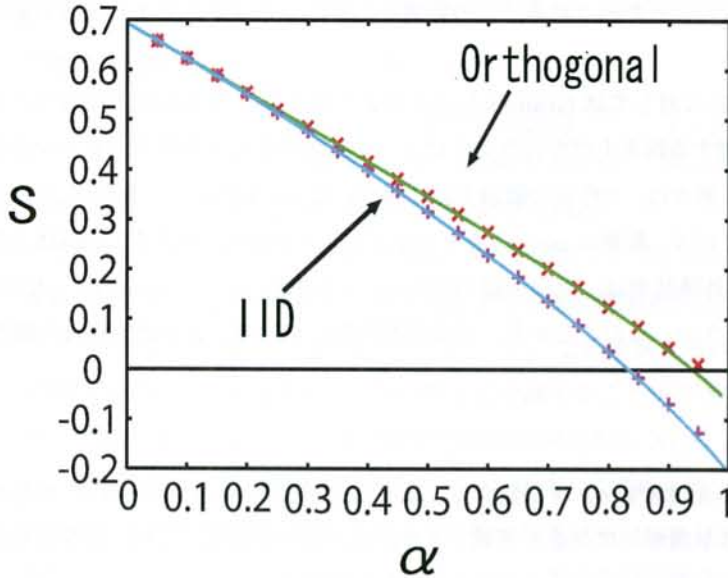


図 5.1: エントロピー S とパターン比 α の数値結果. 下側が i.i.d. 例題のときのエントロピーであり, 上側が直交例題のエントロピーである. 実線がレプリカ法から得られた理論解であり, マーカはそれに対応する TAP 法の結果である. 記憶容量 α_C はエントロピーが 0 に漸近するパターン比として特徴付けられる. i.i.d. 例題 ($\alpha_C \simeq 0.833$) と比較して直交例題 ($\alpha_C \simeq 0.940$) の方が多くのパターンを記憶できることが分かる.

本稿で記したレプリカ解析は「レプリカ対称解 (以下 RS 解)」と呼ばれる最も単純な解に限定している. しかしながら詳細な議論を行うと, $\alpha = \alpha_{AT} \simeq 0.810$ を境に解の熱力学的な安定性が失われ, $\alpha > \alpha_{AT}$ ではより複雑なレ

$$\int_{-\infty}^{\infty} Dyy \tilde{H}^{m+1}(\gamma y) e^{-\frac{y^2 \gamma^2}{2}} = -\frac{\gamma(m+1)}{1+\gamma^2} \frac{1-e^{-\beta}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} Dz \tilde{H}^m(\gamma z) e^{-\gamma^2 z^2} \quad (5.13)$$

数え上げ実験の手順

$t = 1$ とし、以下の数え上げ実験を行う（試行回数 $T = \mathbf{s3}$. 結合加重の集合 Ω_p の中に残っている結合加重の個数 $\#\Omega_p$ に対して、

- i1. 入力パターン $\vec{x}_\mu := \{x_{1\mu}, x_{2\mu}, \dots, x_{N\mu}\}^T \in \mathbf{R}^N$ を生成し、それに対してランダムに出力 y_μ を割り当てる.
- # $\Omega_p > 0$ のとき $p = p + 1$ として **s2.** に戻る.
 # $\Omega_p = 0$ のとき $p_C = p - 1$ として **s4.** に進む.

s4. $\alpha_C(N, t) = p_C/N$ と代入して

- i2. 入力行列 X の自己相関行列 $X^T X$ の固有値分布が $\rho(\lambda)$ を満たすように参照例題 $\xi^\infty := \{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_p, y_p), \dots\}$ を用意する.

$t < T$ のとき $t = t + 1$ として **i1.** に戻る.

$t = T$ のとき 有限サイズ N での記憶容量の推定量 $\alpha_C(N)$ を

$$\alpha_C(N) := \frac{1}{T} \sum_{t=1}^T \alpha_C(N, t) \quad (5.17)$$

- s1.** $p = 1$ とし、 2^N 個の結合加重の集合を Ω_0 とする.

によって定義する.

- s2.** 式 (2.4) で定義されたパーセプトロンを用いて、結合加重の集合 Ω_{p-1} の中で p 番目の入出力関係 $(\vec{x}_p, y_p)$ を満たす結合加重 $\vec{w}$ をすべて集合 Ω_p に代入する.

有限サイズ $N = 2, 3, \dots, 20$ の数え上げ実験を行い、その $\alpha_C(N)$ を用いて熱力学的極限での記憶容量 α_C を予測する.

表 5.1: 記憶容量 α_C を算出するための数え上げ実験の概要. 外挿結果は図 5.1 にて示す.

ブリカ非対称 (RSB) 解を採用しなければならないことが示される. この影響を比較的小きな系に対する直接実験によって調べた.

比較的小きなシステムサイズ $N (= 2, 3, \dots, 20)$ に対しては Gram-Schmidt の直交化アルゴリズムなどに基づいて参照例題 ξ^p を生成し、結合加重 $\vec{w} \in \{\pm 1\}^N$ に関する数え上げを行うことによって記憶できる平均的なパターン数の上限 $\bar{p}_C$ を評価することが可能である (表 5.1). 我々は、 10^6 回の数値実験に基づいて $\bar{p}_C$ を評価し、有限の N に対する臨界値 $\alpha_C(N) := \bar{p}_C/N$ を算出した. 横軸に $1/N$, 縦軸に $\alpha_C(N)$ をとったプロットが図 5.2 である. 曲線はこの結果を $1/N$ の 2 次関数で外挿したものである. 外挿結果は、直交例題の場合 $\alpha_C \simeq 0.938$, i.i.d. の場合 $\alpha_C \simeq 0.819$ を与えており、RS 解による予想値とのずれは小さい. このことから、記憶容量問題におけるレプリカ対称性の破れの影響はそれほど大きくないことが示唆される.

5.2.3 連続値の場合

ここでは直交例題に対する、絶対零度の場合の記憶容量 α_C を議論する. まず固有値分布 $\rho(\lambda)$ は式 (5.3) で与えられており、i.i.d. 例題と同様に α に対して q_w は単調増加になると予測できるため、以下では $q_w \rightarrow 1$ になる α を求める. $\tilde{\chi}_u - \tilde{q}_u \rightarrow 0$ と仮定して、 $w \gg 0$ のとき $\log H(w) \simeq -\frac{w^2}{2} - \frac{1}{2} \log w$ であるので、

$$\alpha \int_{-\infty}^{\infty} Dz \log H \left(z \sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u}} \right) \simeq \alpha \int_0^{\infty} Dz \left(-\frac{z^2}{2} \frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u} - \frac{1}{2} \log \frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u} \right) = -\frac{\alpha}{4} \frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u} - \frac{\alpha}{4} \log \frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u} \quad (5.18)$$

となる. ただし定数は省略した. また式 (5.11) と式 (5.12) から

$$q_u \simeq -\frac{\tilde{q}_u}{2(\tilde{\chi}_u - \tilde{q}_u)^2}, \quad \chi_u \simeq \frac{\tilde{q}_u}{2\tilde{\chi}_u(\tilde{\chi}_u - \tilde{q}_u)} \quad (5.19)$$

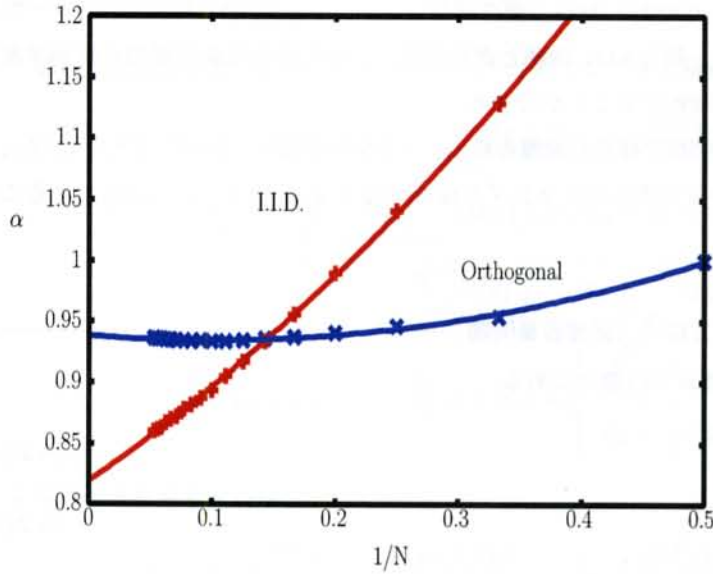


図 5.2: 有限の $N (= 2, 3, \dots, 20)$ に対しての典型的な記憶容量 $\alpha_C(N)$ を概算し (試行回数 10^6), その結果を横軸 $1/N$, 縦軸 $\alpha_C(N)$ でプロットした. 各点での誤差棒はマーカより小さい. 得られた外挿結果は i.i.d. 例題は $\alpha_C(\infty) \simeq 0.819$, 直交例題は $\alpha_C(\infty) \simeq 0.938$ となり, それぞれに対する理論解による予想値に近い (i.i.d. 例題で $\alpha_C \simeq 0.833$, 直交例題で $\alpha_C \simeq 0.940$).'

が得られる. ここで $\gamma := \frac{\bar{\chi}_u}{\bar{q}_u} \simeq 1$ とおくと, $\bar{\chi}_u - \bar{q}_u \simeq \frac{1}{2\chi_u\gamma}$ であり, 式 (5.5) から $\bar{\chi}_u - \bar{q}_u = \frac{1-\Lambda_u}{\chi_u} \simeq \frac{1}{2\chi_u\gamma}$ を用いて, $\Lambda_u \simeq 1 - \frac{1}{2\gamma}$ と $\Lambda_w = 1 + \alpha(\Lambda_u - 1) \simeq 1 - \frac{\alpha}{2\gamma}$ が得られる. $F(\chi_w, \chi_u)$ の極値条件を満たす補助変数は定義から $\Lambda_w, \Lambda_u > 0$ を満たすため,

$$\alpha < 2 = \alpha_C, \quad \gamma \simeq 1 \quad (5.20)$$

となり記憶容量 $\alpha_C = 2$ が得られる. またより一般的に $X \in \mathcal{M}_{p \times N}$ のランクが N_ϱ , ($0 \leq \varrho \leq 1$) の任意のランダム行列に対する, 絶対零度の場合の記憶容量 α_C を議論する. ランクが N_ϱ なので $X^T X$ の固有値分布 $\rho(\lambda)$ は,

$$\rho(\lambda) = (1 - \varrho)\delta(\lambda) + \varrho\tilde{\rho}(\lambda) \quad (5.21)$$

で表せるため, $\Lambda_u^{-1} = \int \frac{d\lambda \rho(\lambda)}{\Lambda_w \Lambda_u + \lambda \chi_w \chi_u}$ から

$$\frac{1}{\Lambda_u} = \frac{1 - \varrho}{\Lambda_w \Lambda_u} + \varrho \int \frac{d\lambda \tilde{\rho}(\lambda)}{\Lambda_w \Lambda_u + \lambda \chi_w \chi_u} \geq \frac{1 - \varrho}{\Lambda_w \Lambda_u}, \quad \Rightarrow \quad \alpha \leq 2\varrho\gamma \simeq 2\varrho \quad (5.22)$$

が得られる. これより記憶容量 $\alpha_C = 2\varrho$ が得られる.

この結果は以下のように解釈できる. $X^T X$ のランク N_ϱ はクラス判別に関わる結合加重 $\vec{w}$ の実効的な次元に対応する. 球状パーセプトロンでは結合加重 $\vec{w}$ を任意の回転行列 O を用いて回転変換しても ($\vec{w}' = O^T \vec{w}$), 球状パーセプトロンの制約を満たす ($\vec{w}^T \vec{w} = \vec{w}'^T \vec{w}' = N$). この性質を用いると, 適当な回転変換を施し結合加重の成分を判別に関わる実効的な部分とそうでない部分に分離することができる. 実効的でない部分に関しては, 参照データから全く情報が得られないので学習後の結合加重は典型的に $N(1 - \varrho)$ の自乗ノルムを持つランダムベクトルと考えてよい. このことは, 考察する問題が自乗ノルムが N_ϱ に制約された N_ϱ 次元の球状パーセプトロンに関するフルランクの例題からの学習問題に還元されることを示している. つまり実効的な次元での変数を $\hat{\alpha}$ とおくと, 元の変数 α に対して $\hat{\alpha} = \frac{\alpha}{N_\varrho} = \frac{\alpha}{\varrho}$ が成り立つ. これより N_ϱ 次元での球状パーセプトロンの臨界値 $\hat{\alpha}_C = 2$ となるため, $\alpha_C = 2\varrho$ と簡単に得られる. 一方, 結合が離散的な場合は, 離散性の制約のためここで述べたような次元の分離ができない. そのため, 固有値分布に依存して異なる値となる.

5.2.4 まとめ

本節では i.i.d. 例題と直交例題, もしくはランクが N_ϱ の任意の入力例題における記憶容量問題について議論し, 以下のことが説明された.

1. 結合加重が離散値の場合, i.i.d. 例題 ($\alpha_C \simeq 0.833$) に比べて直交例題 ($\alpha_C \simeq 0.940$) のほうが多くのパターンを覚えることができることを解析的に示した. 特に i.i.d. 例題と直交例題の結果の差分は前処理の有無が学習効率に与える影響を定量的に評価していると解釈することができる.
2. 結合加重が連続値の場合, i.i.d. 例題と直交例題では同じ記憶容量 $\alpha_C = 2$ が得られた. さらに $X^T X \in \mathcal{M}_{N \times N}$ のランクが N_ϱ である任意の固有値分布から生成されるランダム長方形列 X に対して, $\alpha_C = 2\varrho$ であることも解析的に示された.

鞍点方程式 (記憶容量問題)

記憶容量問題の場合の秩序変数の鞍点方程式は以下の通りである.

$$\tilde{\chi}_w = \frac{\Lambda_w - 1}{\chi_w} + \tilde{q}_w \quad (5.23)$$

$$\tilde{q}_w = 2q_w \frac{\partial^2 F(\chi_w, \chi_u)}{\partial \chi_w^2} + 2q_u \frac{\partial^2 F(\chi_w, \chi_u)}{\partial \chi_w \partial \chi_u} \quad (5.24)$$

$$\tilde{\chi}_u = \frac{1 - \Lambda_u}{\chi_u} + \tilde{q}_u \quad (5.25)$$

$$\tilde{q}_u = -\frac{2q_w}{\alpha} \frac{\partial^2 F(\chi_w, \chi_u)}{\partial \chi_w \partial \chi_u} - \frac{2q_u}{\alpha} \frac{\partial^2 F(\chi_w, \chi_u)}{\partial \chi_u^2} \quad (5.26)$$

$$\Lambda_w = 1 + \alpha(\Lambda_u - 1) \quad (5.27)$$

$$\Lambda_u = \left(\int \frac{d\lambda \rho(\lambda)}{\Lambda_w \Lambda_u + \lambda \chi_w \chi_u} \right)^{-1} \quad (5.28)$$

$$q_w = \begin{cases} \int_{-\infty}^{\infty} Dz \tanh^2 \left(z \sqrt{\tilde{q}_w} \right) & \text{離散値の場合} \\ \frac{\tilde{q}_w}{\Lambda^2}, \quad \Lambda = \frac{1}{1 - q_w} & \text{連続値の場合} \end{cases} \quad (5.29)$$

$$q_u = -\frac{1}{2\pi} \frac{(1 - e^{-\beta})^2}{(1 + e^{-\beta})^2} \frac{1}{\tilde{\chi}_u - \tilde{q}_u} \int_{-\infty}^{\infty} Dy \left(\frac{(1 + e^{-\beta}) e^{-\frac{y^2}{2} \frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u}}}{\tilde{H} \left(y \sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u}} \right)} \right)^2 \quad (5.30)$$

$$\chi_w = 1 - q_w \quad (5.31)$$

$$\chi_u = q_u \left(\frac{\tilde{q}_u}{\tilde{\chi}_u} - 1 \right) \quad (5.32)$$

$$\begin{aligned} \phi = & F(\chi_w, \chi_u) + q_w \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} + q_u \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} + \alpha \int_{-\infty}^{\infty} Dz \log \tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u}} \right) + \frac{\alpha}{2} \tilde{\chi}_u (\chi_u + q_u) \\ & - \frac{\alpha}{2} \tilde{q}_u q_u + \begin{cases} -\frac{\tilde{q}_w}{2} (1 - q_w) + \int_{-\infty}^{\infty} Dz \log \cosh \left(z \sqrt{\tilde{q}_w} \right) & \text{離散値の場合} \\ \frac{1}{2} \log(1 - q_w) + \frac{q_w}{2(1 - q_w)} & \text{連続値の場合} \end{cases} \end{aligned} \quad (5.33)$$

5.3 結果 2 : 教師推定問題

この節では教師推定問題を取り上げる.

5.3.1 レプリカ解析

教師推定問題は式 (4.26) に $n \rightarrow 1$ を行うことでレプリカ解析が得られる。Extr を省略して ϕ は

$$\begin{aligned} \phi = & F(\chi_w, \chi_u) + q_w \frac{\partial F(\chi_w + q_w, \chi_u + q_u)}{\partial \chi_w} + q_u \frac{\partial F(\chi_w + q_w, \chi_u + q_u)}{\partial \chi_u} + \frac{\alpha \tilde{\chi}_u}{2} (\chi_u + q_u) + \frac{\alpha}{2} \tilde{q}_u q_u \\ & + \alpha \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \log \int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})} \\ & - \frac{\tilde{\chi}_w}{2} (\chi_w + q_w - 1) - \frac{\tilde{q}_w}{2} (1 + q_w) + \begin{cases} \int_{-\infty}^{\infty} Dz \log \cosh(z\sqrt{\tilde{q}_w} + \tilde{q}_w) & \text{離散値の場合} \\ \tilde{q}_w - \frac{1}{2} \log(1 + \tilde{q}_w) & \text{連続値の場合} \end{cases} \end{aligned} \quad (5.34)$$

となる。また共役変数は

$$\begin{aligned} \tilde{\chi}_w &= 2 \frac{\partial F(\chi_w + q_w, \chi_u + q_u)}{\partial \chi_w}, & \tilde{q}_w &= 2 \frac{\partial F(\chi_w + q_w, \chi_u + q_u)}{\partial \chi_w} - 2 \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} \\ \tilde{\chi}_u &= -\frac{2}{\alpha} \frac{\partial F(\chi_w + q_w, \chi_u + q_u)}{\partial \chi_u}, & \tilde{q}_u &= -\frac{2}{\alpha} \frac{\partial F(\chi_w + q_w, \chi_u + q_u)}{\partial \chi_u} + \frac{2}{\alpha} \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} \end{aligned} \quad (5.35)$$

を満たす。式 (4.27) と式 (4.28) から

$$\chi_u + q_u = - \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right) \left(\frac{\int_{-\infty}^{\infty} Dxg''(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})}{\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})} \right)}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)} \quad (5.36)$$

$$q_u = - \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right) \left(\frac{\int_{-\infty}^{\infty} Dxg'(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})}{\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u})} \right)^2}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)} \quad (5.37)$$

が得られ、特に $g(y, w) = e^{-\beta} + (1 - e^{-\beta})\Theta(yw)$ のとき、 $\sum_y P(y) \int_{-\infty}^{\infty} Dz Dxg(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) = \frac{1 + e^{-\beta}}{2}$

であるので、式 (5.36) と式 (5.37) から

$$\chi_u + q_u = \frac{2}{1 + e^{-\beta}} \int_{-\infty}^{\infty} Dz Dxz \frac{(1 - e^{-\beta})\sqrt{\tilde{q}_u}}{\sqrt{2\pi(\tilde{\chi}_u - \tilde{q}_u)^3}} e^{-\frac{z^2 \tilde{q}_u}{2(\tilde{\chi}_u - \tilde{q}_u)}} = 0 \quad (5.38)$$

$$q_u = -\frac{(1 - e^{-\beta})^2}{2\pi(\tilde{\chi}_u - \tilde{q}_u)} \frac{2}{(1 + e^{-\beta})} \int_{-\infty}^{\infty} Dz \frac{e^{-\frac{z^2 \tilde{q}_u}{2(\tilde{\chi}_u - \tilde{q}_u)}}}{\tilde{H}\left(z\sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u - \tilde{q}_u}}\right)} = -\frac{(1 - e^{-\beta})^2}{\pi(1 + e^{-\beta})} \frac{1}{\sqrt{\tilde{\chi}_u - \tilde{q}_u}\sqrt{\tilde{\chi}_u}} \int_{-\infty}^{\infty} Dy \frac{e^{-\frac{\tilde{q}_u}{2\tilde{\chi}_u} y^2}}{\tilde{H}\left(y\sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u}}\right)} \quad (5.39)$$

と求まる. 式 (5.38) から $\tilde{\Lambda}_w = \tilde{\Lambda}_u = 1$ であり³, $\frac{\partial F(\chi_w + q_w, \chi_u + q_u)}{\partial \chi_w} = 0$, $\frac{\partial F(\chi_w + q_w, \chi_u + q_u)}{\partial \chi_u} = -\frac{\langle \lambda \rangle}{2}$ が得られる. ただし $\langle A(\lambda) \rangle := \int d\lambda \rho(\lambda) A(\lambda)$ を用いた. これより

$$\tilde{\chi}_w = 0, \quad \tilde{\chi}_u = \frac{\langle \lambda \rangle}{\alpha}, \quad \tilde{q}_w = \frac{1 - \Lambda_w}{\chi_w}, \quad \tilde{q}_u = \frac{\langle \lambda \rangle}{\alpha} - \frac{1 - \Lambda_u}{\chi_u} \quad (5.41)$$

が得られる. よって

$$\begin{aligned} \phi = & F(\chi_w, \chi_u) + \frac{\alpha \chi_u}{2} \left(\frac{\langle \lambda \rangle}{\alpha} - \tilde{q}_u \right) + \frac{2\alpha}{1 + e^{-\beta}} \int_{-\infty}^{\infty} Dz \tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\frac{\langle \lambda \rangle}{\alpha} - \tilde{q}_u}} \right) \log \tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\frac{\langle \lambda \rangle}{\alpha} - \tilde{q}_u}} \right) \\ & - \frac{\tilde{q}_w}{2} (1 + q_w) + \begin{cases} \int_{-\infty}^{\infty} Dz \log \cosh \left(z \sqrt{\tilde{q}_w} + \tilde{q}_w \right) & \text{離散値の場合} \\ \tilde{q}_w - \frac{1}{2} \log(1 + \tilde{q}_w) & \text{連続値の場合} \end{cases} \end{aligned} \quad (5.42)$$

となる. ただし $\chi_w = 1 - q_w$ である.

