Department of Complexity Science and Engineering
Graduate School of Frontier Sciences
The University of Tokyo
東京大学大学院新領域創成科学研究科
複雑理工学専攻

2019
平成 30 年度

Master's Thesis
修士論文

A Comprehensive Study on Multiclass Classification with Rejection
棄却付き多クラス分類問題の包括的研究

Keyword: Statistical Machine Learning, Classification
キーワード：統計的機械学習, 分類

Supervisor: Professor Masashi Sugiyama
指導教員：杉山　将　教授

2019 年 2 月 19 日提出
Submitted on February 19th, 2019

Chenri Ni
倪　辰日
47-176121

Abstract

Machine learning is a process of discovering important hidden structures or patterns from data. Recently, machine learning methods have been widely adopted in various real world applications including medical diagnosis and automatic driving technologies. In spite of tremendous progress in machine learning, making wrong predictions is an inevitable issue. Making mistakes in prediction can be costly and dangerous, or sometimes even life-threatening.

To cope with this problem, a framework called *learning with rejection* has gained much attention recently. This is a framework where a learner is given additional option to reject the data, or to refrain from making a decision if it is unsure about the correctness of its prediction. In this thesis, we give a comprehensive study on learning with rejection under multiclass classification setting from two approaches.

First, we consider the *separation-based approach*, where the roles of the classifier and the rejector are separated, and both models are trained simultaneously. Although previous work revealed that calibration is achievable, that is, Bayes-optimal rejection and classification can be obtained in the binary case, we show that extensions of this approach to the multiclass case leads to failure in achieving calibration.

Secondly, we focus on the *confidence-based approach*, where the rejector is determined after training a classifier. We analyze this approach for the well-known losses, including the one-versus-all loss and the cross-entropy loss. We prove that both losses are calibrated to the Bayes-optimal solution by providing excess risk bounds.

In summary, we give a broad investigation into multiclass classification with rejection, and reveal the possibility and impossibility of achieving calibration from two different approaches. We conclude that these methods, especially the confidence-based method, can be a good baseline for this problem, and worth a further study.