Supervised Learning using Interpretable Ensemble Methods

Qi Xiao

November 2021

Abstract

In this dissertation, we focus on supervised learning using interpretable ensemble methods. We design interpretable supervised learning algorithms by introducing an interpretable model or manipulation input features. Chapter 1 will briefly introduce the background of supervised learning, ensemble methods, and model interpretability. The next three chapters contain three research projects. In chapter 2, I will introduce an efficient random ensemble method. The interpretability comes from the simple base learners and straightforward selection. The random generation saves computational time and makes the method efficient.

In chapter 3 and 4, we will focus on feature selection as data manipulation for interpretability. Both chapters will use the same ensemble model, mixture of experts. In chapter 3, I propose to use a mixture of experts for instance-wise feature selection. By assuming the group-ness structure in the feature selection, the mixture of experts can divide the original problem into several sub-problems. The mixture structure makes the selection compact and is easy to interpret. We can use end-to-end stochastic gradient descent to train the model accurately. In chapter 4, I will present a new feature selection problem setting called group-wise feature selection. It sits between global feature selection and instance-wise feature election and seeks a trade-off between expressiveness and interpretation. We propose to use clustering and mixture of experts to build a group-wise feature selector. The proposed feature selector structure can easily be combined with different discriminative models. Finally, in chapter 5, I will conclude the projects and provide some further directions.