5.3.2 離散値の場合

式 (5.42) から

$$q_w = \int_{-\infty}^{\infty} Dz \tanh^2 \left(z \sqrt{\tilde{q}_w} + \tilde{q}_w \right) \quad (5.43)$$

が得られる. それらを用いてエントロピー S とパターン比 α の関係を導出した結果が図 5.3 に示される [186]. TAP 法の実験条件は 5.2.2 小節の記憶容量問題と同じである. 図 5.3 より直交例題の GD 容量が $\alpha_{\text{GD}} \simeq 1.101$ となり, i.i.d. 例題の $\alpha_{\text{GD}} \simeq 1.245$ と比較して小さい. つまり少ない参照例題によって教師ベクトル $\vec{w}_0$ を同定可能であることが確認できる.

詳細な解析を行うと RS 解は, 直交例題の場合 $\alpha < \alpha_{\text{AT}} \simeq 1.305$ で, i.i.d. 例題では $\alpha < \alpha_{\text{AT}} \simeq 1.492$ の範囲で安定であることがわかる. このことは, $S = 0$ で特徴付けられる α_{GD} の評価において, RS 解の予想値が正しいことを示唆している. 臨界点に近い比較的大きな α に対して, TAP 法の結果がレプリカ法の結果の隔たりがやや目立つのは, 有限サイズ効果の影響であると考えられる. これは臨界現象で見られる特有の現象である. 最後に図 5.2 と同様に有限のサイズ $N (= 6, 7, \dots, 19)$ での GD 容量 $\alpha_{\text{GD}}(N)$ を統計的に求め (試行回数 10^6), $1/N$ でプロットした結果が図 5.4 である. 外挿関数の決定に際しては交差検定法にしたがい, 2 次関数から 10 次関数までの多項式関数の中で 1 つ抜き誤差 (Leave one out error) の総和がもっとも小さな 4 次多項式で外挿を行った (図 5.3). この外挿結果 (i.i.d. 例題は $\alpha_{\text{GD}}(\infty) \simeq 1.245$, 直交例題は $\alpha_{\text{GD}}(\infty) \simeq 1.111$) はレプリカ法や TAP 法による理論予想 (i.i.d. 例題は $\alpha_{\text{GD}}(\infty) \simeq 1.245$, 直交例題は $\alpha_{\text{GD}}(\infty) \simeq 1.101$) と近く, 我々が導出した 2 つの理論を支持していることが伺える.

5.3.3 連続値の場合

式 (5.41) と式 (5.42) から

$$q_w = 1 - \frac{1}{1 + \tilde{q}_w}, \quad \tilde{q}_w = \frac{1 - \Lambda_w}{\chi_w}, \quad \chi_w = 1 - q_w \quad \implies \quad \chi_w = \Lambda_w = 1 - q_w \quad (5.45)$$

³

$$\chi_u = \frac{(1 - e^{-\beta})^2}{\pi(1 + e^{-\beta})} \frac{1}{\sqrt{\chi_u - \tilde{q}_u} \sqrt{\chi_u}} \int_{-\infty}^{\infty} Dy \frac{e^{-\frac{\tilde{q}_u}{2\chi_u} y^2}}{\tilde{H} \left(y \sqrt{\frac{\tilde{q}_u}{\chi_u}} \right)} \quad (5.40)$$

数え上げ実験の手順

$t = 1$ とし, 教師ベクトル $\vec{w}_0$ を適当に選び, 以下の数 **s3**. 結合加重の集合 Ω_p の中に残っている結合加重の数え上げ実験を行う (試行回数 $T = 10^6$).

個数 $\#\Omega_p$ に対して,

i1. 入力パターン $\vec{x}_\mu := \{x_{1\mu}, x_{2\mu}, \dots, x_{N\mu}\}^T \in \mathbf{R}^N$ を生成し, それに対して教師ベクトル $\vec{w}_0$ により出力 y_μ を割り当てる.

$\#\Omega_p > 0$ のとき $p = p + 1$ として **s2**. に戻る.
 $\#\Omega_p = 0$ のとき $p_{GD} = p - 1$ として **s4**. に進む.

s4. $\alpha_{GD}(N, t) = p_{GD}/N$ と代入して

i2. 入力行列 X の自己相関行列 $X^T X$ の固有値分布が $\rho(\lambda)$ を満たすように参照例題 $\xi^\infty := \{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_p, y_p), \dots\}$ を用意する.

$t < T$ のとき $t = t + 1$ として **i1**. に戻る.

$t = T$ のとき 有限サイズ N での GD 容量の推定量 $\alpha_{GD}(N)$ を

十分に生成された参照例題セット ξ^∞ に対して, 次の操作によって入出力関係を満たす結合加重を探索する.

$$\alpha_{GD}(N) := \frac{1}{T} \sum_{t=1}^T \alpha_{GD}(N, t) \quad (5.44)$$

s1. $p = 1$ とし, 2^N 個の結合加重の集合を Ω_0 とする.

によって定義する.

s2. 式 (2.4) で定義されたパーセプトロンを用いて, 結合加重の集合 Ω_{p-1} の中で p 番目の入出力関係 $(\vec{x}_p, y_p)$ を満たす結合加重 $\vec{w}$ をすべて集合 Ω_p に代入する.

有限サイズ $N = 6, 7, \dots, 19$ の数え上げ実験を行い, その $\alpha_{GD}(N)$ を用いて熱力学的極限での GD 容量 α_{GD} を予測する.

表 5.2: GD 容量 α_{GD} を算出するための数え上げ実験の概要. 外挿結果は図 5.4 にて示す.

が得られる. さらに式 (5.38), 式 (5.40) と式 (5.41) から

$$\frac{\tilde{q}_u}{\tilde{\chi}_u} = 1 - \frac{\alpha(1 - \Lambda_u)}{\chi_u \langle \lambda \rangle}, \quad \chi_u = \frac{(1 - e^{-\beta})^2}{\pi(1 + e^{-\beta})} \frac{\alpha}{\langle \lambda \rangle} \sqrt{\frac{\chi_u \langle \lambda \rangle}{\alpha(1 - \Lambda_u)}} \int_{-\infty}^{\infty} Dy \frac{e^{-\frac{y^2}{2}(1 - \frac{\alpha(1 - \Lambda_u)}{\chi_u \langle \lambda \rangle})}}{\tilde{H}\left(y \sqrt{1 - \frac{\alpha(1 - \Lambda_u)}{\chi_u \langle \lambda \rangle}}\right)} \quad (5.46)$$

となる. 以下では $\alpha < 1$ と $\alpha > 1$ の場合で各々議論していく.

$\alpha < 1$ の場合 式 (5.16) より $\Lambda_u = 1 - \chi_u$ が得られ, 式 (5.3) より $\alpha = \langle \lambda \rangle$ なので, 式 (5.46) に代入することで,

$$\chi_u = \frac{2(1 - e^{-\beta})^2}{\pi(1 + e^{-\beta})^2}, \quad \chi_w = 1 - \frac{2(1 - e^{-\beta})^2}{\pi(1 + e^{-\beta})^2} \alpha, \quad q_w = \frac{2(1 - e^{-\beta})^2}{\pi(1 + e^{-\beta})^2} \alpha \quad (5.47)$$

が得られ, q_w が α に比例していることが分かる. これより

$$\phi = \alpha \log \left(\frac{2}{1 + e^{-\beta}} \right) \quad (5.48)$$

が得られる⁴.

$\alpha > 1$ の場合 ここでは $\alpha \gg 1$ での $\varepsilon = 1 - q_w = \chi_w$ の漸近形を議論する. そこで式 (3.51) の関数形に倣い,

$$\varepsilon \simeq \frac{\gamma}{\alpha^s}, \quad (s, \gamma > 0) \quad (5.50)$$

⁴秩序変数は

$$\chi_u = \frac{2(1 - e^{-\beta})^2}{\pi(1 + e^{-\beta})^2}, \quad \Lambda_u = 1 - \chi_u, \quad \tilde{q}_u = 0, \quad \chi_w = 1 - \alpha\chi_u, \quad q_w = \alpha\chi_u, \quad \Lambda_w = 1 - \alpha\chi_u, \quad \tilde{q}_w = \frac{\alpha\chi_u}{1 - \alpha\chi_u}, \quad F(\chi_w, \chi_u) = -\alpha\chi_u - \frac{1}{2} \log(1 - \alpha\chi_u) \quad (5.49)$$

であり, それらを式 (5.42) に代入すると式 (5.48) が得られる.

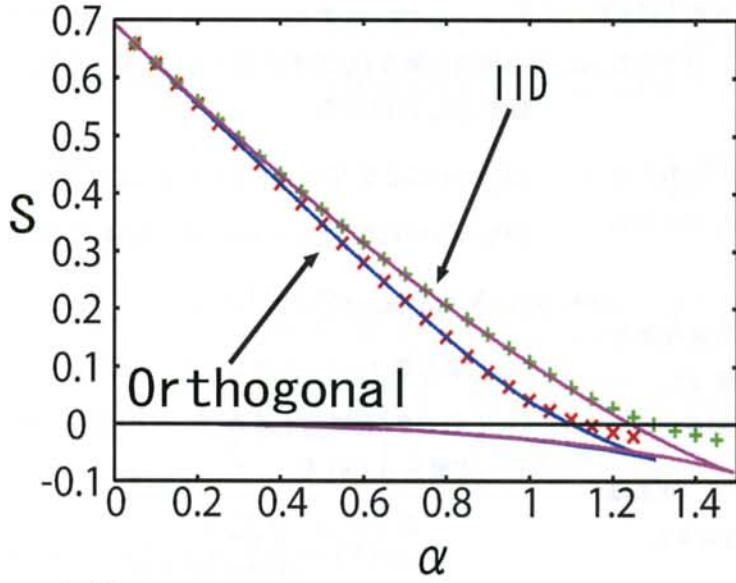


図 5.3: エントロピー S とパターン比 α の数値結果. 上側が i.i.d. 例題のときのエントロピーであり, 下側が直交例題のエントロピーである. 実線がレプリカ解析の結果であり, マーカはそれに対応する TAP 法の結果である. エントロピーが 0 に漸近するパターン比として GD 容量を特徴付けることができる. i.i.d. 例題 ($\alpha_{GD} \simeq 1.245$) に比べて直交例題 ($\alpha_{GD} \simeq 1.101$) のほうが教師を同定しやすいことが分かる.

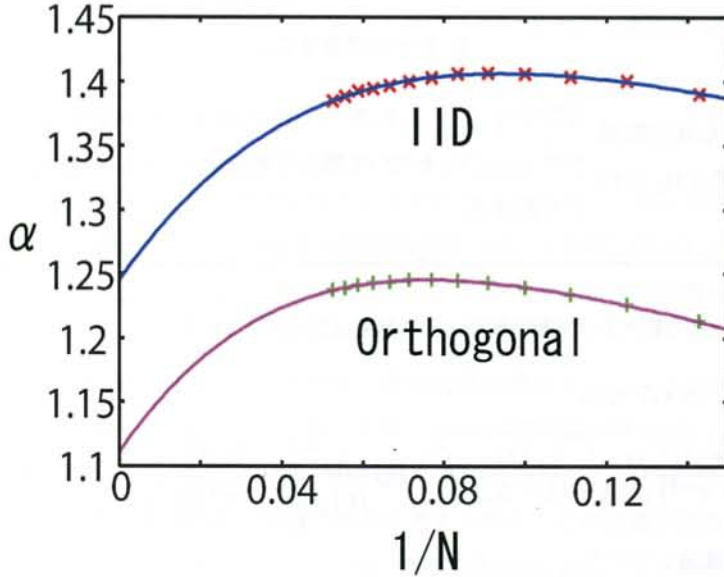


図 5.4: 有限の $N (= 6, 7, \dots, 19)$ に対して GD 容量 $\alpha_{GD}(N)$ を概算し (試行回数 10^6), その結果を横軸 $1/N$, 縦軸 $\alpha_{GD}(N)$ でプロットした. 各点での誤差棒はマーカより小さい. 得られた外挿結果は i.i.d. 例題は $\alpha_{GD}(\infty) \simeq 1.245$, 直交例題は $\alpha_{GD}(\infty) \simeq 1.111$ となり, それぞれに対する理論解による予想値に近い (i.i.d. 例題で $\alpha_{GD} \simeq 1.245$, 直交例題で $\alpha_{GD} \simeq 1.101$). ちなみに表 5.3 から 4 次関数で外挿を行った.

と仮定する. $\Lambda_u = 1 + \frac{\Lambda_w - 1}{\alpha}$ から $\Lambda_u \simeq 1 - \frac{1}{\alpha}$ であり $\tilde{q}_u = 1 - \frac{1 - \Lambda_u}{\chi_u} = 1 - \frac{\chi_w}{\Lambda_u} \simeq 1 - \frac{\gamma}{\alpha^s}$ の漸近形となる. また χ_u は

$$\chi_u = \frac{(1 - e^{-\beta})^2}{\pi(1 + e^{-\beta})} \frac{1}{\sqrt{1 - \tilde{q}_u}} \int_{-\infty}^{\infty} Dy \frac{e^{-\frac{\tilde{q}_u}{2} y^2}}{\tilde{H}(y\sqrt{\tilde{q}_u})} \simeq \frac{(1 - e^{-\beta})^2}{(1 + e^{-\beta})^2} \frac{\alpha^{\frac{s}{2}}}{\pi\sqrt{\gamma}} c(\beta), \quad c(\beta) := \int_{-\infty}^{\infty} Dy \frac{(1 + e^{-\beta})e^{-\frac{y^2}{2}}}{\tilde{H}(y)} \quad (5.51)$$

となり $\tilde{q}_w = \frac{\alpha\chi_u}{\Lambda_u} \simeq \alpha\chi_u$ であり $\tilde{q}_w = \frac{1}{1 - q_w} - 1 \simeq \frac{\alpha^s}{\gamma}$ なので, $\varepsilon = \chi_w$ の漸近形のパラメータは

$$\gamma = \frac{\pi^2}{c^2(\beta)} \left(\frac{1 + e^{-\beta}}{1 - e^{-\beta}} \right)^4, \quad s = 2 \quad (5.52)$$

となる (図 5.5). 特に絶対零度では $\gamma \simeq 1.926$ なので式 (3.51) と一致することがわかる.

より一般的に $X \in \mathcal{M}_{p \times N}$ のランクが $N\varrho$, ($0 \leq \varrho \leq 1$) の任意のランダム行列アンサンブルに対する, $\alpha \gg 1$ のときの $\varepsilon = 1 - q_w$ の漸近形を考える. ランクが $N\varrho$ なので $X^T X \in \mathcal{M}_{N \times N}$ の固有値分布 $\rho(\lambda)$ は

$$\rho(\lambda) = (1 - \varrho)\delta(\lambda) + \varrho\tilde{\rho}(\lambda) \quad (5.53)$$

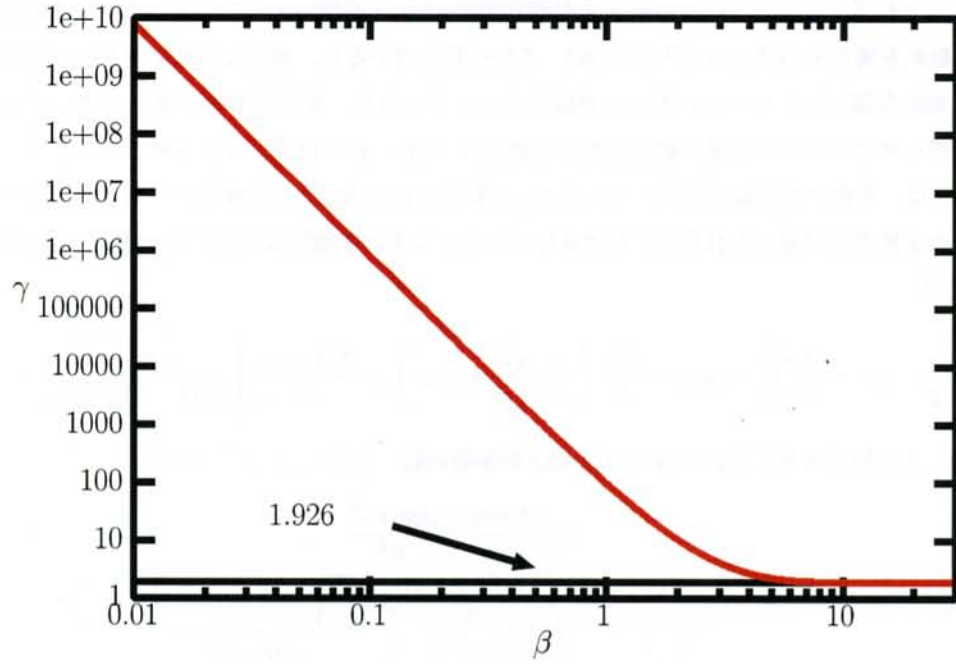


図 5.5: 逆温度 β と γ の図. $\beta \rightarrow \infty$ で $\gamma \rightarrow 1.926$ になる.

で表すことにする. ここで $\varepsilon (= 1 - q_w)$ の関数形を

$$\varepsilon \simeq \delta + \frac{\gamma}{\alpha^s} \quad (5.54)$$

で仮定する. さらに $\Lambda_w = 1 + \alpha(\Lambda_u - 1)$ を用いると, $\Lambda_u = 1 + \frac{\Lambda_w - 1}{\alpha} \simeq 1 - \frac{1 - \delta}{\alpha}$ が得られ,

$$\frac{1}{\Lambda_u} = \int \frac{d\lambda \rho(\lambda)}{\Lambda_w \Lambda_u + \lambda \chi_u \chi_u} \Rightarrow \chi_w = \int \frac{d\lambda \rho(\lambda)}{1 + \frac{\lambda \chi_u}{\Lambda_u}} = 1 - \varrho + \varrho \int \frac{d\lambda \tilde{\rho}(\lambda)}{1 + \frac{\lambda \chi_u}{\Lambda_u}} \geq 1 - \varrho + \frac{\varrho}{1 + \frac{\chi_u}{\Lambda_u} \langle \tilde{\lambda} \rangle} \quad (5.55)$$

ただし $\langle \tilde{\lambda} \rangle := \int d\lambda \tilde{\rho}(\lambda) \lambda$, $\langle \lambda \rangle = \int d\lambda \rho(\lambda) \lambda = \varrho \langle \tilde{\lambda} \rangle$ とおき, 最後の不等式は凸関数 (convex function) の性質を用いた. 式 (5.46) から

$$\frac{\tilde{q}_u}{\tilde{\chi}_u} = 1 - \frac{\alpha(1 - \Lambda_u)}{\chi_u \langle \lambda \rangle} \simeq 1 - \frac{1 - \delta}{\chi_u \langle \lambda \rangle} \quad (5.56)$$

が得られる. これより χ_u の漸近形は

$$\chi_u \simeq \frac{(1 - e^{-\beta})^2}{\pi(1 + e^{-\beta})^2} \frac{\alpha}{\langle \lambda \rangle} \sqrt{\frac{\chi_u \langle \lambda \rangle}{1 - \delta}} c(\beta) \Rightarrow \chi_u \langle \lambda \rangle \simeq \frac{\alpha^2 c^2(\beta)}{\pi^2 (1 - \delta)} \times \left(\frac{1 - e^{-\beta}}{1 + e^{-\beta}} \right)^4 \quad (5.57)$$

となる. また

$$1 + \frac{\chi_u}{\Lambda_u} \langle \tilde{\lambda} \rangle = 1 + \frac{\chi_u \langle \lambda \rangle}{\Lambda_u \varrho} \simeq \frac{\alpha^2 c^2(\beta)}{\pi^2 (1 - \delta) \varrho} \left(\frac{1 - e^{-\beta}}{1 + e^{-\beta}} \right)^4 \quad (5.58)$$

を用いれば, $\alpha \gg 1$ の極限で ε の各パラメータ δ, s, γ は

$$s = 2, \quad \delta = 1 - \varrho, \quad \gamma = \frac{\pi^2 \varrho^3}{c^2(\beta)} \left(\frac{1 + e^{-\beta}}{1 - e^{-\beta}} \right)^4 \quad (5.59)$$

となり, $\beta \rightarrow \infty$ で $\gamma \simeq 1.926 \varrho^3$ となる.

