The Complexity of Algorithmic Hypothesis Class

Tongliang Liu

Faculty of Engineering and Information Technology

University of Technology Sydney

A thesis submitted for the degree of

Doctor of Philosophy

2016

To my loving parents
Lianhua Han and *Baojun Liu*

Certificate of Original Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Tongliang Liu

Acknowledgements

I would like to thank everyone who has helped me to finish my doctoral studies.

First of all, I am extremely grateful to my supervisor Prof Dacheng Tao. Every aspect of this thesis has benefited from his supervision, and I feel lucky to have had him as my advisor. He has given me the trust and freedom to pursue my research interests, always replied to my emails promptly, and provided timely and constructive suggestions to overcome difficulties. I will most miss our face-to-face meetings. His insight and passion have always excited and energized me.

I also wish to express my sincere appreciation to Prof Zengfu Wang, Dr Liheng Zhao, and Dr Jun Yu, who encouraged me into academia when I was at the University of Science and Technology of China (USTC). I was also fortunate to spend time with Prof Gábor Lugosi and Dr Gergely Neu, who carefully supervised me when I visited the Statistics and Random Structures research group (STARS) at the Barcelona Graduate School of Economics. Chapters 4 and 6 grew out of this collaboration.

I also would like to give special thanks to my excellent collaborators: Prof Stephen J. Maybank, Prof Dong Xu, Prof Yun Raymond Fu, Prof Mingli Song, Prof Thomas S. Huang, Prof Tieniu Tan, Prof Junjie Wu, Prof Qingshan Liu, A/Prof Xinmei Tian, A/Prof Jie Gui, Assistant Professor Yubao Sun, Dr Yong Luo, Dr Xiaoyan Li, Ms Yanan Lu, Mr Chang Xu, Mr Chen Gong, Mr Hongfu Liu, Mr Ming Shao, Mr Sheng Li, Mr Ya Li, and Mr Hao Xiong.

I have been fortunate to work and have discussions with many other brilliant researchers: A/Prof Weifeng Liu, Dr Naiyang Guan, Dr Nan-

nan Wang, Dr Fei Gao, Assistant Professor Chaohui Wang, Dr Yunlong Feng, A/Prof Kaibing Zhang, Dr Lianyang Ma, A/Prof Chenping Hou, Dr Xiao Liu, Prof Lianwen Jin, Dr Weilong Hou, A/Prof Bo Du, A/Prof Tao Lei, A/Prof Wankou Yang, A/Prof Shigang Liu, and A/Prof Xianye Ben. Their support and kind company have been a constant source of confidence, strength, and courage in my research life. I wish to express appreciation to all of them.

I am so grateful to my friends in Sydney: Mingming Gong, Ruxin Wang, Zhibin Hong, Shaoli Huang, Meng Fang, Qiang Li, Maoying Qiao, Zhe Xu, Changxing Ding, Wei Bian, Tianyi Zhou, Jun Li, Zhiguo Long, Guodong Long, Jing Jiang, Long Lan, Chunyang Liu, Bozhong Liu, Shirui Pan, Jia Wu, Baosheng Yu, Huan Fu, Zhe Chen, Xiyu Yu, Liu Liu, Yali Du, and Jiankang Deng. I am especially deeply indebted to Mingming, who has provided strong motivation and critical guidance in both my research and daily life. I would also like to express my sincere thanks to Moe, who has tolerated me and given me confidence. All my friends here have provided strong support during both happy and stressful times. I am also incredibly grateful to Kede Ma, Haifeng Liu, Zhangyang Wang, Yiming Qian, Xu Shen, Puyu Liu, Xuemeng Song, Yifei Wang, Di Che, Jiaji Pan, Xiao Zhang, and Zhiyong Chen, who have accompanied me from my bachelor's degree to doctoral studies. I owe my deepest thanks to all of them!

Finally, I would like to express deep-felt gratitude to my family: my parents, my elder sister, my grandparents, my uncles, and my aunties for their endless love, trust, encouragement and full support throughout my studies and life.

I dedicate this thesis to them.

Abstract

Statistical learning theory provides the mathematical and theoretical foundations for statistical learning algorithms and inspires the development of more efficient methods. It is observed that learning algorithms may not output some hypotheses in the predefined hypothesis class. Therefore, in this thesis, we focus on statistical learning theory and study how to measure the complexity of the algorithmic hypothesis class, which is a subset of the predefined hypothesis class that a learning algorithm will (or is likely to) output. By designing complexity measures for the algorithmic hypothesis class, we provide new generalization bounds for k-dimensional coding schemes and multi-task learning and propose two frameworks to derive tighter generalization bounds than the current state-of-the-art.

We take k-dimensional coding schemes, a set of unsupervised learning algorithms, and multi-task learning, a set of supervised learning algorithms, as examples to demonstrate that learning algorithm outputs may have special properties and are therefore included in a subset of the predefined hypothesis class. By analyzing the subsets (or the algorithmic hypothesis classes), we shed new light on learning problems and derive tighter generalization bounds than the current state-of-the-art. Specifically, for k-dimensional coding schemes, we show that the induced algorithmic loss function classes are sets of Lipschitz-continuous hypotheses and that a dimensionality-dependent complexity measure helps to derive small Lipschitz constants and thus improve the generalization bounds. For multi-task learning, we prove that tasks can act as regularizer and that feature structures can contribute to a small algorithmic hypothesis class and also help to improve the generalization bounds.

To more precisely exploit algorithmic hypothesis class complexity by considering the hypothesis and feature structure properties, we extend algorithmic robustness and stability to complexity measures for the hypothesis class.

Inspired by the idea of algorithmic robustness, we propose the complexity measure of uniform robustness. Compared to the Rademacher complexity, our measure more finely considers the geometric information of data. For example, when the sample space is covered by a small number of small radius and widely separated balls, the uniform robustness can be very small while the Rademacher complexity can be very large. Moreover, based on the definition of uniform robustness, we also provide a framework to derive generalization bounds for a very general class of learning algorithms.

We exploit the algorithmic hypothesis class of stable algorithms by studying the definition of algorithmic stability. Stable learning algorithms have the property that their outputs will not change much when one training example is changed. This implies that their outputs will not be sufficiently far apart, even though the training sample is completely altered. Thus, stable learning algorithms often have small algorithmic hypothesis classes. However, since measuring the complexity of the small algorithmic hypothesis class is unknown, we design a novel complexity measure called the algorithmic Rademacher complexity to measure the algorithmic hypothesis class of stable learning algorithms and provide sharper error bounds than the current state-of-the-art.

Contents

List of Figures