論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Generalization Analysis on dependency of Model Complexity in Bayesian Deep Neural Network
著者(和文)	永安修也
Author(English)	Shuya Nagayasu
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12656号, 授与年月日:2024年3月26日, 学位の種別:課程博士, 審査員:渡邊 澄夫,金森 敬文,山下 真,中野 張,髙邉 賢史
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第12656号, Conferred date:2024/3/26, 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： Department of, Graduate major in	数理計算科学 系 数理計算科学 コース		申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of （ 理学 ）
学生氏名： Student's Name	永安修也		審査員主査： Chief Examiner	渡辺澄夫

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

Deep Neural Networks are widely used as machine learning models for computer vision, natural language processing, robotics, and many other areas. The feature of modern practical neural network is the large number of parameters. In machine learning, the generalization error is important for measuring the performance of learning model. In the generalization error, bias term and variance term exist. It is well known that if the number of parameters of a model increase, bias term corresponding to approximation ability decreases but the variance term corresponding to fit to noise increases. This property is called bias-variance trade-off. However, in practical neural network, no matter how many the number of parameters increase, the generalization error decreases. It is inconsistent with the bias-variance trade-off. One theory on this problem focuses on Bayesian inference. Recent studies revealed the relationship between stochastic gradient descent and stochastic gradient Langevin dynamics which realize Bayesian learning. From the perspective of Bayesian learning, these studies explain the high performance of neural networks. On the other hand, many studies try to realize the Bayesian Deep Neural Networks. From these viewpoints, Bayesian learning in Deep Neural Networks has become more important.

In this thesis we theoretically show that the performance of deep neural networks is different from the classical statistical models in perspective of Bayesian learning through the statistical analysis of generalization error in two cases.

One is the case that there exist multiple optimal probability distributions which are nearest to data generating process. In conventional learning theory, the optimal probability distribution was assumed to be one point in parameter space or multiple point but one probability distribution. Former case is called regular and later case is called singular. But in case that optimal probability distributions are multiple, all things were left unknown. In our study we analyze the generalization error in the case that optimal probability distribution is not unique as an extension of the theory in singular case. In this case, we proved that the generalization error increases while the number of data increases. In conventional learning theory, the number of data increases, the generalization error decreases. The new term of generalization error depends on the maximum value of multivariate normal distribution. This case occurs when the model has relatively smaller complexity than the data-generating process. This result suggests that if the bias exists on the large model, the relationships between the generalization error and increment of complexity of the model or the number of data are different from those in smaller classical statistical model. These cases are counterexamples to the bias-variance trade-off. We show sufficient condition of such counterexamples.

The other is the case that the model has larger number of parameters than necessary in convolutional neural networks with the ReLU activation function. We show theoretical value of variance term of the generalization error in convolutional neural networks on the case of including and not including skip connection. In former studies of Bayesian learning, the variance of Neural networks is only known in case of one hidden layer network with analytic activation function. In our study we assume that convolutional neural networks have any number of convolutional layers and fully connected layers. In case without skip connection, the variances increase if the number of parameters increases. On the other hand, in case with skip connection, the variance does not increase even if the number of parameters increases. This result means that in convolutional neural networks with skip connection, redundant parameters than sufficient large network which can approximate the data do not affect the

generalization error in Bayesian learning. This is also an example contrary to bias-variance trade-off.

Our studies theoretically show that these cases are different from bias-variance trade-off in Bayesian learning and why generalization errors of neural networks are smaller than classical learning machine.