この結果は以下のように解釈できる。 $X^T X$ のランク N_ϱ はクラス判別に関わる結合加重 $\vec{w}$ の実効的な次元に対応する。球状パーセプトロンでは結合加重 $\vec{w}$ を任意の回転行列 O を用いて回転変換しても ($\vec{w}' = O^T \vec{w}$)、球状パーセプトロンの制約を満たす ($\vec{w}^T \vec{w} = \vec{w}'^T \vec{w}' = N$)。この性質を用いると、適当な回転変換を施し結合加重の成分を判別に関わる実効的な部分とそうでない部分に分離することができる。実効的でない部分に関しては、参照データから全く情報が得られないので学習後の結合加重は典型的に $N(1 - \varrho)$ の自乗ノルムを持つランダムベクトルと考えてよい。このことは、考察する問題が自乗ノルムが N_ϱ に制約された N_ϱ 次元の球状パーセプトロンに関するフルランクの例題からの学習問題に還元されることを意味している。つまり実効的な次元での変数を $\hat{\alpha}, \hat{q}_w$ とおくと、元の変数 α, q_w に対して

$$\hat{\alpha} = \frac{p}{N_\varrho} = \frac{\alpha}{\varrho}, \quad q_w = \frac{1}{N} \sum_{k=1}^N w_k w_{k0} = \frac{N_\varrho}{N} \left(\frac{1}{N_\varrho} \sum_{k=1}^{N_\varrho} w'_k w_{k0} \right) + \frac{N(1-\varrho)}{N} \left(\frac{1}{N(1-\varrho)} \sum_{k=N_\varrho+1}^N w'_k w_{k0} \right) = \varrho \hat{q}_w \quad (5.60)$$

が成り立つ。また絶対零度において $\alpha \gg 1$ での $\hat{\varepsilon}$ の漸近形は

$$\hat{\varepsilon} \simeq \frac{1.926}{\hat{\alpha}^2} = \frac{1.926 \varrho^2}{\alpha^2} \quad (5.61)$$

となるため、

$$\varepsilon = 1 - q_w = 1 - \varrho(1 - \hat{\varepsilon}) \simeq 1 - \varrho + \frac{1.926 \varrho^3}{\alpha^2} \quad (5.62)$$

と簡単に得られる。一方、結合が離散的な場合は、離散性の制約のためここで述べたような次元の分離ができない。そのため、固有値分布に依存して異なる値となる。

5.3.4 まとめ

本節では i.i.d. 例題と直交例題の教師推定問題について議論した。以下のことが解明された。

1. 結合加重が離散値の場合の教師推定問題に対して、i.i.d. 例題 ($\alpha_{GD} \simeq 1.245$) に比べて直交例題 ($\alpha_{GD} \simeq 1.101$) のほうが少ないパターンで教師ベクトルを同定できることを解析的に示した。特に i.i.d. 例題と直交例題の結果の差分は前処理の有無が学習効率に与える影響を定量的に評価していると解釈することができる。
2. 結合加重が連続値の場合の教師推定問題に対して、i.i.d. 例題と直交例題では α が大きい極限で、同じく $q_w = 1$ からのズレの漸近形 $\varepsilon = 1 - q_w \simeq \frac{1.926}{\alpha^2}$ が得られた。さらに $X^T X \in \mathcal{M}_{N \times N}$ のランクが N_ϱ である任意の固有値分布から生成されるランダム長方形列 X に対して、 $\varepsilon \simeq 1 - \varrho + \frac{1.926 \varrho^3}{\alpha^2}$ であることも解析的に示された。

5.4 結果3：通信方式への応用

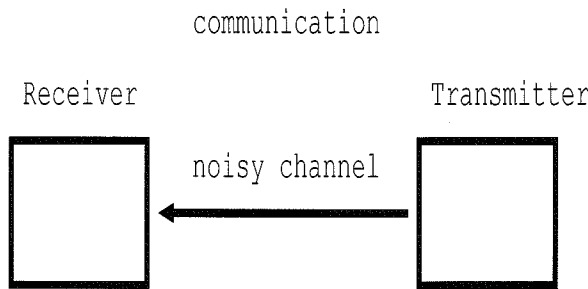


図 5.6: 通信モデル

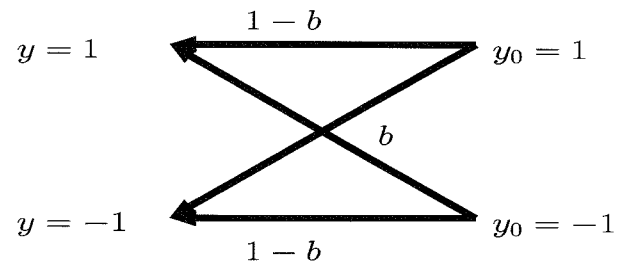


図 5.7: 反転確率 b の 2 元対称通信路 (BSC) の劣化過程の模式図

最後に本研究の手法の応用として、通信方式である「誤り訂正符号」に用いた例を紹介する。以下では、

鞍点方程式（教師推定問題）

教師推定問題の場合の秩序変数の鞍点方程式は以下の通りである。

$$\tilde{\chi}_w = 0 \quad (5.63)$$

$$\tilde{q}_w = \frac{1 - \Lambda_w}{\chi_w} \quad (5.64)$$

$$\tilde{\chi}_u = \frac{\langle \lambda \rangle}{\alpha} \quad (5.65)$$

$$\tilde{q}_u = \frac{\langle \lambda \rangle}{\alpha} - \frac{1 - \Lambda_u}{\chi_u} \quad (5.66)$$

$$\Lambda_w = 1 + \alpha(\Lambda_u - 1) \quad (5.67)$$

$$\Lambda_u = \left(\int \frac{d\lambda \rho(\lambda)}{\Lambda_w \Lambda_u + \lambda \chi_w \chi_u} \right)^{-1} \quad (5.68)$$

$$q_w = \begin{cases} \int_{-\infty}^{\infty} Dz \tanh^2 \left(z \sqrt{\tilde{q}_w} + \tilde{q}_w \right) & \text{離散値の場合} \\ 1 - \frac{1}{1 - \tilde{q}_w} & \text{連続値の場合} \end{cases} \quad (5.69)$$

$$\chi_u = \frac{1(1 - e^{-\beta})^2}{\pi(1 + e^{-\beta})^2} \frac{1}{\tilde{\chi}_u \sqrt{1 - \frac{\tilde{q}_u}{\tilde{\chi}_u}}} \int_{-\infty}^{\infty} Dy \frac{(1 + e^{-\beta}) e^{-\frac{y^2}{2} \frac{\tilde{q}_u}{\tilde{\chi}_u}}}{\tilde{H} \left(y \sqrt{\frac{\tilde{q}_u}{\tilde{\chi}_u}} \right)} \quad (5.70)$$

$$\chi_w = 1 - q_w \quad (5.71)$$

$$q_u = -\chi_u \quad (5.72)$$

$$\begin{aligned} \phi = & F(\chi_w, \chi_u) + \frac{\alpha \chi_u}{2} \left(\frac{\langle \lambda \rangle}{\alpha} - \tilde{q}_u \right) + \frac{2\alpha}{1 + e^{-\beta}} \int_{-\infty}^{\infty} Dz \tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\frac{\langle \lambda \rangle}{\alpha} - \tilde{q}_u}} \right) \log \tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\frac{\langle \lambda \rangle}{\alpha} - \tilde{q}_u}} \right) \\ & - \frac{\tilde{q}_w}{2} (1 + q_w) + \begin{cases} \int_{-\infty}^{\infty} Dz \log \cosh \left(z \sqrt{\tilde{q}_w} + \tilde{q}_w \right) & \text{離散値の場合} \\ \tilde{q}_w - \frac{1}{2} \log(1 + \tilde{q}_w) & \text{連続値の場合} \end{cases} \end{aligned} \quad (5.73)$$

- なぜ誤り訂正符号という技術が必要なのか（誤り訂正符号の一般論）
- 通信路モデルの定式化（通信路のモデル化）
- Shannon の第2定理（通信路符号化定理）

ど、基礎的な部分も含めて説明を行う。より詳しい記述は [40, 158] を参照されたい。

5.4.1 誤り訂正符号の一般論

誤り訂正符号とは、符号理論の中で「信頼性向上のための符号化」⁵という位置付けがなされている符号化のことを言う。この「信頼性向上」とは多くの場合、誤りの発生する確率のある通信路を通して情報伝達を行う「通信問題（情報伝送問題）」を想定していることが多い（図 5.6）⁶。誤り訂正符号の説明をはじめる前に、ここで言う「信頼性向上」技術と聞いて、まずどういうものを思い浮かぶかを考えてみよう。情報伝達で誤りが発生してはいけなないのであれば、「誤りを引き起こしている ノイズ（雑音）の発生源（不純物など）をできるだけ取り除いたデバイスを開発し、それを用いて情報伝達を行えばよい」と言う発想がはじめに思いつくだろう。しかしながら、純度の高い様な物質⁷を作ることは技術的に難しく、製作費用（コスト）もかかるため、高純度の物質を用いるという方策は、技

⁵他には「効率向上のための符号化（圧縮）」、「セキュリティ向上のための符号化（暗号）」などがある。

⁶またはハードディスクに書き込む場合（歪みなし、歪みありデータ圧縮）にも使われているようである。

⁷熱雑音であれば、低温にすることによって「熱振動」による雑音を取り除くことができると考えることができる。しかし費用がかかってしまう。

外挿関数の 次数	1つ抜き誤差の平均 (i.i.d. 例題)	外挿値 (i.i.d. 例題)	1つ抜き誤差の平均 (直交例題)	外挿値 (直交例題)
0	0.004101	1.392704	0.003438	1.234770
1	0.004076	1.414710	0.001321	1.268553
2	0.000256	1.335408	0.000137	1.205526
3	0.000262	1.295237	0.000048	1.158679
4	0.000103	1.245047	0.000001	1.111172
5	0.000145	1.282783	0.000158	1.118968
6	0.007534	1.251661	0.000290	1.275169
7	4.164808	1.300419	0.000053	1.000935
8	2.414898	-6.608969	11.556052	1.329810
9	69.426278	-2.924355	3.854332	0.552925

表 5.3: 交差検定法 (cross validation method) の1つである, 1つ抜き誤差 (Leave one out error) を用いて, 有限サイズでの観測データを使って $N \rightarrow \infty$ の値を予測するための外挿関数 (多項式) の選択を行った. i.i.d. 例題と直交例題の両方で, 4 次関数で外挿することが1つ抜き予測誤差が少ないという結論が得られ, 4 次関数を用いて $N \rightarrow \infty$ を予測すると, レプリカ解析で得られた解析結果と整合性の取れていることが確認できる.

術的・費用的に現実的ではない.

そこで別の方法を考える必要がある. たとえばコストをかけて「ノイズののりにくい物質を開発する」よりも, 決まりごとをつくり, 誤りの箇所を見つけ, それを訂正し, 受信側で送信信号を推定できる「仕掛け」を作るほうが, コスト的にも安く済むのではないかと考えを変えてみよう. すなわち誤りが発生したとしても, それを訂正できる「仕組み」を開発できれば低コストで誤りのない通信が行えると, 発想を転換するのである. 誤り訂正符号とはこのような発想の元で考え出された技術である.

ここで誤り訂正符号の簡単な例を1つ紹介する. 情報源のアルファベットを $\{0, 1\}$ の2値 (2元, Binary とも言う) とし, 反転確率⁸ b で, 送信信号が反転してしまうノイズ通信路を使って通信 (情報伝送) を行うことを考える (図 5.6). たとえば送信者が「1」という情報 1bit 分を送ってみたとする. このとき受信側では, 送信者が送った「1」を正しく受け取ることができる確率が $1-b$ となる. そして誤った情報「0」を受け取る確率が b となる. では次に「1」を3回続けて送って, その受信信号の多数決で送信信号を推定してみることを考える. このとき受信側で「1」と判断できる bit 列は, 「111」, 「011」, 「101」, 「110」の場合であり, 逆に誤って判断してしまう bit 列は, 「100」, 「010」, 「001」, 「000」の場合である. そしてこの場合の「誤る確率」は $O(b^2)$ 程度となる. 明らかに 1bit 分を送る場合よりも誤ってしまう確率は小さくなっていることがわかる. 一般に $2n+1$ 回続けて送って, 単純な多数決の「仕掛け」を使って誤りを訂正する場合, 誤ってしまう確率は $O(b^{n+1})$ となる. このように, 送信する符号語の長さを大きくすればするほど, 誤り訂正能力は向上していくことがわかる. この例は一番単純な例だが, 誤り訂正符号とは, 「多数決のような”仕掛け”をうまく使って誤りを訂正し, 受信側で送信信号を推定を可能にする技術のことである.

5.4.2 雑音あり通信路のモデル化

まず通信問題を扱う場合, 送信信号が外的要因などで確率的に「信号反転する」という, 現実の通信路の統計的性質を抽出して, 数学的に扱いやすく「通信路のモデル化」を行うことが, 通信問題の理論研究の第一歩である. ここでは代表的な通信路モデルである, 2元対称無記憶通信路 (binary symmetric channel ; BSC) における「劣化

⁸ここでいう「反転確率」とは, 送信信号が 0 から 1 へ, または 1 から 0 へ受信信号が変わる確率のことである.

過程の特徴付け」について説明する．次に雑音あり通信路に対して，もっとも効率的に情報伝送を可能にする，送信信号の「割合」について議論にする．またこれ以降で議論の本質は変えないが，パーセプトロン学習の手法を用いるため，送信信号と受信信号の2元アルファベットを ± 1 と表すことにする．

さてBSCとは，通報(送信信号) $y_0 \in \{\pm 1\}$ によらず，反転確率 b で $1 \rightarrow -1$ もしくは $-1 \rightarrow 1$ に反転する通信路のことを指す(図5.7)．つまり送信信号が y_0 の時，通信路を通して受け取る符号語(受信信号) y を受け取る確率が条件付き確率の形で，

$$p(y|y_0) = \begin{cases} b & yy_0 < 0 \\ 1-b & \text{otherwise} \end{cases} \quad (5.74)$$

で表すことができ，これを階段関数 $\Theta(u)$ を用いてまとめると

$$p(y|y_0) = b + (1-2b)\Theta(yy_0) \quad (5.75)$$

となる．この通信路は，その表現(式(5.75))の単純さゆえに，理論の面でも実用の面でも多くの研究が行われてきた通信路モデルである．また定義・導出は後述するが，この通信路の通信容量 C は

$$C = 1 + b \log_2 b + (1-b) \log_2 (1-b) \quad (5.76)$$

となる．またこの通信路に対して，最も情報伝送効率のよい符号は $y_0 = +1$ の生成確率 $\tau := p(y_0 = +1)$ が $\frac{1}{2}$ のとき，つまり等確率のときになる場合であることも知られている．

ここでは難しい説明は抜きにして(詳しくは次小節で行う)，なぜ $\tau = \frac{1}{2}$ になるのかを直感的に説明する．図5.7を見ると，この通信路には対称性があり，(1)どの送信信号 ± 1 を送ったとしても，同じ割合でノイズが入ってしまう性質がある．(2)情報の多様性という面で見ると，等確率で送信信号を生成したほうが表現のバリエーションが増えることが知られている．それらを踏まえると，送信信号を $\tau = \frac{1}{2}$ の割合に従って発生させたほうが効率がよいから， $\tau = \frac{1}{2}$ が妥当であると考えられる．詳細は以下で説明する．

5.4.3 通信容量

通信容量 C は， y_0 の生成確率 $p(y_0 = +1) = \tau$ と $p(y_0 = -1) = 1 - \tau$ と，通信路のノイズ反転確率 b で決まる量であり，通信容量の定義は

$$C = \max_{\tau} I, \quad I := \sum_{y, y_0} p(y, y_0) \log_2 \frac{p(y, y_0)}{p(y)p(y_0)} \quad (5.77)$$

である．今回のBSCにおいては

$$I = - \sum_{y_0} p(y_0) \log_2 p(y_0) + \sum_{y_0} p(y_0) \sum_y p(y|y_0) \log_2 p(y|y_0) = \mathcal{H}(\tau) - \mathcal{H}(b) \leq 1 - \mathcal{H}(b) \quad (5.78)$$

となる．ただしShannonの2値エントロピー関数 $\mathcal{H}(u) := -u \log_2 u - (1-u) \log_2 (1-u)$ を用い， $\tau = 0.5$ のとき $\mathcal{H}(\tau)$ が最大になることが分かる．ここで I は相互情報量(mutual information)と呼び，定義から確率変数 y, y_0 の重なり具合(相関)を定量化した量であり，まったく独立の場合は0となる[40]．通信の文脈で言えば，送信信号と受信信号を表す確率変数をそれぞれ y_0, y で表したとき，相互情報量は受信信号 y を受け取ることで得られる送信信号 y_0 の情報量の大きさを表すことに相当する．これより相互情報量 I の最大値により定義される通信容量(channel capacity)とは，ある通信路を通して単位時間あたりに送り得る最大の情報量のことであり，上の定義より通信路の性質，つまり条件付き確率 $p(y|y_0)$ で決まる量である．また最大化を行う自由度は送信信号 y_0 の生成確率であり，それを調整することで効率のよい情報通信が期待できる．

5.4.4 通信路符号化定理

Shannon は「通信の数学的理論」の中で次の重要な結論（存在定理）を導き出している [182]⁹。その説明に入る前に伝送速度（もしくは符号化率）と呼ばれる 1 つの指標を導入する。

— 伝送速度 R —

原情報のビット長 N とそれを符号化した送信信号（符号語）のビット長 p との比を $R := \frac{N}{p}$ とし、伝送速度（もしくは符号化率）と定義する。伝送速度は、単位時間あたり p 文字列の中に N 文字列分の情報を含まれることを表す指標であり、伝送能力を示す目安となる量である。ちなみに R が大きいほど情報を送る際の伝送効率が良いと解釈される。

これを用いて Shannon の通信路符号化定理を説明する。

— 通信路符号化存在定理 —

ある通信路と符号化方法が与えられたとし、その符号化の伝送速度を R とする。このとき、与えられた通信路の性質で決まるある有限の臨界値 C が存在し、

- $R < C$ なら 1bit あたりの最大誤り率を 0 に限りなく近づき符号化方法が、 $N, p \rightarrow \infty$ の極限で存在する。
- 逆に $R > C$ なら、どのような符号化方法を持ってきたとしても、誤り率を 0 になる符号化は存在しない。このときの臨界値 C のことを通信容量と定義する。

実は Shannon の論文以前の認識では、誤りが発生する確率のある通信路を用いて通信を行い、できるだけ情報を正確に伝えるためには、あるいはこの通信方式の誤り訂正能力を高めるためには、一般に冗長度 α （伝送速度 R の逆数）をできるだけ大きくし、冗長したところをうまく使って、通信過程で生じた誤りを訂正できるのだと考えられていた。だが伝送速度 R は冗長度 $\alpha = \frac{p}{N}$ の逆数なので、どうしても情報を正確に伝えるには長い時間もしくは通信コストがかかる（ R が小さくなるので）と考えられていた。訂正能力のためには冗長を必要とする。それが Shannon の論文以前の常識（ジレンマ）だったのである。しかしながら Shannon は上の定理で、通信路の持つ固有の通信容量 C を少しでも小さい伝送速度であれば、いくらでも誤りを小さくすることのできる符号が存在することを示したのである。

ここで 1 つコメントする。上の定理はあくまで、符号化の存在（可能性）しか言及していないということである。なぜなら Shannon が理論限界を説明するために導入した「ランダム符号化」は、数学的な操作としては性質の良い符号化ではあるが、構成論的な符号ではないことが知られており実用的ではない。そのためそれに代わる、ランダム符号化と同等の誤り訂正能力を持つ構成論的な符号が、特殊な通信路を除いて現在もなお提示されているとは言い難い。そこで我々は次小節でパーセプトロンを用いた誤り訂正符号を提案し、性能評価を行う¹⁰。

5.4.5 誤り訂正符号の提案

これまで見てきたように、結合加重が離散値の場合に関する教師推定問題は、ノイズなしの 2 元対称通信路 (BSC) に対する誤り訂正符号と解釈することができる。つまりこの解釈では ω_0 が送信情報、パターン行列 X が符号鍵、 $\vec{y}_0$ が送信する符号語、伝送速度は $R = N/p (= 1/\alpha)$ となる。本研究の提案手法であるパーセプトロンは、誤り訂正符

⁹情報源符号化定理に関しては韓・小林が詳しい。しかし、ここではすでにそれが行われたことにし、情報源符号化された符号語を改めて原情報と見なすことにする。

¹⁰Shannon が理論的に証明したのは、 $N, p \rightarrow \infty$ の極限、つまり熱力学的極限に相当する bit 長での評価でしかない。実際、多くの場合、復号において bit 長 N, p が長くなるにつれ、送信信号推定の計算量が指数関数的に増え、復号困難になるという問題を抱えている。従来の符号理論研究では、それを回避するために、計算量を少なくなるように bit 長を短くし、その代わりに、代数幾何や整数論の枠組みを使い、誤り訂正の「仕掛け」を工夫する、という戦略がとられてきた。しかしながら、近年計算機の性能が上がり、大規模統計モデル（確率モデル）で定式化を行った符号に対して、ある種の近似を用いて、現実の実行時間で、高い精度で Shannon 限界に接近する誤り訂正符号の復号方法の研究が数多く報告されている。

号の符号器/復号器に対応することができ、送信情報を完全に同定できる限界の符号化率 $R_{\text{GD}} := 1/\alpha_{\text{GD}}$ はこのシステムの通信容量を表す。

以下では誤り訂正符号の流れを説明する。まず原情報を $\vec{w}_0 \in \{\pm 1\}^N$ で表し、パターン $\vec{x} := \{x_1, x_2, \dots, x_N\} \in \mathbf{R}^N$ に対して送信信号 $y_0 \in \{\pm 1\}$

$$y_0 = \text{sgn} \left(\frac{\vec{x}^T \vec{w}_0}{\sqrt{N}} \right) \quad (5.79)$$

を割り当てるパーセプトロンを符号器とする。このとき p 個の $\vec{x}_\mu$, $(\mu = 1, 2, \dots, p)$ に対して、式 (5.79) を用いて符号語 $\vec{y}_0 := \{y_{10}, y_{20}, \dots, y_{p0}\}$ をつくり、受信側に送信信号 $\vec{y}_0$ を送り（送受信側で $\vec{x}_\mu$ を共有していると仮定する）、受信側では p 個の受信信号 $\vec{y}$ を受け取る。このとき通信路のノイズの統計性により、 $\vec{y}$ の条件付き確率は

$$P(\vec{y}|\vec{y}_0) = \prod_{\mu=1}^p p(y_\mu|y_{\mu 0}) \quad (5.80)$$

に比例する。これより $\vec{y}$ を受け取ったときの $\vec{w}_0$ に対する事後確率は Bayes の定理より、

$$P(\vec{w}_0|\xi^p) := \frac{P(\vec{w}_0) \prod_{\mu=1}^p p(y_\mu|y_{\mu 0})}{\sum_{\vec{w}_0} P(\vec{w}_0) \prod_{\mu=1}^p p(y_\mu|y_{\mu 0})} \quad (5.81)$$

と表される。ここで $P(\vec{w}_0)$ は $\vec{w}_0$ の事前確率を表し、情報源符号化を行った原情報であることを仮定する [40]。さらに条件付き確率 $p(y|y_0)$ は反転確率 b を用いて $p(y|y_0) = b + (1-2b)\Theta(y y_0) = b + (1-2b)\Theta\left(\frac{y \vec{x}^T \vec{w}_0}{\sqrt{N}}\right) =: g\left(y, \frac{\vec{x}^T \vec{w}_0}{\sqrt{N}}\right)$ と表すことができるため、この誤り訂正符号の性能解析はそのまま本学位論文の提案手法を用いて評価することができる。これよりエントロピー S は

$$S = F(\chi_w, \chi_u) + \frac{\alpha \chi_u}{2} \left(\frac{\langle \lambda \rangle}{\alpha} - \tilde{q}_u \right) + 2\alpha \int_{-\infty}^{\infty} Dz \tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\frac{\langle \lambda \rangle}{\alpha} - \tilde{q}_u}} \right) \log \tilde{H} \left(z \sqrt{\frac{\tilde{q}_u}{\frac{\langle \lambda \rangle}{\alpha} - \tilde{q}_u}} \right) - \frac{\tilde{q}_w}{2} (1 + q_w) + \int_{-\infty}^{\infty} Dz \log 2 \cosh \left(z \sqrt{\tilde{q}_w} + \tilde{q}_w \right) + \alpha \log 2\mathcal{H}(b) \quad (5.82)$$

$$\tilde{H}(u) := b + (1-2b)H(u) \quad (5.83)$$

となる。ただし式 (5.82) の $\alpha \log 2\mathcal{H}(b)$ は $\chi_w = 0, \tilde{q}_u = \frac{\langle \lambda \rangle}{\alpha}$ のときに $S = 0$ になるように決めてある¹¹。次小節ではこの誤り訂正符号の有効性を検証するために $\vec{x}$ が i.i.d. 例題と直交例題のそれぞれについて数値実験を行った。

5.4.6 数値結果

ノイズ反転がないとき ($b = 0$) の R_{GD} を比較した場合、i.i.d. 例題に比べて直交例題のほうが Shannon 限界 $C = 1$ に近い、つまり通信効率が優れていることが分かる (5.3.2 小節の結果を用いると $C \geq 1/1.101 > 1/1.245$ となる)。我々の提案手法を用いればノイズのある状況においても解析することが可能であり、図 5.8 は反転確率 b の BSC に対して R_{GD} を評価した結果を表す。復号方法については 73 ページの TAP 法アルゴリズムを用いればよく、5.3.2 小節などの議論から分かるようにレプリカ解析の結果と TAP 法の典型的な振る舞いの整合性が取れているため、図 5.8 ではレプリカ法の結果のみをプロットした。 $0 \leq b < 0.5$ の全域で、i.i.d. 例題より直交例題の方が Shannon 限界に近いことから、直交化により優れた通信効率を確保できることが確認できる。

¹¹

$2\alpha \int_{-\infty}^{\infty} Dz Dx \left(b + (1-2b)\Theta(z\sqrt{\tilde{q}_u}) \right) \log \int_{-\infty}^{\infty} Dx \left(b + (1-2b)\Theta(z\sqrt{\tilde{q}_u}) \right) = 2\alpha \times \left(\frac{b}{2} \log b + \frac{1-b}{2} \log(1-b) \right) = -\alpha \log 2\mathcal{H}(b)$

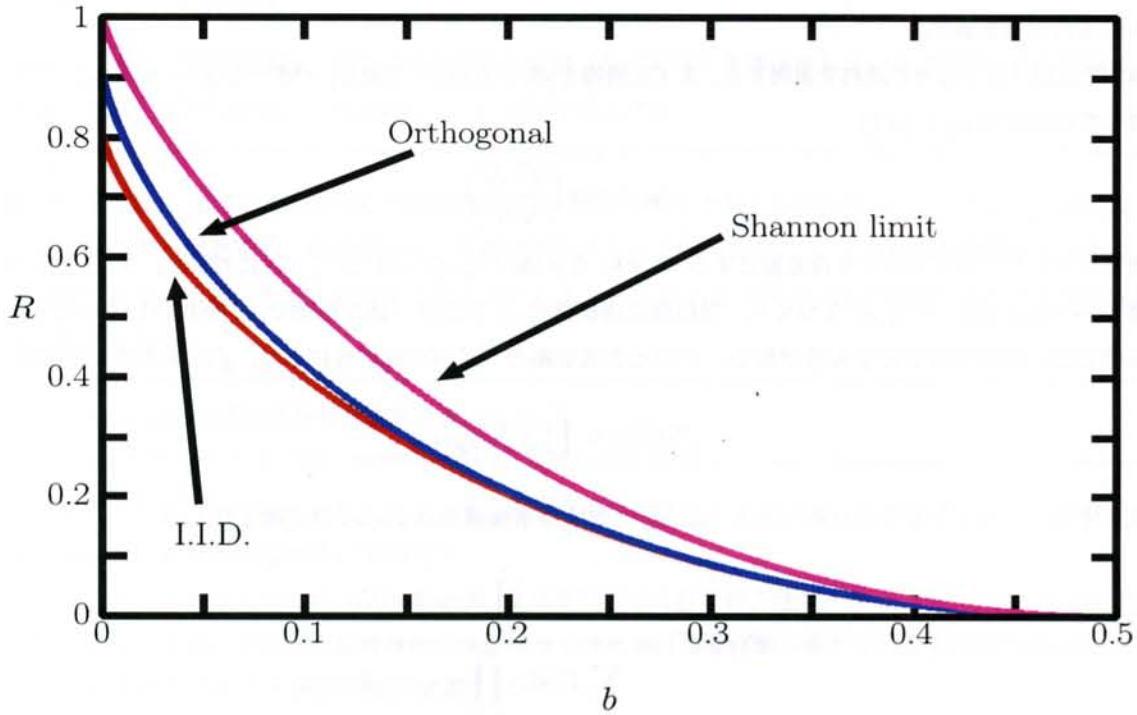


図 5.8: ω_0 を同定できる符号化率 R の上限 R_{GD} (通信容量) と反転確率 b の関係. i.i.d. 例題に比べて直交例題 (通信では Welch 限界を達成する系列) のほうが Shannon 限界 $C = 1 - \mathcal{H}(b)$ に近い.

