

論文 / 著書情報  
Article / Book Information

題目(和文)	パラメータ制約付き特異モデルの統計的学習理論
Title(English)	Statistical Learning Theory of Parameter-Restricted Singular Models
著者(和文)	林 直輝
Author(English)	Naoki Hayashi
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12028号, 授与年月日:2021年6月30日, 学位の種別:課程博士, 審査員:渡邊 澄夫,高安 美佐子,金森 敬文,山下 真,澄田 範奈
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第12028号, Conferred date:2021/6/30, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

(博士課程)  
Doctoral Program

## 論文要旨

THESIS SUMMARY

系・コース：	数理・計算科学	系
Department of, Graduate major in	数理・計算科学	コース
学生氏名：	林 直輝	
Student's Name		

申請学位 (専攻分野)：	博士	( 理学 )
Academic Degree Requested	Doctor of	
指導教員 (主)：	渡邊 澄夫	
Academic Supervisor(main)		
指導教員 (副)：		
Academic Supervisor(sub)		

要旨 (英文 800 語程度)  
Thesis Summary (approx.800 English Words )

Statistical models used in machine learning are called learning machines. Learning machines are widely applied to the prediction of events and the discovery of knowledge in many fields. Indeed, machine learning has grown over the last several decades, resulting in that publications and applications have been rapidly increasing in recent years. In order to estimate complex structures of parts in the real world from examples, learning machines often have hierarchical structures or hidden variables. The purpose of statistical learning theory is to clarify the generalization performances of such learning machines.

Singular learning theory is a mathematical foundation for statistical inference using singular models. Typical hierarchical models, such as neural networks, tree and forest models, mixture models, matrix factorizations, and topic models, are statistically singular since a map from a parameter to a probability density function is not one-to-one. Hence, almost all learning machines are singular models. To clarify the generalization behaviors of such models is a foundation to estimate sufficient sample sizes, design models, and tune hyperparameters. The conventional statistical theory assuming the regularity condition cannot be applied to these models because their likelihoods cannot be approximated by any normal distribution. Singular learning theory provides a general view for this problem; birational invariants of an analytic set (a.k.a. algebraic variety) determine the generalization error. That set is defined by zero of a Kullback-Leibler (KL) divergence between the data-generating distribution and the model. Algebraic structures of statistical models are essential in singular learning theory; thus, that theory can be interpreted as an intersection between algebraic statistics and statistical learning theory.

One of such invariants is a real log canonical threshold (RLCT). An RLCT is a negative-maximum pole of a zeta function defined by an integral of a KL divergence. Determining an RLCT of a concrete model is performed by resolution of singularities. In fact, algebraic statisticians and machine learning researchers have derived the exact values or upper bounds of the RLCTs for several singular models. The theoretical value of the RLCT is effective in statistical model selection methods such as sBIC proposed by Drton and Plummer. Besides, Nagata proposed a

tuning method using RLCTs for exchange Monte Carlo.

On the other hand, from the practical point of view, the parameter region of the model is often restricted to improve interpretability. Non-negative matrix factorization (NMF) and latent Dirichlet allocation (LDA) are well-known examples of parameter-restricted singular models. In general, such constraints make the generalization error changed. However, for each singular model and condition, the quantitative effect of those constraints has not yet been clarified because the singularities in the above analytic set are also changed by the restriction to the parameter region.

In this dissertation, in order to establish a foundation of a singular learning theory of parameter-restricted statistical models, we theoretically study the asymptotic behavior of the Bayesian generalization error in NMF and LDA, which are two typical singular models whose parameter regions are constrained and show the following results.

- In NMF, the restricted parameters are on the non-negative region. We derive an upper bound of the RLCT and a lower bound of the variational approximation error. This theoretical analysis for NMF shows a phase transition structure; there is a critical line of hyperparameters. The Bayesian generalization error and the variational approximation error drastically changes beyond that line. The phase transition line we found is different from that of variational Bayesian NMF. That is because the variational posterior of NMF is different from the Bayesian posterior distribution of NMF.
- In LDA, the constrained parameters are on the simplex region. We prove its RLCT is equal to that of matrix factorization with simplex restriction and clarify the exact asymptotic form of the generalization error, i.e. we determine the exact value of the RLCT of LDA. This mathematical study for LDA shows the RLCT of LDA is much smaller than that of a regular model whose parameter dimension is the same as LDA. Besides, when the number of topics increases, the RLCT monotonically and non-linearly grows but is bounded, whereas the parameter dimension linearly does and is not bounded.
- These results also provide quantitative differences of generalization errors from matrix factorization whose parameter space is not restricted. The Bayesian generalization error in NMF becomes strictly larger than that of non-restricted matrix factorization when the entries of the true parameter matrices are zero. On the other hand, the Bayesian generalization error in LDA is strictly smaller than that in an LDA-like model whose parameter region would have no constraint.

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note: Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1 copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ (T2R2) にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).