5.5 まとめ

本章では i.i.d. 例題と直交例題の記憶容量問題と教師推定問題などについて議論し、以下のことが解明された.

1. 結合加重が離散値の場合の記憶容量問題に対して, i.i.d. 例題 ($\alpha_C \simeq 0.833$) に比べて直交例題 ($\alpha_C \simeq 0.940$) のほうが多くのパターンを覚えることができることを解析的に示した.
2. 結合加重が連続値の場合の記憶容量問題に対して, i.i.d. 例題と直交例題では同じ記憶容量 $\alpha_C = 2$ が得られた. さらに $X^T X \in \mathcal{M}_{N \times N}$ のランクが N_ϱ である任意の固有値分布から生成されるランダム長方形列 X に対して, $\alpha_C = 2\varrho$ であることも解析的に示された. 直感的に説明すると拘束条件を保ちながら, 適当な座標変換を用いて新たな座標系を定義しなおしたとき, その結合加重の空間を, クラス識別に寄与する N_ϱ の「有効次元」の空間と, クラス識別に寄与しない $N(1 - \varrho)$ の「無効次元」の空間に分けることができるため, 実効的な結合加重の次元スケールは N_ϱ となり, $\alpha_C = 2\varrho$ が得られる. これは $\vec{w}$ が連続値の場合の特有の性質であるが, $\vec{w}$ が離散値の場合は新しい変数を取り直すことができるかどうか自明ではないため, 記憶容量 α_C が固有値分布 $\rho(\lambda)$ に強く依存することも示された.
3. 結合加重が離散値の場合の教師推定問題に対して, i.i.d. 例題 ($\alpha_{GD} \simeq 1.245$) に比べて直交例題 ($\alpha_{GD} \simeq 1.101$) のほうが少ないパターンで教師ベクトルを同定できることを解析的に示した.
4. 結合加重が連続値の場合の教師推定問題に対して, i.i.d. 例題と直交例題では α が大きい極限で, 同じく $q_w = 1$ からのズレの漸近形 $\varepsilon = 1 - q_w \simeq \frac{1.926}{\alpha^2}$ が得られた. さらに $X^T X \in \mathcal{M}_{N \times N}$ のランクが N_ϱ である任意の固有値分布から生成されるランダム長方形列 X に対して, $\varepsilon \simeq 1 - \varrho + \frac{1.926\varrho^3}{\alpha^2}$ であることも解析的に示された. 直感的に説明すると新たな座標系で N_ϱ の「有効次元」の空間と $N(1 - \varrho)$ の「無効次元」の空間に分けることができ, 実効的な結合加重での各変数 $\hat{\varepsilon}, \hat{q}_w, \hat{\alpha}$ と元の変数 ε, q_w, α はそれぞれ $\hat{\varepsilon} = 1 - \hat{q}_w \simeq \frac{1.926}{\hat{\alpha}^2}$, $\frac{\hat{q}_w}{\hat{\alpha}} = \frac{q}{\alpha}$, $\hat{\alpha} = \frac{\alpha}{\varrho}$ の関係を満たすため, $\varepsilon = 1 - q_w \simeq 1 - \varrho + \frac{1.926\varrho^3}{\alpha^2}$ が得られる. これは $\vec{w}$ が連続値の場合の特有の性質である.

が、 $\vec{w}$ が離散値の場合は新しい変数を取り直すことができるかどうか自明ではないため、記憶容量 α_C が固有値分布 $\rho(\lambda)$ に強く依存することも示された。

5. 上記の4つの問題において、システムサイズ無限大の極限を仮定して導出した「レプリカ解析」の結果とシステムサイズ有限のサンプルに対する「TAP 法」の典型的な振る舞いが数値的に確認した。2つの結果が一致することは、それらの仮定が正しいことを裏付けているとも言える。
6. さらにシステムサイズ N が比較的小さな場合に（記憶容量問題の場合 $2 \leq N \leq 20$ ，教師推定問題の場合 $6 \leq N \leq 19$ ），全数探索を用いて求めた結果を，大システム極限に外挿することによって，レプリカ解の結果と整合する結論が得られた。今回のパーセプトロン学習に限定して言えば理論解と数え上げ実験の結果の整合性が取れるということは，理論展開のために用いたいくつかの仮定が保証される根拠の1つと考えることができる。
7. 結合加重が離散値の場合の教師推定問題は，ちょうどノイズ無しの2元対称通信路を用いた，情報通信の誤り訂正符号と読み替えることができる。それを踏まえて，反転確率 b , ($0 \leq b < 0.5$) の時に対してのパーセプトロンを用いた誤り訂正符号の誤り訂正能力を解析的に評価し，i.i.d. 例題の場合に比べて，直交例題の場合が情報理論で知られている Shannon 限界に近づいていることを明らかにした。

第6章 結論と今後の展望

本章では本学位論文の結論と、今後の展望として提案手法のいくつかの応用例を述べる。

6.1 本学位論文の結論

本学位論文では入力間に相関が含まれる参照例題に対するパーセプトロンのクラス識別能力を評価するために既存の解析手法の拡張を行い、相関の有無がクラス識別能力に与える影響を定量的に評価した。特に入力の組を並べた入力行列 X を用いた自己相関行列 $X^T X$ の固有値分布で特徴付けられる、ランダム長方形行列の積分公式 $F(x, y)$ を用いることで、以下の2つのクラス識別問題の既存手法を体系的に拡張することができた。その1つはシステムサイズが無限大の極限における、システムのマクロな変数の間で成り立つ関係式を導き、その関係式を満たすマクロ変数を用いることで、パーセプトロンの典型的な学習性能を評価することのできる「レプリカ解析」であり、もう1つは、平均場近似の考えを下にして、実際に与えられた参照例題に対して、それらを説明するミクロな変数を導くことができる Thouless-Anderson-Palmer 法であり、通信技術の文脈では「復号アルゴリズム」として見ることもできる。

第2章では、パーセプトロン学習を Bayes 推定の枠組みで定式化し、その事後確率を Boltzmann 分布として解釈することで、統計力学的なアプローチが可能になることを述べた。また大自由度複雑系のモデルの雛形である、Sherrington-Kirkpatrick 模型を用いて、レプリカ解析と TAP 法を紹介し、さらに本研究のキーになるランダム行列の積分公式 $F(x, y)$ を導出した。本研究では詳しく触れていないが、ランダム正方形行列の積分公式 $G(x)$ を用いたいくつかの研究も行われている [109, 191, 192]。第3章では、既存研究で扱われてきた、入力が独立同分布を仮定した学習例題に対する、パーセプトロン学習の理論的な結果を紹介した。これは、パーセプトロン学習の統計力学的な定式化の意味でも、本研究で扱った直交例題と比較する意味でも、重要な部分である。第4章は、本研究の核となる解析手法を説明した。特に2.3節で導入したランダム行列の積分公式を用いることで、既存手法では解析が困難な、入力間に相関が含まれる例題に対しても、統計力学的手法が拡張できることを説明した。第5章では第4章を受けて、独立同分布に従う例題集合と各々直交している例題集合の「記憶容量問題」と「教師推定問題」に対して、結合加重が離散値の場合と拘束条件付きの連続値の場合に分けて検証を行った。特に連続値の場合は、 X の統計性というよりはむしろ、行列のランクに強く依存するという結論を導き出した。なぜなら拘束条件付きとはいえ、適当な回転行列を用いて座標系を取り直すことで、フルランクの解析に帰着させることができるため、ランクの次元が解析に最も寄与することを明らかにした。一方離散値では連続値と違って、自明な回転行列を定義して新たな離散値を取るベクトルを取り直すことがいつもできるとは限らないため、パーセプトロンの学習能力が固有値分布に強く寄与することを示した。さらに誤り訂正符号に対する本手法の有効性を検証した。

しかしながら現実的なデータの相関は必ずしも固有値分布のみで特徴付けられるわけではない。一方で、相関情報を詳細に考慮することは一般に膨大な計算量を要する。本研究で示した手法は、相関情報の一部を実効的な計算量で比較的容易に取り入れることができる簡便法と位置づけられ、バイオインフォマティクスやデータマイニングなどの分野において大量の計算資源を投入する本格的な解析を行う前のデータスクリーニング手法としての利用などが期待される。

6.2 応用；機械学習

ここでは本学位論文で提案した手法の、パーセプトロン学習以外の機械学習への応用の可能性についていくつか言及する。

6.2.1 アンサンブル学習

機械学習の学習方法の1つである「アンサンブル学習」とは、性能が低い学習器を複数用意し、ある課題に対して各々が意思決定を行い（もしくは単純にクラスを判別し）、そのクラスラベルの多数決によってクラス識別を行う学習方法である。特に性能の高い学習器を1つ用意するよりも、集団として意思決定するほうが、高い学習能力を保証する状況が存在することが知られている [125, 142, 143]。また本研究の関連として述べれば、多くのアンサンブル学習では、学習器が学習する過程で、各々独立に学習する場合を扱ってきたが、我々の得られた知見から各々が独立に学習するよりも相関を持たせるように、もしくは各々が分業的、相互補完的に学習するほうが、学習能力が向上することが容易に想像される。またこのような学習方法はセンサーネットワークや分散処理などの比較的安価な端末によって情報伝達や情報処理を行うことと類似しており [12]、相関が含まれる場合のアンサンブル学習の知見を通信分野に応用することも考えられる。

6.2.2 バイオインフォマティックス

本研究で議論した結合加重は、2つの値をとる（離散値の場合）と、拘束条件付きの実数値をとる（連続値の場合）に限定して有効性の検証を行ってきた。しかしながら提案手法はより広い結合加重のクラスに対しても、表現することは可能である（レプリカ法やTAP法では任意の結合加重の確率分布 $P(w)$ を用いて定式化していたことに注意）。たとえば $w = 0$ にデルタピークを持ち、それ以外はL1ノルムの指数分布に従うような場合、つまり事前確率が

$$P(w) = \begin{cases} c & w = 0 \\ \frac{\nu(1-c)}{2} e^{-\nu|w|} & \text{otherwise} \end{cases} \quad (6.1)$$

のように表現できる場合を考えることもできる。特に c が1に近い場合（ほとんどすべての結合加重が0を示すことを意味する）の解析は、バイオインフォマティックスの分野において重要である。なぜなら多くのマイクロアレイデータなどは [203]、発症（クラス識別の文脈で言えばクラスラベル）に寄与する成分が大自由度なデータ次元に比べて少数であることが知られており（つまりほとんど $w = 0$ である）、またマイクロアレイデータセットを入力間に相関が含まれるサンプルセットであると見なせば、本提案手法で解析することが可能である。すなわちバイオインフォマティックスへの新たな解析方法としての応用が期待される¹。

6.3 応用；情報技術

続いて情報理論への応用をいくつか紹介する。

6.3.1 歪みありデータ圧縮

T. Hosaka, Y. Kabashima and H. Nishimori [96], T. Hosaka and Y. Kabashima [97] や T. Murayama [147] らにより、歪みありデータ圧縮に対して統計力学的解析が行われている。特に [96, 97] ではデータ圧縮器として、我々と同様にパーセプトロンを用いており、式 (1.21) の閾値 κ を調整することや、出力関数を別の関数形、たとえば $y = \text{sgn} \left(\left| \frac{\vec{x}^T \vec{w}}{\sqrt{N}} \right| - \kappa \right)$ などに変えることで、データ圧縮として必要な偏りのある出力 y を生成させ、パーセプトロンを用いた歪みありデータ圧縮に対する圧縮効率を評価している。ここでも本研究の入力行列 X に相当する部分が暗黙のうちに独立同分布に従うと仮定されており、彼らは出力関数を工夫することで理論限界に接近する歪みあり圧縮アルゴリズムを提案した。しかしながら出力関数の非線形性の効果が比較的大きいため、摂動論の仮定の下で導出された反復アルゴリズムではなかなか解が収束しないことが知られている。そこで我々の提案する方法を用いることで、出力関数を比較的非線形性の小さな式 (1.21) を用いてかつ、相関のある例題にすることで、実効的な時間でデータ圧縮のできるアルゴリズムが提案できると期待される。

¹提案手法を用いて与えられた参照例題 ξ^P に対して実際に結合加重の推定量をTAP法を用いて導くことができ、またレプリカ解析を用いてモデルパラメータ c, ν を適応的に決めることができる。

6.3.2 符号分割多元接続方式

現在の無線通信技術の一つである「符号分割多元接続方式 (Code Division Multiple Access: CDMA)」とは、複数の通信端末から送られてきた信号を単一の基地局で受信し、重畳した信号から受信側で各端末の送信情報を推定し、復調する通信方式である [106, 133]。またいわゆる第三世代の携帯電話システムに広く採用されていることもあり、情報工学の分野で現在もなお、精力的に研究・開発が行われている通信技術である。これまでこの CDMA に対して統計力学的な解析によりいくつかの知見が得られている [76, 109, 191, 194]。特に本研究の入力行列に相当する部分（拡散符号系列など）に対する知見は広がりつつあるが、一般に重畳した信号に加わるノイズは白色ノイズと暗黙のうちに仮定している場合が少なくない。実際は信号の直接波と環境に依存したマルチパスから来る間接波の重ね合わせになるため、「白色ノイズの仮定」の下で得られた知見では不十分である。そこで環境ノイズを仲上・ライスフェージングや、より一般化されたレイリーフェージングを用いてマルチパスから来る影響を特徴づけ、より現実に近い状況下での CDMA の通信効率を評価する必要がある。幸いなことにこの問題は Bayes 推論の枠組みで定式化することができるため [76, 109, 191, 194]、相関ノイズ下での CDMA の性能解析に対する本提案手法の有効性が期待される。

まとめ

以上に述べたように我々が提案した手法は、大自由度大規模統計データを扱うような、さらにベイズ推定の枠組みで定式化される問題に対して、有効性が期待できる。今後、機械学習や情報理論などと統計力学との学際領域的研究を推し進め、有機的な研究により、列挙したような未解決問題を解析していきたい。

付録 A $H(u)$ について

ここでは $H(u)$ について知られている性質を説明する． $H(u)$ の定義は，

$$H(u) := \int_{-\infty}^{\infty} Dx \Theta(x-u) = \int_u^{\infty} Dx, \quad Dx := \frac{dx}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (\text{A.1})$$

である (図 A.1)．また Euler の補誤差関数 $\text{erfc}(u) := \frac{2}{\sqrt{\pi}} \int_u^{\infty} e^{-x^2} dx$ を用いて，

$$H(u) = \frac{1}{2} \text{erfc}\left(\frac{u}{\sqrt{2}}\right) = \frac{1}{2} \left(1 - \text{erf}\left(\frac{u}{\sqrt{2}}\right)\right), \quad H(u) = \gamma \implies u = \sqrt{2} \text{erfc}^{-1}(2\gamma) \quad (\text{A.2})$$

で表すことができる．定義から

$$H(u) + H(-u) = 1 \quad (\text{A.3})$$

を満たす．さらに $H(u)$ の漸近形は

$$H(u) \sim \begin{cases} \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \left(u - \frac{u^3}{6} + \dots\right) & u \rightarrow 0 \\ \frac{1}{\sqrt{2\pi}u} e^{-\frac{u^2}{2}} \left(1 - \frac{1}{u^2} + \frac{3}{u^4} + \dots\right) & u \rightarrow \infty \\ 1 + \frac{1}{\sqrt{2\pi}u} e^{-\frac{u^2}{2}} \left(1 - \frac{1}{u^2} + \frac{3}{u^4} + \dots\right) & u \rightarrow -\infty \end{cases} \quad (\text{A.4})$$

となる．これより

$$\frac{e^{-\frac{u^2}{2}}}{H(u)} \sim \begin{cases} \sqrt{2\pi}u & u \gg 1 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.5})$$

が得られる (図 A.2)．さらに $a \in \mathbf{R}$ に対して¹

$$\int_{-\infty}^{\infty} Dx H(ax) = \frac{1}{2} \quad (\text{A.11})$$

$$\int_{-\infty}^{\infty} Dxx H(ax) = -\frac{a}{\sqrt{2\pi}(1+a^2)} \quad (\text{A.12})$$

$$\int_{-\infty}^{\infty} Dxx^2 H(ax) = \frac{1}{2} \quad (\text{A.13})$$

1

$$\int_{-\infty}^{\infty} Dxx^{2n} H(ax) = \frac{(2n)!}{2^{n+1}n!} \quad (\text{A.6})$$

$$\int_{-\infty}^{\infty} Dxx^{2n+1} H(ax) = -\frac{2^{n+\frac{1}{2}}}{\pi} \Gamma\left(n + \frac{3}{2}\right) \int_0^{\tan^{-1}(a)} d\theta \cos^{2n+1} \theta = -\frac{1}{\sqrt{2\pi}} \frac{(2n+2)!}{2^{n+1}(n+1)!} \frac{a}{\sqrt{1+a^2}} \sum_{k=0}^n \frac{n C_k}{2k+1} \left(\frac{-a^2}{1+a^2}\right)^k \quad (\text{A.7})$$

や

$$\int_{\alpha}^{\infty} Dxe^{\beta x} = e^{\frac{\beta^2}{2}} H(\alpha - \beta) \quad (\text{A.8})$$

$$\int_{-\infty}^{\infty} Dxe^{-\frac{\alpha}{2}(x-\beta)^2} = \exp\left[-\frac{1}{2} \log(1+\alpha) - \frac{\alpha\beta^2}{2(1+\alpha)}\right] \quad (\text{A.9})$$

$$\int_{-\infty}^{\infty} Dx H(\alpha + \beta x) = H\left(\frac{\alpha}{\sqrt{1+\beta^2}}\right) \quad (\text{A.10})$$

が得られる．

や

$$\int_0^\infty Dx H(ax) = \frac{1}{2\pi} \tan^{-1} \left(\frac{1}{a} \right) = \frac{1}{2\pi} \left(\frac{\pi}{2} - \tan^{-1}(a) \right) \quad (\text{A.14})$$

$$\int_0^\infty Dxx H(ax) = \frac{1}{2\sqrt{2\pi}} \left(1 - \frac{a}{\sqrt{1+a^2}} \right) \quad (\text{A.15})$$

$$\int_0^\infty Dxx^2 H(ax) = \frac{1}{2\pi} \tan^{-1} \left(\frac{1}{a} \right) - \frac{a}{2\pi(1+a^2)} \quad (\text{A.16})$$

が知られている [55].

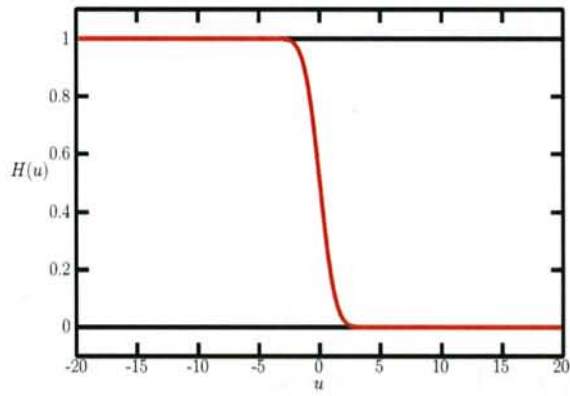


図 A.1: $H(u) = \int_u^\infty Dx$ の図

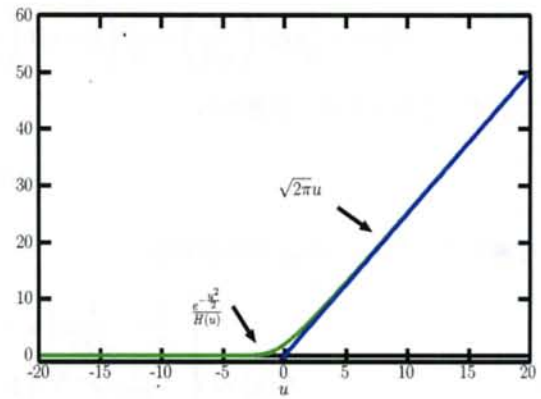


図 A.2: $\frac{e^{-\frac{u^2}{2}}}{H(u)}$ の図. $u \gg 1$ において $\sqrt{2\pi}u$ に漸近する.

付録B 式(4.9)の鞍点条件

ここでは式(4.9)の鞍点条件を議論する。Extr を省略して、

$$\begin{aligned} \psi(n) = & \mathcal{F}(Q_w, Q_u) + \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \prod_{a=1}^n g(y, \Delta_a) e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}} + \frac{\alpha}{2} \text{Tr} \tilde{Q}_u Q_u - \frac{1}{2} \text{Tr} \tilde{Q}_w Q_w \\ & + \log \sum_{\vec{w}} P(\vec{w}) e^{\frac{1}{2} \vec{w}^T \tilde{Q}_w \vec{w}} \end{aligned} \quad (\text{B.1})$$

$$\mathcal{F}(Q_w, Q_u) = \frac{1}{2} \text{Tr}(\Lambda_w - I) + \frac{\alpha}{2} \text{Tr}(\Lambda_u - I) - \frac{\alpha - 1}{2} \log \det \Lambda_u - \frac{1}{2} \int d\lambda \rho(\lambda) \log \det (\Lambda_w \Lambda_u + \lambda Q_w Q_u) \quad (\text{B.2})$$

で表す。ただし $\vec{w}, \vec{u}, \vec{\Delta}$ は n 次元ベクトルであり、 $Q_w, Q_u, \tilde{Q}_w, \tilde{Q}_u, \Lambda_w, \Lambda_u$ は n 次正方行列である。

一般解

ここで式(B.2)の極値を示す Λ_w, Λ_u は

$$\Lambda_u^{-1} = \int d\lambda \rho(\lambda) (\Lambda_w \Lambda_u + \lambda Q_w Q_u)^{-1} \quad (\text{B.3})$$

$$\Lambda_w = \alpha (\Lambda_u - I) + I \quad (\text{B.4})$$

を満たす。さらに

$$\frac{\partial \mathcal{F}(Q_w, Q_u)}{\partial Q_w} = -\frac{1}{2} \int d\lambda \rho(\lambda) \lambda Q_u (\Lambda_w \Lambda_u + \lambda Q_w Q_u)^{-1} = \frac{1}{2} Q_w^{-1} (\Lambda_w - I) \quad (\text{B.5})$$

$$\frac{\partial \mathcal{F}(Q_w, Q_u)}{\partial Q_u} = -\frac{1}{2} \int d\lambda \rho(\lambda) (\Lambda_w \Lambda_u + \lambda Q_w Q_u)^{-1} \lambda Q_w = \frac{\alpha}{2} (\Lambda_u - I) Q_u^{-1} \quad (\text{B.6})$$

を用いると、

$$\tilde{Q}_w = Q_w^{-1} (\Lambda_w - I) \quad (\text{B.7})$$

$$\tilde{Q}_u = (I - \Lambda_u) Q_u^{-1} \quad (\text{B.8})$$

が得られる。また $\tilde{Q}_w, \tilde{Q}_u$ の微分により

$$Q_w = \frac{\sum_{\vec{w}} P(\vec{w}) \vec{w} \vec{w}^T e^{\frac{1}{2} \vec{w}^T \tilde{Q}_w \vec{w}}}{\sum_{\vec{w}} P(\vec{w}) e^{\frac{1}{2} \vec{w}^T \tilde{Q}_w \vec{w}}} \quad (\text{B.9})$$

$$Q_u = \frac{\sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \vec{u} \vec{u}^T \prod_{a=1}^n g(y, \Delta_a) e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}}}{\sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \prod_{a=1}^n g(y, \Delta_a) e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}}} \quad (\text{B.10})$$

が導出される。特に $u e^{iu\Delta} = \frac{\partial}{\partial i\Delta} e^{iu\Delta}$ を用いれば式(B.10)の各成分は

$$q_{uab} = \frac{\sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \prod_{a=1}^n g(y, \Delta_a) \frac{\partial}{\partial i\Delta_a} \frac{\partial}{\partial i\Delta_b} e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}}}{\sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \prod_{a=1}^n g(y, \Delta_a) e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}}} \quad (\text{B.11})$$

であり, 部分積分

$$\int_{-\infty}^{\infty} d\Delta g(\Delta) f'(\Delta) = [g(\Delta) f(\Delta)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} d\Delta g'(\Delta) f(\Delta) \quad (\text{B.12})$$

を用いれば, 式 (B.11) が簡単化される. ただし式 (B.12) の第 1 項は, $\vec{u}$ の積分を実行すれば, 被積分関数の両端がガウス積分の裾にあたるため, 0 と見なすことができる. よって

$$q_{uab} = - \frac{\sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \frac{g'(y, \Delta_a)}{g(y, \Delta_a)} \frac{g'(y, \Delta_b)}{g(y, \Delta_b)} \left(\prod_{c=1}^n g(y, \Delta_c) \right) e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}}}{\sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \prod_{a=1}^n g(y, \Delta_a) e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}}} \quad (\text{B.13})$$

$$q_{uaa} = - \frac{\sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \frac{g''(y, \Delta_a)}{g(y, \Delta_a)} \left(\prod_{c=1}^n g(y, \Delta_c) \right) e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}}}{\sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \prod_{a=1}^n g(y, \Delta_a) e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}}} \quad (\text{B.14})$$

となる.

鞍点方程式

$P(y), g(y, \Delta), P(w), \rho(\lambda)$ を用い,

$$\Lambda_u^{-1} = \int d\lambda \rho(\lambda) (\Lambda_w \Lambda_u + \lambda Q_w Q_u)^{-1} \quad (\text{B.15})$$

$$\Lambda_w = \alpha (\Lambda_u - I) + I \quad (\text{B.16})$$

$$\tilde{Q}_w = Q_w^{-1} (\Lambda_w - I) \quad (\text{B.17})$$

$$\tilde{Q}_u = (I - \Lambda_u) Q_u^{-1} \quad (\text{B.18})$$

$$Q_w = \frac{\sum_{\vec{w}} P(\vec{w}) \vec{w} \vec{w}^T e^{\frac{1}{2} \vec{w}^T \tilde{Q}_w \vec{w}}}{\sum_{\vec{w}} P(\vec{w}) e^{\frac{1}{2} \vec{w}^T \tilde{Q}_w \vec{w}}} \quad (\text{B.19})$$

$$Q_u = \frac{\sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \vec{u} \vec{u}^T \prod_{a=1}^n g(y, \Delta_a) e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}}}{\sum_y P(y) \int_{-\infty}^{\infty} \frac{d\vec{\Delta} d\vec{u}}{(2\pi)^n} \prod_{a=1}^n g(y, \Delta_a) e^{i\vec{u}^T \vec{\Delta} - \frac{1}{2} \vec{u}^T \tilde{Q}_u \vec{u}}} \quad (\text{B.20})$$

を満たす秩序変数 $Q_w, Q_u, \tilde{Q}_w, \tilde{Q}_u, \Lambda_w, \Lambda_u$ によってシステムの典型的な振る舞いを評価することができる.

レプリカ対称解

さらに

$$(q_{wab}, q_{uab}) = \begin{cases} (\chi_w + q_w, \chi_u + q_u) & a = b \\ (q_w, q_u) & \text{otherwise} \end{cases}, \quad (\tilde{q}_{wab}, \tilde{q}_{uab}) = \begin{cases} (\tilde{\chi}_w, \tilde{\chi}_u) & a = b \\ (\tilde{q}_w, \tilde{q}_u) & \text{otherwise} \end{cases} \quad (\text{B.21})$$

のような RS 解を仮定すると, 67 ページの議論から

$$\mathcal{F}(Q_w, Q_u) = (n-1)F(\chi_w, \chi_u) + F(\chi_w + nq_w, \chi_u + nq_u) \quad (\text{B.22})$$

が得られる。また

$$\frac{\partial \mathcal{F}(Q_w, Q_u)}{\partial \chi_w} = (n-1) \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} + \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_w} \quad (\text{B.23})$$

$$\frac{\partial \mathcal{F}(Q_w, Q_u)}{\partial \chi_u} = (n-1) \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} + \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_u} \quad (\text{B.24})$$

$$\frac{\partial \mathcal{F}(Q_w, Q_u)}{\partial q_w} = n \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_w} \quad (\text{B.25})$$

$$\frac{\partial \mathcal{F}(Q_w, Q_u)}{\partial q_u} = n \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_u} \quad (\text{B.26})$$

と

$$\frac{\alpha}{2} \text{Tr} Q_u \tilde{Q}_u - \frac{1}{2} \text{Tr} Q_w \tilde{Q}_w = \frac{n\alpha}{2} \tilde{\chi}_u (\chi_u + q_u) + \frac{n(n-1)\alpha}{2} \tilde{q}_u q_u - \frac{n}{2} \tilde{\chi}_w (\chi_w + q_w) - \frac{n(n-1)}{2} \tilde{q}_w q_w \quad (\text{B.27})$$

を用いて

$$\tilde{\chi}_w = \frac{2}{n} \left((n-1) \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} + \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_w} \right) = \frac{\Lambda_w - 1}{\chi_w} + \frac{1}{n} \left(\frac{\tilde{\Lambda}_w - 1}{\chi_w + nq_w} - \frac{\Lambda_w - 1}{\chi_w} \right) \quad (\text{B.28})$$

$$\tilde{\chi}_u = -\frac{2}{n\alpha} \left((n-1) \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} + \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_u} \right) = -\frac{\Lambda_u - 1}{\chi_u} - \frac{1}{n} \left(\frac{\tilde{\Lambda}_u - 1}{\chi_u + nq_u} - \frac{\Lambda_u - 1}{\chi_u} \right) \quad (\text{B.29})$$

$$\tilde{q}_w = \frac{2}{n} \left(-\frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} + \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_w} \right) = \frac{1}{n} \left(\frac{\tilde{\Lambda}_w - 1}{\chi_w + nq_w} - \frac{\Lambda_w - 1}{\chi_w} \right) \quad (\text{B.30})$$

$$\tilde{q}_u = -\frac{2}{n\alpha} \left(-\frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} + \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_u} \right) = -\frac{1}{n} \left(\frac{\tilde{\Lambda}_u - 1}{\chi_u + nq_u} - \frac{\Lambda_u - 1}{\chi_u} \right) \quad (\text{B.31})$$

となる。ただし $F(\chi_w, \chi_u)$ の共役な変数を Λ_w, Λ_u とし、 $F(\chi_w + nq_w, \chi_u + nq_u)$ の共役な変数を $\tilde{\Lambda}_w, \tilde{\Lambda}_u$ とすると、

$$-2 \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_w} = \frac{1 - \Lambda_w}{\chi_w} \quad (\text{B.32})$$

$$-\frac{2}{\alpha} \frac{\partial F(\chi_w, \chi_u)}{\partial \chi_u} = \frac{1 - \Lambda_u}{\chi_u} \quad (\text{B.33})$$

$$-2 \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_w} = \frac{1 - \tilde{\Lambda}_w}{\chi_w + nq_w} \quad (\text{B.34})$$

$$-\frac{2}{\alpha} \frac{\partial F(\chi_w + nq_w, \chi_u + nq_u)}{\partial \chi_u} = \frac{1 - \tilde{\Lambda}_u}{\chi_u + nq_u} \quad (\text{B.35})$$

の関係が成り立ち、共役変数は

$$\Lambda_w = \alpha(\Lambda_u - 1) + 1 \quad (\text{B.36})$$

$$\frac{1}{\Lambda_u} = \int \frac{d\lambda \rho(\lambda)}{\Lambda_w \Lambda_u + \lambda \chi_w \chi_u} \quad (\text{B.37})$$

$$\tilde{\Lambda}_w = \alpha(\tilde{\Lambda}_u - 1) + 1 \quad (\text{B.38})$$

$$\frac{1}{\tilde{\Lambda}_u} = \int \frac{d\lambda \rho(\lambda)}{\tilde{\Lambda}_w \tilde{\Lambda}_u + \lambda(\chi_w + nq_w)(\chi_u + nq_u)} \quad (\text{B.39})$$

を満たす。さらに式 (B.9), 式 (B.13) と式 (B.14) から

$$\chi_w + q_w = \frac{\int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n \left(\frac{\sum_w w^2 P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}}}{\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}}} \right)}{\int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n} \quad (\text{B.40})$$

$$q_w = \frac{\int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n \left(\frac{\sum_w w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}}}{\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}}} \right)^2}{\int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + zw\sqrt{\tilde{q}_w}} \right)^n} \quad (\text{B.41})$$

と

$$\chi_u + q_u = - \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg \left(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u} \right) \right)^n \left(\frac{\int_{-\infty}^{\infty} Dxg'' \left(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u} \right)}{\int_{-\infty}^{\infty} Dxg \left(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u} \right)} \right)}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg \left(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u} \right) \right)^n} \quad (\text{B.42})$$

$$q_u = - \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg \left(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u} \right) \right)^n \left(\frac{\int_{-\infty}^{\infty} Dxg' \left(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u} \right)}{\int_{-\infty}^{\infty} Dxg \left(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u} \right)} \right)^2}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dxg \left(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u} \right) \right)^n} \quad (\text{B.43})$$

が得られる。

注意 47 ページでもコメントしたように我々の提案手法はパーセプトロン学習に限定するものではない。なぜなら文脈に合わせて $P(y), g(y, \Delta), P(w), \rho(\lambda)$ を変えることで任意の解析が可能であるからだ。すなわち大システム極限での Bayes 推定一般に対して有効であることが分かる。

付 録 C 線形回帰問題の統計力学的解析

ここでは線形回帰問題についての解析を説明する [41, 65]. 線形回帰問題は本編で考察した 2 クラス識別問題に対する多クラスの極限として,

$$y = \frac{\vec{x}^T \vec{w}}{\sqrt{N}} \quad (\text{C.1})$$

のように出力 y が連続値を取る線形パーセプトロンの識別問題に対応し [65], 回帰分析などに広く使われている [41]. 以下では入力行列 $X := \frac{1}{\sqrt{N}} \{\vec{x}_1, \dots, \vec{x}_p\}^T \in \mathcal{M}_{p \times N}$ で定義される自己相関行列 $X^T X \in \mathcal{M}_{N \times N}$ の固有値分布が $\rho(\lambda)$ で特徴づけられるランダム行列アンサンブルに対する, 参照例題 $\xi^p = \{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_p, y_p)\}$ を用いた線形回帰問題を考える. 本付録では議論を簡単にするために, 出力 $\vec{y}$ は入力行列 X が与えられたもとの式 (3.9) で定義される条件付き確率 $P(\vec{y}|X)$ に従うと仮定し, レプリカ解析において, $n = 0$ のとき記憶容量問題に, $n = 1$ のときオラクル推定問題にそれぞれ対応する.

C.1 線形回帰問題のレプリカ解析

この問題のハミルトニアンは誤差の統計性により

$$\mathcal{H}(\vec{w}|\xi^p) := \begin{cases} \sum_{\mu=1}^p \left| y_\mu - \frac{\vec{x}_\mu^T \vec{w}}{\sqrt{N}} \right| & L^1 \text{ ノルム} \\ \frac{1}{2} \sum_{\mu=1}^p \left(y_\mu - \frac{\vec{x}_\mu^T \vec{w}}{\sqrt{N}} \right)^2 & L^2 \text{ ノルム} \end{cases} \quad (\text{C.2})$$

などで与えられる. 離散値と連続値のそれぞれに対して, キュムラント母関数 $\psi(n)$ の結合加重に関係する部分を計算すると,

$$\begin{aligned} & -\frac{n\tilde{\chi}_w}{2} - \frac{n(n-1)q_w\tilde{q}_w}{2} + \log \int_{-\infty}^{\infty} Dz \left(\sum_w P(w) e^{\frac{\tilde{\chi}_w - \tilde{q}_w}{2} w^2 + z w \sqrt{\tilde{q}_w}} \right)^n \\ = & \begin{cases} -\frac{n\tilde{q}_w}{2} - \frac{n(n-1)}{2} q_w \tilde{q}_w + \log \int_{-\infty}^{\infty} Dz \cosh^n(z \sqrt{\tilde{q}_w}) & \text{離散値} \\ \frac{n-1}{2} \log(1 - q_w) + \frac{1}{2} \log(1 - q_w + nq_w) & \text{連続値} \end{cases} \end{aligned} \quad (\text{C.3})$$

となる. 以下では L^1 と L^2 のそれぞれに対してキュムラント母関数の残りの部分の解析を行う.

L^1 ノルム 式 (C.2) から $g(y, u) := e^{-\beta|y-u|}$ であるので,

$$\begin{aligned} & \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dx g\left(y, x \sqrt{\tilde{\chi}_u - \tilde{q}_u} + z \sqrt{\tilde{q}_u}\right) \right)^n \\ = & \frac{\alpha n \beta^2}{2} (\tilde{\chi}_u - \tilde{q}_u) + \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \tilde{H}^n\left(y - z \sqrt{\tilde{q}_u}\right) \end{aligned} \quad (\text{C.4})$$

となる. ただし

$$\tilde{H}\left(y - z \sqrt{\tilde{q}_u}\right) := H\left(\frac{y - z \sqrt{\tilde{q}_u}}{\sqrt{\tilde{\chi}_u - \tilde{q}_u}} - \beta \sqrt{\tilde{\chi}_u - \tilde{q}_u}\right) e^{-\beta(y - z \sqrt{\tilde{q}_u})} + H\left(-\frac{y - z \sqrt{\tilde{q}_u}}{\sqrt{\tilde{\chi}_u - \tilde{q}_u}} - \beta \sqrt{\tilde{\chi}_u - \tilde{q}_u}\right) e^{\beta(y - z \sqrt{\tilde{q}_u})} \quad (\text{C.5})$$

を用いた。これよりキュムラント母関数は

$$\begin{aligned}\psi(n) &= (n-1)F(\chi_w, \chi_u) + F(\chi_w + nq_w, \chi_u + nq_u) + \frac{n\alpha}{2}\tilde{\chi}_u(\chi_u + q_u) + \frac{n(n-1)\alpha}{2}\tilde{q}_u q_u \\ &\quad + \frac{\alpha n\beta^2}{2}(\tilde{\chi}_u - \tilde{q}_u) + \alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \tilde{H}^n(y - z\sqrt{\tilde{q}_u}) \\ &\quad + \begin{cases} -\frac{n\tilde{q}_w}{2} - \frac{n(n-1)}{2}q_w\tilde{q}_w + \log \int_{-\infty}^{\infty} Dz \cosh^n(z\sqrt{\tilde{q}_w}) & \text{離散値} \\ \frac{n-1}{2}\log(1-q_w) + \frac{1}{2}\log(1-q_w + nq_w) & \text{連続値} \end{cases}\end{aligned}\quad (\text{C.6})$$

となる。ただし $\chi_w = 1 - q_w$ であり¹,

$$q_u = -\beta^2 \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \tilde{H}^n(y - z\sqrt{\tilde{q}_u}) \tanh^2 \left(\beta \left(z\sqrt{\tilde{q}_u} - y \right) + \frac{1}{2} \log \frac{H \left(\frac{y - z\sqrt{\tilde{q}_u}}{\sqrt{\tilde{\chi}_u} - \tilde{q}_u} - \beta \sqrt{\tilde{\chi}_u - \tilde{q}_u} \right)}{H \left(-\frac{y - z\sqrt{\tilde{q}_u}}{\sqrt{\tilde{\chi}_u} - \tilde{q}_u} - \beta \sqrt{\tilde{\chi}_u - \tilde{q}_u} \right)} \right)}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \tilde{H}^n(y - z\sqrt{\tilde{q}_u})} \quad (\text{C.7})$$

$$\chi_u = -q_u - \beta^2 - \frac{2\beta}{\sqrt{2\pi(\tilde{\chi}_u - \tilde{q}_u)}} \frac{\sum_y P(y) \int_{-\infty}^{\infty} Dz \tilde{H}^{n-1}(y - z\sqrt{\tilde{q}_u})}{\sum_y P(y) \int_{-\infty}^{\infty} Dz \tilde{H}^n(y - z\sqrt{\tilde{q}_u})} \quad (\text{C.8})$$

が得られる。

L^2 ノルム 式 (C.2) から $g(y, u) := e^{-\frac{\beta}{2}(y-u)^2}$ とおくことで,

$$\begin{aligned}&\alpha \log \sum_y P(y) \int_{-\infty}^{\infty} Dz \left(\int_{-\infty}^{\infty} Dx g(y, x\sqrt{\tilde{\chi}_u - \tilde{q}_u} + z\sqrt{\tilde{q}_u}) \right)^n \\ &= -\frac{\alpha(n-1)}{2} \log(1 + \beta(\tilde{\chi}_u - \tilde{q}_u)) - \frac{\alpha}{2} \log(1 + \beta(\tilde{\chi}_u - \tilde{q}_u) + n\beta\tilde{q}_u) + \alpha \log \sum_y P(y) e^{-\frac{n\beta y^2}{2(1+\beta(\tilde{\chi}_u - \tilde{q}_u) + n\beta\tilde{q}_u)}}\end{aligned}\quad (\text{C.9})$$

となり、キュムラント母関数は

$$\begin{aligned}\psi(n) &= (n-1)F(\chi_w, \chi_u) + F(\chi_w + nq_w, \chi_u + nq_u) + \frac{n\alpha}{2}\tilde{\chi}_u(\chi_u + q_u) + \frac{n(n-1)\alpha}{2}\tilde{q}_u q_u \\ &\quad - \frac{\alpha(n-1)}{2} \log(1 + \beta(\tilde{\chi}_u - \tilde{q}_u)) - \frac{\alpha}{2} \log(1 + \beta(\tilde{\chi}_u - \tilde{q}_u) + n\beta\tilde{q}_u) + \alpha \log \sum_y P(y) e^{-\frac{n\beta y^2}{2(1+\beta(\tilde{\chi}_u - \tilde{q}_u) + n\beta\tilde{q}_u)}} \\ &\quad + \begin{cases} -\frac{n\tilde{q}_w}{2} - \frac{n(n-1)}{2}q_w\tilde{q}_w + \log \int_{-\infty}^{\infty} Dy \cosh^n(z\sqrt{\tilde{q}_w}) & \text{離散値} \\ \frac{n-1}{2}\log(1-q_w) + \frac{1}{2}\log(1-q_w + nq_w) & \text{連続値} \end{cases}\end{aligned}\quad (\text{C.10})$$

となる。ただし $\chi_w = 1 - q_w$ を用いた²。さらに秩序変数は

$$q_u = -\frac{\beta^2 \tilde{q}_u}{(1 + \beta(\tilde{\chi}_u - \tilde{q}_u))(1 + \beta(\tilde{\chi}_u - \tilde{q}_u) + n\beta\tilde{q}_u)} - \frac{\beta^2 \langle y^2 \rangle}{(1 + \beta(\tilde{\chi}_u - \tilde{q}_u) + n\beta\tilde{q}_u)^2} \quad (\text{C.12})$$

$$\chi_u = \frac{\beta}{1 + \beta(\tilde{\chi}_u - \tilde{q}_u)} \quad (\text{C.13})$$

$$\langle y^2 \rangle = \frac{\sum_y P(y) y^2 e^{-\frac{n\beta y^2}{2(1+\beta(\tilde{\chi}_u - \tilde{q}_u) + n\beta\tilde{q}_u)}}}{\sum_y P(y) e^{-\frac{n\beta y^2}{2(1+\beta(\tilde{\chi}_u - \tilde{q}_u) + n\beta\tilde{q}_u)}}} \quad (\text{C.14})$$

¹ $\frac{\partial |\theta|}{\partial \theta} = \text{sgn}(-\theta)$ なので $\frac{\partial}{\partial \theta} e^{-\beta|\theta|} = \beta \text{sgn}(\theta) e^{-\beta|\theta|}$ や $\frac{\partial^2}{\partial \theta^2} e^{-\beta|\theta|} = (\beta^2 + 2\beta\delta(\theta)) e^{-\beta|\theta|}$ が得られる。

$$\chi_u + q_u = \frac{(n-1)\beta^2 \tilde{q}_u + \beta + \beta^2(\tilde{\chi}_u - \tilde{q}_u)}{(1 + \beta(\tilde{\chi}_u - \tilde{q}_u))(1 + \beta(\tilde{\chi}_u - \tilde{q}_u) + n\beta\tilde{q}_u)} - \frac{\beta^2 \langle y^2 \rangle}{(1 + \beta(\tilde{\chi}_u - \tilde{q}_u) + n\beta\tilde{q}_u)^2} \quad (\text{C.11})$$

を満たす.

C.2 線形回帰問題の TAP 法

73 ページの TAP アルゴリズムの中で変更するところだけをここに示す.

L^1 ノルム 式 (C.2) から $g(y, u) = e^{-\beta|y-u|}$ であるので

$$\int_{-\infty}^{\infty} Dxg\left(y_\mu, x\sqrt{\tilde{\chi}_u} + h_{u\mu}\right) = e^{\frac{\beta^2 \tilde{\chi}_u}{2}} \left\{ H\left(\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right) e^{-\beta(y_\mu - h_{u\mu})} + H\left(-\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right) e^{\beta(y_\mu - h_{u\mu})} \right\}$$

となる. これを 73 ページの TAP 法のアルゴリズムに代入することで

$$\begin{aligned} m_{u\mu} &= \beta \frac{H\left(\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right) e^{-\beta(y_\mu - h_{u\mu})} - H\left(-\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right) e^{\beta(y_\mu - h_{u\mu})}}{H\left(\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right) e^{-\beta(y_\mu - h_{u\mu})} + H\left(-\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right) e^{\beta(y_\mu - h_{u\mu})}} \\ &= \beta \tanh \left(\beta(h_{u\mu} - y_\mu) + \frac{1}{2} \log \frac{H\left(\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right)}{H\left(-\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right)} \right) \end{aligned} \quad (C.15)$$

$$\begin{aligned} \delta_{u\mu} &= \beta^2 + \frac{2\beta}{\sqrt{2\pi\tilde{\chi}_u}} \left(H\left(\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right) e^{\frac{1}{2}\left(\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right)^2} \right. \\ &\quad \left. + H\left(-\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right) e^{\frac{1}{2}\left(-\frac{y_\mu - h_{u\mu}}{\sqrt{\tilde{\chi}_u}} - \beta\sqrt{\tilde{\chi}_u}\right)^2} \right)^{-1} \end{aligned} \quad (C.16)$$

が得られる.

L^2 ノルム 同様に式 (C.2) から $g(y, u) = e^{-\frac{\beta}{2}(y-u)^2}$ であるので

$$\int_{-\infty}^{\infty} Dxg\left(y_\mu, x\sqrt{\tilde{\chi}_u} + h_{u\mu}\right) = \frac{e^{-\frac{\beta}{2(1+\beta\tilde{\chi}_u)}(y_\mu - h_{u\mu})^2}}{\sqrt{1+\beta\tilde{\chi}_u}} \quad (C.17)$$

となり

$$m_{u\mu} = \chi_u(y_\mu - h_{u\mu}), \quad \chi_u = \frac{\beta}{1 + \beta\tilde{\chi}_u} \quad (C.18)$$

が得られる.

C.3 拘束条件付き線形回帰問題

特に結合加重 $\vec{w}$ が連続値でかつ拘束条件 $\vec{w}^T \vec{w} = N$ がある場合は, Lagrange 未定乗数法を用いて, ハミルトニアンを最小にする結合加重を決定することもできる. Lagrangian は

$$\mathcal{L}(\vec{w}) := \frac{\tilde{\chi}}{2} (\vec{w}^T \vec{w} - N) + \mathcal{H}(\vec{w}|\xi^p) \quad (C.19)$$

で定義される. ここでは L^2 ノルムの場合について議論する. 式 (C.19) より

$$\frac{\partial \mathcal{L}(\vec{w})}{\partial \vec{w}} = \tilde{\chi} \vec{w} + X^T X \vec{w} - X^T \vec{y} = 0, \quad \implies \quad \vec{w} = (\tilde{\chi} + X^T X)^{-1} X^T \vec{y} \quad (C.20)$$

が得られる. さらに $\tilde{\chi}$ は拘束条件を満たすように決める. もしくは十分小さな学習率 $\varepsilon, \varepsilon' > 0$ を用いて最急降下法 (steepest descent method) で求めることができる.

$$\vec{w}_{t+1} = \vec{w}_t - \varepsilon ((\tilde{\chi}_t + X^T X) \vec{w}_t - X^T \vec{y}) \quad (C.21)$$

$$\tilde{\chi}_{t+1} = \tilde{\chi}_t - \varepsilon' (\vec{w}_t^T \vec{w}_t - N) \quad (C.22)$$

または式 (D.15) のアルゴリズムを用いて逆行列を解かずに結合加重 $\vec{w}$ を求めることもできる.

付 録 D 求解アルゴリズム

レプリカ法や TAP 法の数値実験を行う際に、反復法（もしくは逐次代入法）を用いて解を求める際、更新手順によって振動する場合があります、アルゴリズムの設計に注意が必要である。ここでは、更新手順からくる振動現象を抑えつつ効率的に数値結果を求めることのできる、いくつかの求解アルゴリズムを紹介する。

D.1 逐次過緩和法

以下の方程式を満たす x を求めることを考える。

$$x = f(x) \quad (\text{D.1})$$

この方程式を満たす x を x^* とする。単純に考えれば、更新時刻 t の状態を x_t で表し、適当な初期値 x_0 から、

$$x_{t+1} = f(x_t) \quad (\text{D.2})$$

によって更新することで $x^* (= \lim_{t \rightarrow \infty} x_t)$ が求まることが期待できる。しかしながらうまく調整しないと式 (D.2) は振動する場合があります、反復法を力学系と見立てたアルゴリズムの「安定性解析」をする必要がある。だが安定性解析の一般論を展開するには本学位論文の程度を超えているため¹、ここでは有効な解決策の 1 つであり、比較的振動を抑えつつ効率よく求めることのできる「逐次過緩和法 (Successive Over Relaxation method; SOR 法)」について説明する。

D.1.1 SOR 法

式 (D.2) の代わりに変数 $\varepsilon \neq 0$ を用いて次の更新式

$$x_{t+1} = \varepsilon f(x_t) + (1 - \varepsilon)x_t \quad (\text{D.3})$$

によって解を求めることを考える²。この方法を SOR 法と呼ぶことにする。 $x = x^*$ のとき式 (D.3) を満たすため、式 (D.2) 同様、式 (D.1) の解法になっていることが分かる。また $\varepsilon = 1$ のとき式 (D.2) を含むことも分かる。直感的には ε が 0 に近いとき、少しずつ状態を更新するため（慣性効果）、発散や振動するような場合に比べて、収束性が保証できることが期待される。さてこの更新式に対して収束性を議論しよう。そのためにまず関数列 $|x_{t+1} - x_t|$ の漸化式を評価する。

$$\begin{aligned} |x_{t+2} - x_{t+1}| &= |\varepsilon| |f(x_{t+1}) - x_{t+1}| = |\varepsilon| |f(x_t) + \varepsilon(f(x_t) - x_t)f'(x_t + \delta) - x_{t+1}| \\ &\leq |\varepsilon| |f(x_t) - x_t| |1 + \varepsilon f'(x_t + \delta)| + |\varepsilon| |x_t - x_{t+1}| = |x_{t+1} - x_t| (|\varepsilon| + |1 + \varepsilon f'(x_t + \delta)|) \end{aligned} \quad (\text{D.5})$$

ただし三角不等式 $|a + b| \leq |a| + |b|$ と $f(x_{t+1})$ を x_t の周りで $\varepsilon(f(x_t) - x_t)$ を用いて Taylor 展開し、1 次までの展開を用いた。これより単調非増加関数列 $|x_{t+1} - x_t|$ にするためには、 ε の満たす十分条件として $|\varepsilon| + |1 + \varepsilon f'| \leq 1$ から

$$f' < -1, \quad \Rightarrow \quad 0 \leq \varepsilon \leq \varepsilon^* = \frac{2}{1 - f'} \quad (\text{D.6})$$

$$f' > 1, \quad \Rightarrow \quad 0 \geq \varepsilon \geq \varepsilon^* = \frac{-2}{1 + f'} \quad (\text{D.7})$$

¹たとえば $f(x) = \tanh(-2x)$ の場合 $x^* = 0$ であるが素朴にやると振動する。

²また振動の周期が 2 の場合は、式 (D.3) の他に

$$x_{t+1} = pf(x_t) + qx_t + rx_{t-1} \quad (\text{D.4})$$

などの改良も考えることができる。ただし $p + q + r = 1, p, q, r > 0$ とした。

が得られる。ただし $f' := f'(x_t + \delta)$ とおいた。さらに $f' < -1$ のとき $0 \leq \varepsilon \leq \varepsilon^*$ の範囲で $|\varepsilon| + |1 + \varepsilon f'|$ を最小にする ε は $\varepsilon = -\frac{1}{f'}$ のときであり、 $f' > 1$ のとき $0 \geq \varepsilon \geq \varepsilon^*$ の範囲で $|\varepsilon| + |1 + \varepsilon f'|$ を最小にする ε は $\varepsilon = -\frac{1}{f'}$ のときである³。また $|f'| < 1$ の場合は Taylor 展開を 2 次まで行い、 $\left|1 + \varepsilon f'(x_t) + \frac{\varepsilon^2}{2}(f(x_t) - x_t)f''(x_t + \delta)\right| + |\varepsilon| \leq 1$ を満たす ε の範囲で左辺を最小にする ε^* を求めればよいが、同様の議論のため省略する。またここでは議論を簡単にするために x は 1 変数としていたが、容易に多変数に対しても拡張できることに注意。

D.1.2 応用例

SOR 法の有効性を検証するために関数 $f(x) := \tanh(\beta x)$ としたときの解起動を考察する。任意の β に対して自明な解 $x^* = 0$ が知られており、さらに $\beta > 1$ のとき $x \neq 0$ となる解が存在する。そこで $\beta = \pm 2$ のそれぞれに対して式 (D.3) の数値実験を行ってみた。 $\beta = -2$ のとき $x^* = 0$ であるが、式 (D.3) において ε を 0.1 間隔にして検証実験を行った。これにより $0 < \varepsilon \leq 0.6$ の範囲で $x^* = 0$ に収束していることが分かる。だが $0.7 \leq \varepsilon$ において振動していることが分かる。一方 $\beta = 2$ においては $0 < \varepsilon$ で $x^* \neq 0$ の解に収束していることが分かる。

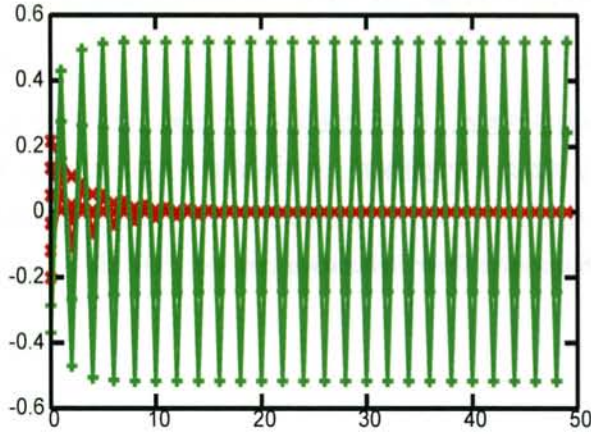


図 D.1: $\beta = -2$ のときの解軌道。横軸を t とし、縦軸は x_t を表す。 ε を 0.1 ずつでプロットした。 $0 < \varepsilon \leq 0.6$ で $x^* = 0$ の解を求められている。しかしながら $0.7 \leq \varepsilon$ では振動していることが分かる。

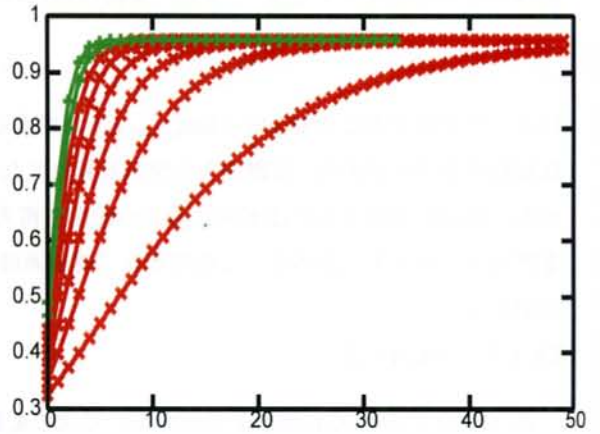


図 D.2: $\beta = 2$ のときの解軌道。横軸を t とし、縦軸は x_t を表す。 $\varepsilon = 0.1$ ずつでプロットした。 $0 < \varepsilon$ で $x^* \neq 0$ の解を求められている。グラフから ε が小さいと収束が遅いことが見て取れる。

本稿のレプリカ法や TAP 法において、システムを特徴づける秩序変数やマイクロ変数は、上記の式 (D.1) のような多変数の連立方程式により決定される。式 (D.2) のように素朴な方法で行うと、ある領域においては収束が遅かったり解が振動してしまい効率的でなくなることが、数値実験を行うことで経験的に分かった。そこで我々はレプリカ法や TAP 法のそれぞれの数値実験において、式 (D.3) と最適な ε を用いて実験を行うことで所望な結果を得ることができた。

D.2 Newton 法

この節では Newton 法を説明する。

D.2.1 1 変数の場合の Newton 法

はじめに $g(x) = 0$ を満たす x を求めることを考える。SOR 法と同様に反復法により更新時刻 t の状態を x_t とし、 $(x_t, g(x_t))$ と傾き $g'(x_t)$ を用いて効率的に最適解を求める x_t の更新式を導出する。そこで $g(x_{t+1})$ を x_t の周りで $x_{t+1} - x_t$ を用いて 1 次までの Taylor 展開を行うと、

$$g(x_{t+1}) = g(x_t + x_{t+1} - x_t) \simeq g(x_t) + (x_{t+1} - x_t)g'(x_t) \quad (\text{D.8})$$

³ $|f'| > 1$ のとき、 $\varepsilon^* = -\frac{1}{f'}$ が最適である。

となる。ここで右辺が 0 になるように x_{t+1} を決定すると、結果的に $t \rightarrow \infty$ で $g(x_t) \rightarrow 0$ と $|x_{t+1} - x_t| \rightarrow 0$ になるため、 x^* が求まることが分かる。よって更新式は、

$$x_{t+1} = x_t - \frac{g(x_t)}{g'(x_t)} \quad (\text{D.9})$$

となる。ちなみに $g(x) = f(x) - x$ とおくことで式 (D.1) の解析も行いうことが出来る。さらに関数 $e(x)$ の極値を求める問題の場合、 $g(x) = e'(x)$ とおくことで、つまり

$$x_{t+1} = x_t - \frac{e'(x_t)}{e''(x_t)} \quad (\text{D.10})$$

によって更新することで極値（極大極小の議論はしない）を求めることができる。

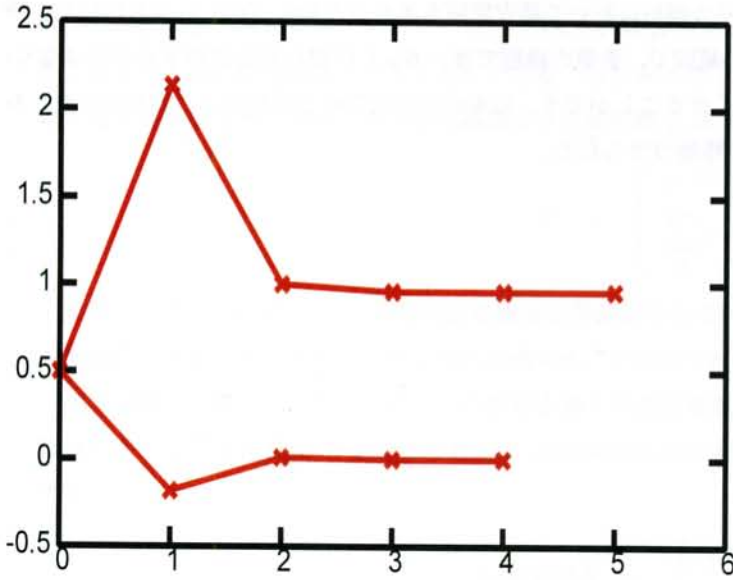


図 D.3: $\beta = \pm 2$ の場合の $x = \tanh(\beta x)$ の固定点を Newton 法 ($g(x) = \tanh(\beta x) - x$) を用いて求めた結果。横軸を反復時刻 t とし、縦軸は x_t を表す。下の曲線が $\beta = -2$ であり、上の曲線が $\beta = 2$ のときである。SOR 法に比べて収束が早いだが、不安定である。また数値実験により $\beta = 2$ のとき、 $x_0 = 0.3$ にすると $x^* = 0$ の解に収束する。つまり初期値によって安定固定点にも不安定固定点にも収束する可能性がある。

D.2.2 多変数の場合の Newton 法

x が多変数の場合に拡張する。 d 次元 $\vec{x} := \{x_1, x_2, \dots, x_d\}^T \in \mathbf{R}^d$ の関数 $g(\vec{x}) = 0$ を扱い、式 (D.8) と同様に $g(\vec{x}_{t+1})$ を $\vec{x}_t$ の周りで Taylor 展開すると、

$$g(\vec{x}_{t+1}) = g(\vec{x}_t + \vec{x}_{t+1} - \vec{x}_t) \simeq g(\vec{x}_t) + (\vec{x}_{t+1} - \vec{x}_t)^T \frac{\partial g(\vec{x}_t)}{\partial \vec{x}_t} \quad (\text{D.11})$$

となり、右辺が 0 になるように更新則を立てると、

$$\vec{x}_{t+1} = \vec{x}_t - \left\{ \frac{g(\vec{x}_t)}{\frac{\partial g(\vec{x}_t)}{\partial x_{t1}}}, \frac{g(\vec{x}_t)}{\frac{\partial g(\vec{x}_t)}{\partial x_{t2}}}, \dots, \frac{g(\vec{x}_t)}{\frac{\partial g(\vec{x}_t)}{\partial x_{td}}} \right\}^T \quad (\text{D.12})$$

が得られる。さらに関数 $e(\vec{x})$ の極値を求める場合、 $g(\vec{x}) = \frac{\partial e(\vec{x})}{\partial \vec{x}}$ とおくことで、すなわち

$$\frac{\partial e(\vec{x})}{\partial \vec{x}} \Big|_{\vec{x}=\vec{x}_{t+1}} \simeq \frac{\partial e(\vec{x})}{\partial \vec{x}} \Big|_{\vec{x}=\vec{x}_t} + \frac{\partial^2 e(\vec{x})}{\partial \vec{x} \partial \vec{x}} \Big|_{\vec{x}=\vec{x}_t} (\vec{x}_{t+1} - \vec{x}_t) \quad (\text{D.13})$$

から更新式

$$\vec{x}_{t+1} = \vec{x}_t - H^{-1} \frac{\partial e(\vec{x})}{\partial \vec{x}} \Big|_{\vec{x}=\vec{x}_t}, \quad H := \{H_{ij}\} \in \mathcal{M}_{d \times d}, \quad H_{ij} := \frac{\partial^2 e(\vec{x})}{\partial x_i \partial x_j} \Big|_{\vec{x}=\vec{x}_t} \quad (\text{D.14})$$

が得られる。ただし H は Hesse 行列と呼ばれており、逆行列 H^{-1} を解析的に評価するには行列のサイズ d に対していくつかの特殊なケースを除いて $O(d^3)$ の計算量が必要である。そこで次の方法を用いて効率的に $H^{-1} \frac{\partial e(\vec{x})}{\partial \vec{x}} \Big|_{\vec{x}=\vec{x}_t}$ を評価する方法を紹介する。

2つの d 次元ベクトル $\vec{a} = H^{-1} \frac{\partial e(\vec{x})}{\partial \vec{x}} \Big|_{\vec{x}=\vec{x}_t}$ と $\vec{b} = \frac{\partial e(\vec{x})}{\partial \vec{x}} \Big|_{\vec{x}=\vec{x}_t}$ とおき、行列 H の対角部分と非対角部分をそれぞれ Λ と $\Sigma = H - \Lambda$ で表すことにする。ここでは議論を簡単にするために $H_{ii} = \Lambda_{ii} \neq 0$ を仮定する⁴。これより $H^{-1} \frac{\partial e(\vec{x})}{\partial \vec{x}} \Big|_{\vec{x}=\vec{x}_t}$ を求める問題は $\vec{b} = H\vec{a}$ を満たす $\vec{a}$ を求める問題に帰着される。これより更新式

$$\vec{b} = H\vec{a} = \Lambda\vec{a} + \Sigma\vec{a}, \quad \implies \quad \Lambda\vec{a} = \vec{b} - \Sigma\vec{a}, \quad \implies \quad \vec{a}_{t+1} = \Lambda^{-1}\vec{b} - \Lambda^{-1}\Sigma\vec{a}_t \quad (\text{D.15})$$

が得られる。 $\Lambda, \Sigma, \vec{b}$ を与えられることで式 (D.15) の更新式により最適な $\vec{a}$ を求めることができる。

本稿では教師推定問題（離散値の場合）のレプリカ解析において準安定解を求めるために Newton 法を用いた。特に⁵RS 解の局所安定性を議論する際に $\alpha < \alpha_{\text{AT}}$ の範囲で、更新の過程で逐一 6×6 の逆行列を計算するのではなく、式 (D.15) の更新則を用いることで計算量を少なくすることができ、効率的に準安定解を導出することができた。ちなみに α_{AT} は RS 解の局所安定性が破れることで特徴づけられる。

⁴もし $H_{ii} = 0$ の場合は適当な $\delta \neq 0$ を用いて、 $\Lambda_{ii} = \delta$ と $\Sigma_{ii} = -\delta$ とおくことで上記の議論が可能である。

⁵i.i.d. 例題の場合 $\alpha_{\text{AT}} \simeq 1.492$ 、直交例題の場合 $\alpha_{\text{AT}} \simeq 1.305$ である。

付録E 大偏差原理とLegendre変換

本付録ではレプリカ解析やThouless-Anderson-Palmer法に対して、数理的なサポートを与える大偏差原理とLegendre変換について解説する。

E.1 大偏差統計もしくは鞍点法

ここではスピングラス理論や学習理論に限定せず、かつ議論を簡単にするために、正準集団の表現を用いて、相転移現象を説明する大偏差原理、もしくは鞍点法を考察する。一般的に逆温度 β と微視的状态 $\vec{S}$ のハミルトニアン $\mathcal{H}(\vec{S}) \in \mathbf{R}$ が与えられ、システムサイズ N が大きい場合の1自由度あたりの自由エネルギー $f = F/N$ の「非解析性」によって相転移を特徴付けることができる。ちなみに F と Z はそれぞれシステム全体の自由エネルギーとそれに対応する分配関数を表す。第1章の議論より分配関数 Z は次のような2種類の違う表現で表すことができる。

$$Z = \begin{cases} e^{-N\beta f}, \\ \sum_{\vec{S}} e^{-\beta\mathcal{H}(\vec{S})}. \end{cases} \quad (\text{E.1})$$

十分大きい N の状況で、もし分配関数を厳密な評価ができない場合に、我々は分配関数の評価に最も寄与する部分分配関数（後述する）を用いて漸近的に評価することが許される。

さて議論を見通しよくするために次の性質を満たす秩序変数（order parameter）を導入する。ある微視的状态 $\vec{S}$ に対して1価関数 $\theta(\vec{S})$ が定義できたとき¹、次の条件を満たす新しい関数、インジケータ関数 $I(\theta, \theta(\vec{S}))$ を用意する。

$$I(\theta, \theta(\vec{S})) = \begin{cases} 1 & \theta = \theta(\vec{S}) \\ 0 & \text{otherwise.} \end{cases} \quad (\text{E.3})$$

インジケータ関数の最もよく使われる例の1つとしてクロネッカーのデルタ（Kronecker's delta）が挙げられる。さらにある $\vec{S}$ に対して式（E.3）を用いて、 $\theta(\vec{S})$ が1価関数であるため恒等式

$$\sum_{\theta} I(\theta, \theta(\vec{S})) = 1, \quad (\text{E.4})$$

が成り立つ。ただし $\sum_{\theta}$ は θ の取り得るすべての可能性²に対して積分（もしくは和）を実行することを表す。これより式（E.3）を用いて、分配関数を次のように分解することができる。

$$Z = \sum_{\vec{S}} \sum_{\theta} I(\theta, \theta(\vec{S})) e^{-\beta\mathcal{H}(\vec{S})} = \sum_{\theta} \left(\sum_{\vec{S}} I(\theta, \theta(\vec{S})) e^{-\beta\mathcal{H}(\vec{S})} \right) = \sum_{\theta} e^{-N\beta f(\theta)}, \quad (\text{E.5})$$

ただし $\theta = \theta(\vec{S})$ を満たす微視的状态 $\vec{S}$ 全体で定義される部分自由エネルギー $f(\theta) \in \mathbf{R}$ は

$$e^{-N\beta f(\theta)} := \sum_{\vec{S}} I(\theta, \theta(\vec{S})) e^{-\beta\mathcal{H}(\vec{S})}. \quad (\text{E.6})$$

¹ $\theta(\vec{S})$ をたとえばイジングスピン $\vec{S} \in \{\pm 1\}^N$ の場合

$$\theta(\vec{S}) := \frac{1}{N} \sum_{k=1}^N S_k, \quad (\text{E.2})$$

など、扱う問題に応じて $\theta(\vec{S})$ の関数形を適宜決める必要がある。

²たとえば式（E.2）の場合 $-1 \leq \theta \leq 1$ となる。

で定義される。また定義から任意の $f(\theta)$ に対して $e^{-N\beta f} \geq e^{-N\beta f(\theta)}$ のため、必ず $f \leq f(\theta)$ が成り立つことが知られている。そして十分大きな N に対して f の推定値は式 (E.6) を用いて

$$f = \min_{\theta} f(\theta). \quad (\text{E.7})$$

となると期待される。なぜなら $f(\theta)$ の最小値を満たす θ の個数がシステムサイズ N の指数関数に比例していない場合、式 (E.1) の分配関数は次のように議論できる。

$$e^{-N\beta f} = e^{-N\beta f(\theta)} \left(1 + \sum_{\theta' (\neq \theta)} e^{-N\beta (f(\theta') - f(\theta))} \right), \quad (\text{E.8})$$

ただし式 (E.8) の括弧の中に含まれる 1 は形式的な 1 であることに注意。もし θ の定義域が離散集合の場合、直感的に式 (E.8) が成り立っていることは明らか。一方 θ の定義域が連続集合の場合、単位”1”を定義することが不可能であるが、積分の微小区間を”1”と形式的に見なしても N が十分大きくなる極限で、以下の議論に影響を与えることはない³。よって N が無限大の極限で式 (E.8) に含まれる括弧は、

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \left(1 + \sum_{\theta' (\neq \theta)} e^{-N\beta (f(\theta') - f(\theta))} \right) = \begin{cases} 0 & f(\theta) = \min_{\tau} f(\tau) \\ \infty & \text{otherwise,} \end{cases} \quad (\text{E.9})$$

となり、式 (E.7) が証明された。ここで 5 点重要なコメントをする。

1. 今までの議論で θ は 1 変数とは限らず、多変数でもかまわない。なぜなら上記の議論は修正なしに多変数に拡張することが可能であるからだ。
2. θ の定義域が連続集合で十分な広さが確保されている場合、次の条件

$$\frac{\partial f(\theta)}{\partial \theta} = 0, \quad (\text{E.10})$$

を用いて部分自由エネルギー $f(\theta)$ を最小にする最適な θ の候補を決定することができる。しかしながら不幸なことに θ の定義域が有界である場合⁴、 θ の定義域の縁 (edge) にも強く注意を払う必要がある。なぜなら簡単のために、たとえば 1 変数関数 $f(x) = (x - a)^2$ を用意し、 x の定義域を $-1 \leq x \leq 1$ としたとき、変数 $a \in \mathbf{R}$ に応じて

$$\min_{-1 \leq x \leq 1} f(x) = \begin{cases} (1 + a)^2 & a < -1 \\ 0 & -1 \leq a \leq 1 \\ (1 - a)^2 & \text{otherwise,} \end{cases} \quad (\text{E.11})$$

となることが分かる。当たり前だが $|a| > 1$ の場合、この単純な問題の最小値を満たす最適解 x は式 (E.10) の必要条件を満たさない。我々が自由度複雑系で扱う問題が上記のおもちゃ問題と同様の簡単な構造をしているわけではないが、大雑把に言えば、レプリカ法や TAP 法で用いた記号 Extr とは、凸最適化の文脈として解釈されるため、定義域の縁から非自明な解が導かれる可能性があることに注意を払う必要がある。さらに θ の定義域が離散集合の場合、式 (E.10) を用いて直接最適解を評価することは難しい。なぜなら部分自由エネルギー $f(\theta)$ の最小化を解くことは NP 困難な問題に含まれる「組み合わせ最適化」の問題と見なされるからだ。紙面の都合上、また学位論文の目的として最適化問題の根幹に関心があるわけではないためこれ以上踏み込まない。もし最適化が気になる方は [29] などを参照されたい。

³ちなみに断りなく使ってきたが、式 (E.3) と式 (E.4) の”1”も形式的な”1”であることが分かる。

⁴たとえば $-1 \leq \theta \leq 1$ など。

3. $f(\theta)$ の最小値を表す θ の個数がシステムサイズ N の指数関数 $e^{N\Sigma}$, ($\Sigma \neq 0$) に比例しているとき（上記の議論ではシステムサイズの指数関数に比例しないと仮定した），一般的に式 (E.9) の上の条件が満たさない．しかしながら今までの議論を再帰的に用いることで，分配関数を次のように表すことができる．

$$Z = \sum_{\theta} e^{-N\beta f(\theta)} = \sum_{f'} e^{-N\beta f' + N\Sigma(f')}, \quad (\text{E.12})$$

$$1 = \sum_{f'} I(f', f(\theta)), \quad (\text{E.13})$$

$$I(f', f(\theta)) := \begin{cases} 1 & f' = f(\theta) \\ 0 & \text{otherwise,} \end{cases} \quad (\text{E.14})$$

$$e^{N\Sigma(f')} := \sum_{\theta} I(f', f(\theta)), \quad (\text{E.15})$$

ちなみに新しいインジケータ関数 $I(f', f(\theta))$ はすでに式 (E.3) で定義されたものと同様のセンスで用いた．また $\Sigma(f')$ は complexity と呼ばれるもので，文脈によっては準最適な解の情報を含む重要な関数である．これより我々が扱うシステムの 1 自由度あたりの自由エネルギーは

$$f = \min_{f'} \left(f' - \frac{1}{\beta} \Sigma(f') \right). \quad (\text{E.16})$$

のように求まる．幸いなことに本学位論文で扱う解析において complexity が顔を出さないため，これ以上は言及しない．

4. 微視的状態 $\vec{S}$ のある物理量 $\mathcal{A}(\vec{S})$ に対しての平均値 $\langle \mathcal{A} \rangle$ を考える．特に上記の変数変換により， $\theta = \theta(\vec{S})$ で特徴付けられる物理量 $\mathcal{A}(\theta)$ とその確率 $P(\theta)$ をそれぞれ

$$\mathcal{A}(\theta) := \sum_{\vec{S}} \mathcal{A}(\vec{S}) \times \frac{I(\theta, \theta(\vec{S})) e^{-\beta \mathcal{H}(\vec{S})}}{\sum_{\vec{S}} I(\theta, \theta(\vec{S})) e^{-\beta \mathcal{H}(\vec{S})}} \quad (\text{E.17})$$

$$P(\theta) := \frac{e^{-N\beta f(\theta)}}{\sum_{\theta} e^{-N\beta f(\theta)}} \quad (\text{E.18})$$

のように定義すると，任意の N において Boltzmann 分布 $P(\vec{S})$ を用いて平均を行った $\langle \mathcal{A} \rangle$ と秩序変数 θ での平均で

$$\langle \mathcal{A} \rangle := \sum_{\vec{S}} \mathcal{A}(\vec{S}) P(\vec{S}) = \sum_{\theta} \mathcal{A}(\theta) P(\theta) \quad (\text{E.19})$$

の関係が成り立つ．また N の大きい極限において，部分自由エネルギー $f(\theta)$ の最小値を示す θ が複数存在する場合，それらの集合を $\Theta_{f_{\min}}$ と定義すれば，一般に $\forall \theta, \theta' \in \Theta_{f_{\min}}$ において $P(\theta) = P(\theta')$ とは限らないが，指数の意味で

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log P(\theta) \rightarrow \begin{cases} = 0 & \theta \in \Theta_{f_{\min}} \\ < 0 & \text{otherwise} \end{cases} \quad (\text{E.20})$$

が成り立つため $W(\theta) := \lim_{N \rightarrow \infty} P(\theta)$ とおくと大システム極限での物理量 $\langle \mathcal{A} \rangle$ の推定値は $\sum_{\theta \in \Theta_{f_{\min}}} \mathcal{A}(\theta) W(\theta)$ によって表される．この問題では混合比 $W(\theta)$ に関する Fisher 情報行列が縮退しているため，最尤法を用いて $W(\theta)$ を決定することは困難である．そこでいくつかの解決策が考えられるが，自由エネルギーを最小にする秩序変数の集合の個数が $\#\Theta_{f_{\min}} \sim e^{N\Sigma}$ の場合は complexity を用いて再帰的に解消されるし，指数関数的でなければ $\theta(\vec{S})$ の定義自体を工夫することで解決される場合もある．たとえばゼロ磁場の無限レンジ強磁性体

イジング模型（伏見-テンパリー模型とも呼ばれる）

$$\mathcal{H}(\vec{S}) := -\frac{J}{N} \sum_{i=1}^N \sum_{j=i+1}^N S_i S_j \quad (\text{E.21})$$

において $\theta = \frac{1}{N} \sum_{k=1}^N S_k$ のように定義すれば転移点 $T_C = J$ 以下では自由エネルギーの展望が単峰ではなくなるが、 $\theta = \left| \frac{1}{N} \sum_{k=1}^N S_k \right|$ とおけば単峰になるため、混合分布の議論を避けることができる。モンテカルロ法などの数値実験においても $\theta(\vec{S})$ の決め方はクリティカルに利くことを付け加えておく。

5. 式 (E.6) から

$$f(\theta) = -\frac{1}{N\beta} \log \sum_{\vec{S}} I(\theta, \theta(\vec{S})) e^{-\beta \mathcal{H}(\vec{S})} \quad (\text{E.22})$$

は β に依存した関数であるため、一般的にシステムの自由エネルギー f を近似する部分自由エネルギー $f(\theta)$ の最小値を満たす θ が β に応じて変化することが直感的にわかる。これより θ を秩序変数と名づけ、 β の変化によって部分自由エネルギーの最小値を示す θ が変化することを「相転移」と一般的に呼ばれている。

以上の議論は「大偏差原理⁵⁾」により数学的にサポートされているが、重要なこととして「相転移」はハミルトニアンに含まれているものではなく、秩序変数の選び方とそれに付随する極限操作に依存する概念であるということである。より詳しい議論は [56] を参照されたい。

E.2 Legendre 変換

Legendre 変換は凸解析の分野で重要な概念であるが、レプリカ法や TAP 法の解析に対しても重要な役割を果たす。

Legendre 変換の定義

ある変数 y に対して、変数 x の凸関数 $\psi(x) \in \mathbf{R}$ （ただし x, y の次元は一致し、各々凸集合に含まれると仮定する。ただし凸集合は共通である必要はない）を用いて、次の Legendre 変換が与えられる。

$$\phi(y) := \sup_x \{xy - \psi(x)\}, \quad (\text{E.23})$$

ただし xy は x が 1 次元の場合は積を、他次元の場合は内積をそれぞれ表し、また xy と $\psi(x)$ は実数値を取りうる。自明なこととして $\phi(y) \in \mathbf{R}$ は変数 y の凸関数である⁶⁾。

これに対して、 x と y の凸集合がそれぞれ十分広い場合、次で定義される逆 Legendre 変換を用いて戻すことができる⁷⁾。

$$\psi(x) = \sup_y \{xy - \phi(y)\}, \quad (\text{E.26})$$

TAP 法の導出でも議論したように、不幸なことに $\psi(x)$ の解析的な評価が困難であり、かつ幸運なことに $\phi(y)$ は解析的に評価することができる場合、Legendre 変換と逆 Legendre 変換（の双対性）を用いることで $\psi(x)$ を評価することができる。

⁵⁾特殊な操作として「鞍点法」と呼ぶこともあるが、大偏差解析には変わりない。

⁶⁾式 (E.23) を用いて $\phi(y) \geq xy - \psi(x)$ が成り立つ。この不等式は Fenchel の不等式、もしくは Young の不等式と呼ばれており、この不等式を用いることで $\phi(y)$ の凸関数の証明が簡単に行える。さて任意の y_1, y_2 と任意の x に対して、また変数 $\forall \lambda \in [0, 1]$ を用いて

$$\lambda \phi(y_1) + (1 - \lambda) \phi(y_2) \geq \lambda (xy_1 - \psi(x)) + (1 - \lambda) (xy_2 - \psi(x)) = (\lambda y_1 + (1 - \lambda) y_2) x - \psi(x), \quad (\text{E.24})$$

が成り立つ。右辺は任意の x に対して成り立つため、

$$\lambda \phi(y_1) + (1 - \lambda) \phi(y_2) \geq \sup_x \{(\lambda y_1 + (1 - \lambda) y_2) x - \psi(x)\} = \phi(\lambda y_1 + (1 - \lambda) y_2). \quad (\text{E.25})$$

より $\phi(y)$ が凸関数であることが証明された。

⁷⁾Fenchel の不等式を用いて証明することができる [197]。

さて先ほど自由エネルギー f の評価に用いた部分自由エネルギー $f(\theta)$ は $\theta = \theta(\vec{S})$ の条件を満たす $\vec{S}$ についてすべて足し合わせる必要があるが、いくつかの特殊な場合を除いて、依然として $f(\theta)$ の評価においても計算量的な困難を避けられているとは言い難い。そこで別の改良を提案する。それは式 (E.3) で定義されたインジケータ関数をシステムサイズが十分大きな極限で解析関数で近似することである。たとえば $\tilde{\theta}$ の指数関数を用いて

$$I(\theta, \theta(\vec{S})) \propto \text{Extr}_{\tilde{\theta}} e^{N\tilde{\theta}(\theta(\vec{S})-\theta)} \quad (\text{E.27})$$

が成り立つと仮定する⁸。ここで N が十分大きい場合、分配関数は

$$\begin{aligned} Z &= \sum_{\theta} \sum_{\vec{S}} I(\theta, \theta(\vec{S})) e^{-\beta \mathcal{H}(\vec{S})} \simeq \sum_{\theta} \text{Extr}_{\tilde{\theta}} \sum_{\vec{S}} e^{N\tilde{\theta}(\theta(\vec{S})-\theta)-\beta \mathcal{H}(\vec{S})} \\ &\simeq \text{Extr}_{\theta, \tilde{\theta}} e^{-N\tilde{\theta}\theta} \left(\sum_{\vec{S}} e^{-\beta \mathcal{H}(\vec{S})+N\tilde{\theta}\theta(\vec{S})} \right) = \text{Extr}_{\tilde{\theta}} \left(\text{Extr}_{\theta} e^{-N\tilde{\theta}\theta - N\beta g(\tilde{\theta})} \right) \end{aligned} \quad (\text{E.29})$$

のようになる。ただし新しい関数を

$$g(\tilde{\theta}) := -\frac{1}{N\beta} \log \sum_{\vec{S}} e^{-\beta \mathcal{H}(\vec{S})+N\tilde{\theta}\theta(\vec{S})}, \quad (\text{E.30})$$

のように定義し、正準集団の場合 Gibbs 自由エネルギーを表すことに注意。この定義から、Gibbs 自由エネルギー $g(\tilde{\theta})$ は $\frac{\partial^2 g(\tilde{\theta})}{\partial \tilde{\theta}^2} \leq 0$ を満たすため、 $\tilde{\theta}$ に対して凸関数であることが分かる。よって $g(\tilde{\theta})$ の評価が $f(\theta)$ の評価に比べて容易に実行できる場合、Helmholtz 自由エネルギーは

$$\max_{\tilde{\theta}} \left\{ -\tilde{\theta}\theta - \beta g(\tilde{\theta}) \right\} = : -\beta f(\theta). \quad (\text{E.31})$$

により導出することができ、結果的に

$$f = -\max_{\theta} \{-f(\theta)\} = \min_{\theta} f(\theta), \quad (\text{E.32})$$

が導かれる。ここで2点コメントをする。式 (E.29) 中の操作 $\sum_{\theta} \text{Extr}_{\tilde{\theta}} \rightarrow \text{Extr}_{\theta, \tilde{\theta}}$ の正当性は、全体の自由エネルギーを部分自由エネルギーで近似するという大偏差原理の枠組みで説明できるため、数理的な保証が得られる。さらに式 (E.27) から見て分かるように、共役な秩序変数 $\tilde{\theta}$ が自然な形として要請される。つまり本編で議論されたレプリカ法や TAP 法で用いた変数は上記の操作を念頭において定義されたものであることに注意。

⁸ただし記号 $\simeq$ は

$$a(N) \simeq b(N) \rightarrow \lim_{N \rightarrow \infty} \frac{a(N)}{b(N)} = 1, \quad (\text{E.28})$$

を満たすときに用いる。

関連図書

- [1] 阿部龍蔵, 統計力学, 東京大学出版会, (1992), 阿部龍蔵, 熱統計力学, 裳華房, (1995)
- [2] D. H. Ackley, G. E. Hinton and T. J. Sejnowski, Cong. Science, Vol. **9**, No. 1, 147-169, (1985)
- [3] M. Aizenman, J. L. Lebowitz and D. Ruelle, Comm. Math. Phys., Vol. **112**, 3-20, (1987)
- [4] M. Aizenman, R. Sims and S. L. Starr, Phys. Rev. B, Vol. **68**, 214403, (2003)
- [5] J. R. L. de Almeida and D. J. Thouless, J. Phys. A, Vol. **11**, No. 5, 983-990, (1978)
- [6] S. Amari, Biol. Cybern, Vol. **26**, No. 3, 175-185, (1977)
- [7] S. Amari, Neural Networks, Vol. **8**, No. 9, 1379-1408, (1995)
- [8] D. J. Amit, H. Gutfreund and H. Sompolinsky, Phys. Rev. A, Vol. **32**, No. 2, 1007-1018, (1985)
- [9] D. J. Amit, H. Gutfreund and H. Sompolinsky, Phys. Rev. Lett., Vol. **55**, No. 14, 1530-1533, (1985)
- [10] D. J. Amit, H. Gutfreund and H. Sompolinsky, Ann. Phys., Vol. **173**, No. 1, 30-67, (1987)
- [11] D. J. Amit, H. Gutfreund and H. Sompolinsky, Phys. Rev. A, Vol. **35**, No. 5, 2293-2303, (1987)
- [12] 安藤繁, 戸辺義人, 田村陽介, 南正輝, センサネットワーク技術, 東京電機大学出版局, (2005)
- [13] J. K. Anlauf and M. Biehl, Europhys. Lett., Vol. **10**, No. 7, 687-692, (1989)
- [14] 荒木不二洋, 統計物理の数理, 岩波書店, (1994)
- [15] 有本卓, 現代情報理論, コロナ社, (1978)
- [16] D. M. L. Barbato and J. F. Fontanari, J. Phys. A, Vol. **26**, No. 8, 1847-1863, (1993)
- [17] T. H. Berlin and M. Kac, Phys. Rev., Vol. **86**, No. 6, 821-835, (1952)
- [18] H. A. Bethe, Proc. Roy. Soc. A, Vol. **150**, No. 871, 552-575, (1935)
- [19] C. Bhattacharyya and S. S. Keerthi, J. Phys. A, Vol. **33**, No. 7, 11307-1312, (2000)
- [20] K. Binder and A. P. Young, Rev. Mod. Phys., Vol. **58**, No. 4, 801-976, (1986)
- [21] C. M. Bishop, Neural Networks for Pattern Recognition, Oxford, (1995)
- [22] C. M. Bishop, Pattern Recognition And Machine Learning, Springer, (2006)
- [23] A. Blandin, J. Phys. Colloq., Vol. **C6-39**, 1499-1516, (1978)
- [24] H. D. Block, Rev. Mod. Phys., Vol. **34**, No. 1, 123-135, (1962)
- [25] M. Blume, V. J. Emery and R. B. Griffiths, Phys. Rev. A, Vol. **4**, No. 3, 1071-1077, (1971)
- [26] L. Boltzmann, Wiener Berichte, Vol. **53**, 195-220, (1866)
- [27] L. Boltzmann, Wiener Berichte, Vol. **66**, 275-370, (1872)
- [28] L. Boltzmann, Wiener Berichte, Vol. **76**, 373-435, (1877)
- [29] S. P. Boyd and L. Vandenberghe, Convex Optimization, Cambridge, (2004)
- [30] D. R. Bowman and K. Levin, Phys. Rev. B, Vol. **25**, No. 5, 3438-3441, (1982)
- [31] A. Braunstein and R. Zecchina, Phys. Rev. Lett., Vol. **96**, No. 3, 030201, (2006)
- [32] R. Brout, Phys. Rev., Vol. **115**, No. 4, 824-835, (1959)
- [33] V. Cannella and J. A. Mydosh, Phys. Rev. B, Vol. **6**, No. 11, 4220-4237, (1972)
- [34] P. Carnevali and S. Patarnello, Europhys. Lett., Vol. **4**, No. 10, 1199-1204, (1987)
- [35] T. Castellani and A. Cavagna, J. Stat. Mech., P05012, (2005)
- [36] J. Chalupa, Phys. Rev. B, Vol. **17**, No. 11, 4335-4338, (1978)
- [37] A. C. C. Coolen and D. Sherrington, Phys. Rev. Lett., Vol. **71**, No. 23, 3886-3889, (1993)
- [38] A. C. C. Coolen and D. Sherrington, Phys. Rev. E, Vol. **49**, No. 3, 1921-1934, (1994)

- [39] T. M. Cover, IEEE Trans. Electron. Comput., Vol. **14**, No. 3, 326-334, (1965)
- [40] T. M. Cover and J. A. Thomas, Elements of Information Theory, Wiley, (2006)
- [41] N. Cristianini and J. S.-Taylor, An Introduction to Support Vector Machines: And Other Kernel-Based Learning Methods, Cambridge, (2000)
- [42] A. P. Dempster, N. M. Laird, and D. B. Rubin, J. of Royal Stat. Soc. B, Vol. **39**, 1-38, (1977)
- [43] B. Derrida, Phys. Rev. Lett., Vol. **45**, No. 2, 79-82, (1980)
- [44] B. Derrida, Phys. Rev. B, Vol. **24**, No. 5, 2613-2626, (1981)
- [45] S. Diederich and M. Oppen, Phys. Rev. Lett., Vol. **58**, No. 9, 949-952, (1987)
- [46] C. De Dominicis, Phys. Rep., Vol. **67**, No. 1, 37-46, (1980)
- [47] C. De Dominicis, M. Gabay, T. Garel and H. Orland, J. Phys. (Paris), Vol. **41**, No. 9, 923-930, (1980)
- [48] S. F. Edwards, Proc. 4th Int. Conf. on Amorphous Materials, Wiley, (1970)
- [49] S. F. Edwards, Polymer Networks, Plenum, (1971)
- [50] S. F. Edwards, J. Phys. A, Vol. **13**, L239-241, (1980)
- [51] S. F. Edwards and P. W. Anderson, J. Phys. F, Vol. **5**, No. 10, 965-974, (1975)
- [52] S. F. Edwards and R. C. Jones, J. Phys. A, Vol. **9**, No. 10, 1595-1603, (1976)
- [53] K. Efetov, Supersymmetry in Disorder and Chaos, Cambridge, (1999)
- [54] V. J. Emery, Phys. Rev. B, Vol. **11**, No. 1, 239-247, (1975)
- [55] A. Engel and C. Van den Broeck, Statistical Mechanics of Learning, Cambridge, (2001)
- [56] R. S. Ellis, Entropy, Large Deviations and Statistical Mechanics, Springer, (2006)
- [57] S. E. Fahlman, G. E. Hinton and T. J. Sejnowski, Proc. Nat. Conf. on AI, AAAI-83, 109, (1983)
- [58] J. D. Fast, Entropy, Palgrave Macmillan, (1970)
- [59] W. Feller, An Introduction to Probability Theory and Its Application, John Wiley & Sons Inc, (1968)
- [60] R. P. Feynmann, Statistical Mechanics, Westview Press, (1998)
- [61] R. D. Fields and B. S. Graham, Science, Vol. **298**, 556-562, (2002)
- [62] L. O. de O. Filho, F. A. da Costa and C. S. O. Yokoi, Phys. Rev. E, Vol. **74**, No. 3, 031117, (2006)
- [63] K. H. Fischer and J. A. Hertz, Spin Glasses, Cambridge, (1991)
- [64] R. A. Fisher, Proc. Cambridge Phil. Soc., Vol. **22**, 700-725, (1925)
- [65] J. F. Fontanari and R. Meir, J. Phys. A, Vol. **26**, No. 5, 1077-1089, (1993)
- [66] B. J. Frey, Graphical Models for Machine Learning and Digital Communication, MIT, (1999)
- [67] F. Fröhlich and F. R. N. Nabarro, Proc. Roy. Soc. A, Vol. **175**, 382-391, (1940)
- [68] 福島邦彦, 神経回路と情報処理, 朝倉書店, (1989)
- [69] A. Gamba, L. Gamberini, G. Palmieri and R. Sanna, Nuovo Cimento, Vol. **20**, No. 2, 112-115, (1961)
- [70] E. Gardner, J. Phys. A, Vol. **19**, No. 16, L1047-L1052, (1986)
- [71] E. Gardner, J. Phys. A, Vol. **21**, No. 1, 257-270, (1988)
- [72] T. Geiszi, Physical Models of Neural Networks, World Scientific, (1990)
- [73] J. W. Gibbs, Trans. Conn. Acad. Scien., Vol. **3**, 108-248, (1876)
- [74] J. W. Gibbs, Trans. Conn. Acad. Scien., Vol. **3**, 343-524, (1878)
- [75] P. D. Giudice, S. Franz and M. A. Virasoro, J. Phys. France, Vol. **50**, No. 2, 121-134, (1989)
- [76] D. Gou and S. Verdú, IEEE IT Trans., Vol. **51**, No. 6, 1983-2010, (2005)
- [77] G. Grinstein and A. H. Luther, Phys. Rev. B, Vol. **13**, No. 3, 1329-1343, (1976)
- [78] F. Guerra and F. L. Toninelli, Comm. Math. Phys., Vol. **230**, 71-79, (2002)
- [79] F. Guerra, Comm. Math. Phys., Vol. **233**, 1-12, (2003)

- [80] G. Györgyi, Phys. Rev. A, Vol. **41**, No. 12, 7097-7100, (1990)
- [81] G. Györgyi and N. Tishby, Neural networks and Spin Glasses, ed. K. Thuemann and R. Koeberle, 3-36, (1990)
- [82] 韓太舜, 小林欣吾, 情報と符号化の数理, 培風館, (1999)
- [83] G. Hardy, J. E. Littlewood and G. Pólya, Inequalities, Cambridge, (1991)
- [84] R. V. L. Hartley, Bell Sys. Tech. J., Vol. **7**, 535, (1928)
- [85] P. G. Haydon, Nature Rev., Vol. **2**, 185-193, (2001)
- [86] D. O. Hebb, Organization of Behavior, Wiley New York, (1949)
- [87] J. L. van Hemmen, Phys. Rev. Lett., Vol. **49**, No. 6, 409-412, (1982)
- [88] J. L. van Hemmen and R. G. Palmer, J. Phys. A, Vol. **12**, No. 4, 563-580, (1979)
- [89] J. A. Hertz, A. S. Krogh and R. G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, (1991)
- [90] G. E. Hinton and J. A. Anderson, Parallel Models of Associative Memory, Lawrence Erlbaum Associates, (1981)
- [91] G. E. Hinton and T. J. Sejnowski, Proc. of IEEE Conf. Computer Vision and Pattern Recognition, 448-453, (1983)
- [92] J. J. Hopfield, Proc. Natl. Acad. Sci., Vol. **79**, No. 8, 2554-2558, (1982)
- [93] J. J. Hopfield, Proc. Natl. Acad. Sci., Vol. **81**, No. 10, 3088-3092, (1984)
- [94] J. J. Hopfield and D. W. Tank, Biol. Cybern., Vol. **52**, No. 3, 141-152, (1985)
- [95] J. J. Hopfield, Proc. Natl. Acad. Sci., Vol. **84**, No. 23, 8429-8433, (1987)
- [96] T. Hosaka, Y. Kabashima and H. Nishimori, Phys. Rev. E, Vol. **66**, 066126, (2002)
- [97] T. Hosaka and Y. Kabashima, J. Phys. Soc. Jpn., Vol. **74**, No. 1, 488-497, (2005)
- [98] 保坂忠明, 歪み有りデータ圧縮の統計力学的研究, 学位論文 (東京工業大学), (2005)
- [99] 伊庭幸人・種村正美・大森裕浩・和合肇・佐藤整尚・高橋明彦, 計算統計 II, 統計科学のフロンティア 12, 岩波書店, (2005)
- [100] 市村浩, 統計力学, 裳華房, (1992)
- [101] 今井秀樹, 情報理論, 昭晃堂, (1984)
- [102] 今井秀樹, 符号理論, 電子情報通信学会, (1990)
- [103] 今田正俊, 統計物理学, 丸善, (2004)
- [104] 井上純一, 情報理論, プレアデス出版, (2008)
- [105] 石井健一郎, 前田英作, 上田修功, 村瀬洋, わかりやすいパターン認識, オーム社, (1998)
- [106] 伊丹誠, わかりやすい OFDM 技術, オーム社, (2005)
- [107] C. Itzykson and J. B. Zuber, J. Math. Phys., Vol. **21**, No. 3, 411-421, (1980)
- [108] D. Jasnow and M. E. Fisher, Phys. Rev. B, Vol. **13**, No. 3, 1112-1118, (1976)
- [109] Y. Kabashima, J. Phys. A, Vol. **36**, No. 43, 11111-11121, (2003)
- [110] Y. Kabashima, J. Phys. Conf. Ser., Vol. **95**, 012001, (2008)
- [111] M. Kac, unpublished, Trondheim Theoretical Physics Seminar Nordita Publ., No. 286, (1968)
- [112] 鎌谷直之, 遺伝統計学入門, 岩波書店, (2007)
- [113] T. Kasuya, Prog. Theor. Phys., Vol. **16**, 45-57, (1956)
- [114] 川村光, 統計物理, 丸善, (2000)
- [115] 久保亮五, 統計力学, 共立出版, (1952)
- [116] 久保亮五, 市村浩, 碓井恒丸, 橋爪夏樹, 統計力学, 裳華房, (1961)
- [117] R. Kikuchi, Phys. Rev., Vol. **81**, No. 6, 988-1003, (1951)
- [118] S. Kirkpatrick and D. Sherrington, Phys. Rev. B, Vol. **17**, No. 11, 4384-4403, (1978)
- [119] K. Kitagawa and T. Tanaka, IEEE, Int. Symp. Info. Theory, 2008, 1373-1377, (2008)
- [120] M. W. Klein and R. Brout, Phys. Rev., Vol. **132**, No. 6, 2412-2426, (1963)

- [121] T. Kohonen, Self-Organization and Associative Memory, Springer-Verlag, (1989) , T. Kohonen (著), 中谷和夫 (翻訳), 自己組織化と連想記憶, シュプリンガー東京, (1993)
- [122] T. Kohonen, Self-Organizing Maps, Springer-Verlag, (2000) , T. Kohonen (著), 徳高平蔵, 堀尾恵一, 大北正昭, 大藪又茂, 藤村喜久郎, 自己組織化マップ, シュプリンガー東京, (2005)
- [123] J. M. Kosterlitz, D. J. Thouless and R. C. Jones, Phys. Rev. Lett., Vol. **36**, No. 20, 1217-1220, (1976)
- [124] W. Krauth and M. Mézard, J. Phys. France, Vol. **50**, 3057-3066, (1989)
- [125] A. Krogh and P. Sollich, Phys. Rev. E, Vol. **55**, No. 1, 811-825, (1997)
- [126] 工藤佳久, 序 グリアーニューロン回路網の概念, 細胞工学 Vol. **22** No. 4 390-392 (2003)
- [127] 熊沢逸夫, 学習とニューラルネットワーク, 森北出版, (1998)
- [128] L. D. Landou and E. M. Lifshitz, Statistical Physics, Butterworth-Heinemann, (1984)
- [129] E. Levin, N. Tishby and S. A. Solla, Proc. IEEE, Vol. **78**, No. 10, 1568-1574, (1990)
- [130] T. F. Lin, J. Math. Phys., Vol. **11**, No. 5, 1584-1590, (1970)
- [131] D. J. C. Mackay, Information Theory, Inference, and Learning Algorithms, Cambridge, (2003)
- [132] V. A. Marčenko and L. A. Pastur, Math. USSR-Sbornik, Vol. **1**, 457-483, (1967)
- [133] 松尾憲一, スペクトラム拡散技術のすべて, 東京電機大学出版局, (2002)
- [134] W. McCulloch and W. Pitts, Bull. Math. Bio., Vol. **5**, No. 4, 115-133, (1943)
- [135] M. L. Mehte, Random Matrices, Academic Press, (2004)
- [136] P. L. Meyer, Introductory Probability and Statistical Applications, Addison-Wesley Educational Publishers, (1970)
- [137] M. Mézard, J. Phys. A, Vol. **22**, No. 21, 2181-2190, (1989)
- [138] M. Mézard, G. Parisi and M. Virasoro, Spin Glass Theory and Beyond, World Scietific, (1987)
- [139] M. L. Minsky and S. Papert, Perceptron, An introduction to Computational Geometry, MIT, (1969)
- [140] 宮川雅巳, 応用統計学, Vol. **16**, No. 1, 1-19, (1987)
- [141] 宮下精二, 熱・統計力学, 培風館, (1993)
- [142] S. Miyoshi, K. Hara and M. Okada, Phys. Rev. E, Vol. **71**, No. 3, 036116, (2005)
- [143] S. Miyoshi and M. Okada, J. Phys. Soc. Jpn., Vol. **75**, No. 2, 024003, (2006)
- [144] 元田浩, 山口高平, 津本周作, 沼尾正行, データマイニングの基礎, オーム社, (2006)
- [145] R. R. Müller, EUSIPCO '04, 189-196, (2004)
- [146] 村田昇, 情報理論の基礎, サイエンス社, (2008)
- [147] T. Murayama, Phys. Rev. E, Vol. **69**, 035105, (2004)
- [148] 室橋弘人・中村健太郎・川端一光・福中公輔・岩間徳兼・堀辺千晴・君島康昭・久保沙織・鈴川由美・竹下恵・池原一哉, マルコフ連鎖モンテカルロ法, 統計ライブラリー, 朝倉書店, (2008)
- [149] P. Nachbar, J. A. Nossek and J. Strobl, IEEE ISCAS '93, Vol. 4, 2152-2155, (1993)
- [150] 長岡洋介, 統計力学, 岩波書店, (1994)
- [151] K. Nakanishi, Phys. Rev. B, Vol. **23**, No. 7, 3514-3522, (1981)
- [152] 中野藤生・服部眞澄, エルゴード性とは何か, 丸善, (1994)
- [153] 中野良平, ニューラル情報処理の基礎数理, 数理工学社, (2005)
- [154] I. Neri, N. S. Skantzios and D. Bollé, J. Stat. Mech., P10018, (2008)
- [155] C. M. Newman and D. L. Stein, Phys. Rev. Lett., Vol. **76**, No. 3, 515-518, (1996)
- [156] N. J. Nilsson, Learning Machine, McGraw-Hill Edu. (1965), Mathematical Foundations of Learning Machines, Morgan Kaufmann Pub., (1990)
- [157] 西川恭治, 森弘之, 統計物理学, 朝倉書店, (2000)
- [158] H. Nishimori, Statistical Physics of Spin Glasses and Information Processing, Oxford, (2001), 西森秀稔, スピングラス理論と情報統計力学, 岩波書店, (1999)

- [159] 西森秀稔, 相転移・臨界現象の統計物理学, 培風館, (2006)
- [160] 西山清, 最適フィルタリング, 培風館, (2001)
- [161] H. Nyquist, Bell Sys. Tech. J., Vol. **3**, 324, (1924)
- [162] 小口武彦, 磁性体の統計理論, 裳華房, (1970)
- [163] 小口武彦, スピングラスとは何か, 物理学最前線 8, 共立出版, (1984)
- [164] 岡部豊, 阿部龍蔵, 川村清, 統計力学, 裳華房, (2000)
- [165] 大平徹, ノイズと遅れの数理, 共立出版, (2006)
- [166] M. Opper and D. Saad, Advanced Mean Field Methods, Theory and Practice, MIT, (2001)
- [167] M. Opper and O. Winther, Phys. Rev. E, Vol. **64**, No. 5, 056131, (2001)
- [168] G. Parisi and M. Potters, Europhys. Lett., Vol. **32**, No. 1, 13-17, (1995)
- [169] G. Parisi and M. Potters, J. Phys. A, Vol. **28**, No. 18, 5267-5285, (1995)
- [170] L. A. Pastur and M. V. Shcherbina, J. Stat. Phys., Vol. **62**, 1-19, (1990)
- [171] T. Plefka, J. Phys. A, Vol. **15**, 1971-1978, (1982)
- [172] M. Prettì, J. Stat. Mech., P11008, (2005)
- [173] R. A. Redner and H. F. Walker, SIAM Review, Vol. **26**, 195-239, (1984)
- [174] F. Rosenblatt, Psych. Rev., Vol. **65**, No. 6, 386-408, (1958)
- [175] M. A. Ruderman and C. Kittel, Phys. Rev., Vol. **96**, 99-102, (1954)
- [176] D. E. Rumelhart, G. E. Hinton and R. J. Williams, Parallel Distributed Processing, Vol. **1**, 318-362, (1986)
- [177] D. E. Rumelhart, G. E. Hinton and R. J. Williams, Nature, Vol. **323**, 533-536, (1986)
- [178] D. E. Rumelhart, B. Widrow and M. A. Lehr, Comm. ACM, Vol. **37**, No. 3, 87-92, (1994)
- [179] S. J. Russell and P. Norvig, Artificial Intelligence A Modern Approach, Pearson Education 2nd, (2003), S. J. Russell and P. Norvig(著), 古川康一(監訳) エージェントアプローチ 人工知能 第2版 共立出版, (2008)
- [180] 佐藤洋, 情報理論, 裳華房, (1983)
- [181] T. Schneider and E. Pytte, Phys. Rev. B, Vol. **15**, No. 3 1519-1522, (1977)
- [182] C. E. Shannon, Bell Sys. Tech. J., Vol. **27**, 379, (1948)
- [183] D. Sherrington and B. W. Southern, J. Phys. F, Vol. **5**, L49-L53, (1975)
- [184] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett., Vol. **35**, No. 26, 1792-1796, (1975)
- [185] T. Shinzato and Y. Kabashima, J. Phys. A, Vol. **41**, No. 32, 324013, (2008)
- [186] T. Shinzato and Y. Kabashima, J. Phys. A, Vol. **42**, No. 1, 015005, (2009)
- [187] H. -J. Sommers, Z. Phys. B, Vol. **31**, 301-307, (1980)
- [188] H. Sompolinsky and N. Barkai, IJCNN '93 Nagoya Tutorial 221-240, (1993)
- [189] B. W. Southern, J. Phys. C, Vol. **9**, 4011-4020, (1976)
- [190] B. W. Southern and A. P. Young, J. Phys. C, Vol. **10**, L79-L83, (1977)
- [191] K. Takeda, S. Uda and Y. Kabashima, Europhys. Lett., Vol. **76**, No. 6, 1193-1199, (2006)
- [192] K. Takeda, A. Hatabu and Y. Kabashima, J. Phys. A, Vol. **40**, No. 47, 14085-14098, (2007)
- [193] 田中和之, 確率モデルによる画像処理技術入門, 森北出版, (2006)
- [194] T. Tanaka, IEEE IT Trans., Vol. **48**, No. 11, 2888-2910, (2002)
- [195] T. Tanaka, J. Phys. Conf. Ser., Vol. **95**, 012002, (2008)
- [196] T. Tanaka and K. Kitagawa, IEEE WiOPT2008, 604-609, (2008)
- [197] 田崎晴明, 熱力学, 培風館, (2000)
- [198] D. J. Thouless, P. W. Anderson and R. G. Palmer, Phil. Mag., Vol. **35**, 593-601, (1977)
- [199] N. Tishby, E. Levin and S. A. Solla, Neural Networks, Vol. **2**, 403-409, (1989)

- [200] E. C. Titchmarsh, *The Theory of Functions* 2nd ed. Oxford, (1939)
- [201] G. Toulouse, *Commun. Phys.*, Vol. **2**, 115-119, (1977)
- [202] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Now Publishers, (2004)
- [203] S. Uda and Y. Kabashima, *J. Phys. Soc. Jpn.*, Vol. **74**, No. 8, 2233-2242, (2005)
- [204] 宇田新介, パーセプトロン型確率モデルにおけるベイズ学習のための平均場アルゴリズムの研究, 学位論文 (東京工業大学), (2006)
- [205] 上坂吉則, ニューロコンピューティングの数学的基礎, 近代科学社, (1993)
- [206] L. G. Valiant, *Comm. ACM*, Vol. **27**, No. 11, 1134-1142, (1984)
- [207] S. S. Venkatesh, *American Inst. Phys. Conf. Proc.*, Vol. **151**, No. 1, 440-445, (1986)
- [208] 汪金芳, 田栗正章, 手塚集, 樺島祥介, 上田修功, 計算統計 I, 統計科学のフロンティア 11, 岩波書店, (2003)
- [209] L. R. Welch, *IEEE Trans. on IT*, Vol. **20**, No. 3, 397-399, (1974)
- [210] A. Wendemuth, M. Oppen and W. Kinzel, *J. Phys. A*, Vol. **26**, No. 13, 3165-3185, (1993)
- [211] B. Widrow, Stanford Elect. Lab., Tech. Rep., 15532, (1960)
- [212] B. Widrow, *Self-Organizing Systems*, 435-461, Spartan Books, (1962)
- [213] B. Widrow and M. A. Lehr, *Proc. IEEE*, Vol. **78**, No. 9, 1415-1442, (1990)
- [214] J. Wishart, *Biometrika*, Vol. **20**, A 32-52, (1928)
- [215] E. Wigner, *Ann. of Math.*, Vol. **62**, 546-564, (1955)
- [216] M. Yasuda and K. Tanaka, *J. Phys. A*, Vol. **40**, No. 33, 9993-10007, (2007)
- [217] J. M. Yeomans, *Statistical Mechanics of Phase Transitions*, Oxford, (1992)
- [218] K. Yoshida, *Phys. Rev.*, Vol. **106**, 893-898, (1957)
- [219] C. Zener, *Phys. Rev.*, Vol. **81**, 440-444, (1951), Vol. **83**, 299-301, (1951)