Ensembles and Methods Constructing Ensembles

Li Zichao
15220162202173

May 5, 2019

Abstract

This report mainly concludes the basic concept of ensemble methods and several methods for constructing ensembles.

1 Ensemble Methods

Ensemble methods are learning algorithms that construct a set of calssifiers and then classify new data points by taking a weighted vote of their prediction. The original ensemble method is Bayesian averaging, but more recent algorithms include error-correcting output coding, Bagging and boosting. Ensembles are often considered to be better than any single reveiwed for three fundamental reasons[1]. Those fundamental issues are the most important ways in which existing learning algorithms fail. But ensemble methods have the promise of reducing these key shortcomings of standard learning algorithms. These issues will be elaborated later in this section.

Let's begin the introduction of ensenmble methods from the standard supervised learning problem. In some unknown function $y = f(x)$, the training examples of a learning program is given in the form of $(\mathbf{x}_1, y1), \ldots, (\mathbf{x}_m, y_m)$. The $\mathbf{x}_i$ values are typically vectors of the form $\langle x_{i,1}, x_{i,2}, \ldots, x_{i,n} \rangle$, whose components are called the *features* of $\mathbf{x}_i$.

Given a set S of training examples, a learning algorithms can output a *classifier*, which is a hypothesis about the true function f. With new $\mathbf{x}$ values, the classifier predicts the corresponding y values. An **ensemble of classifier** is a set of classifiers

[1] Including **statistical**, **computational** and **representiational**.

whose individual decisions are combined in some way, typically by weighted or un weighted voting, to classify new examples.

Accuracy and **diversity** are two necessary and sufficient condition for and an ensemble of classifiers to be mor accurate than any of its individual members. Accuracy means that the classifier has an error rate of better than random guessing on new $\mathbf{x}$ values. Diversity means each classifier makes different errors on new data points. The importance of accuracy is intuitive. To demonstrate the importance of diversity, imagine the case that now we have three totally mutually indepentdent classifier with identical error rates equal to 0.3. We predict y as *true* if at least two of the three classifiers predict true. Then, the overall error rate is $0.3^3 + 3 \times 0.3^2 \times 0.7 = 0.216 < 0.3$. More precisely, if the error rates of L hypotheses h_l are all equal to p, then the probability that the majority vote will be rong will be the area under the binomial distribution where more than $L/2$ hypothese are wrong.

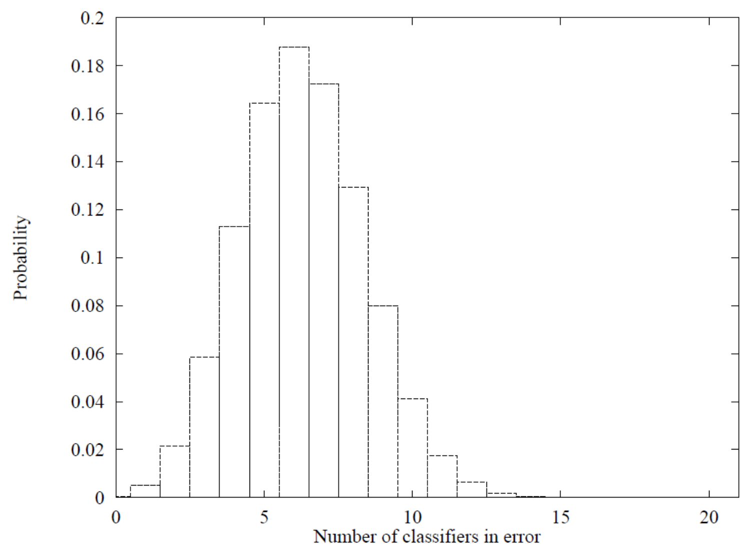

Figure 1: The probability that exactly l hypotheses will make an error, assuming each hypothesis has an error rate of 0.3 and makes its errorss independently of the other hypotheses.

As we mentioned before, there are three fundamental reasons make it often possible to construct very good ensembles, including:

1. **Statistical:** Ensembles can "average" acurate classifiers' votes and reduce the risk of choosing the wrong classifier.

2. **Computational:** An ensemble constructed by running the local search from many different starting points may provide a better approximation to the true unknown function.

3. **Representational:** By forming weighted sums of hypotheses it may be possible to expand the space of representable functions.

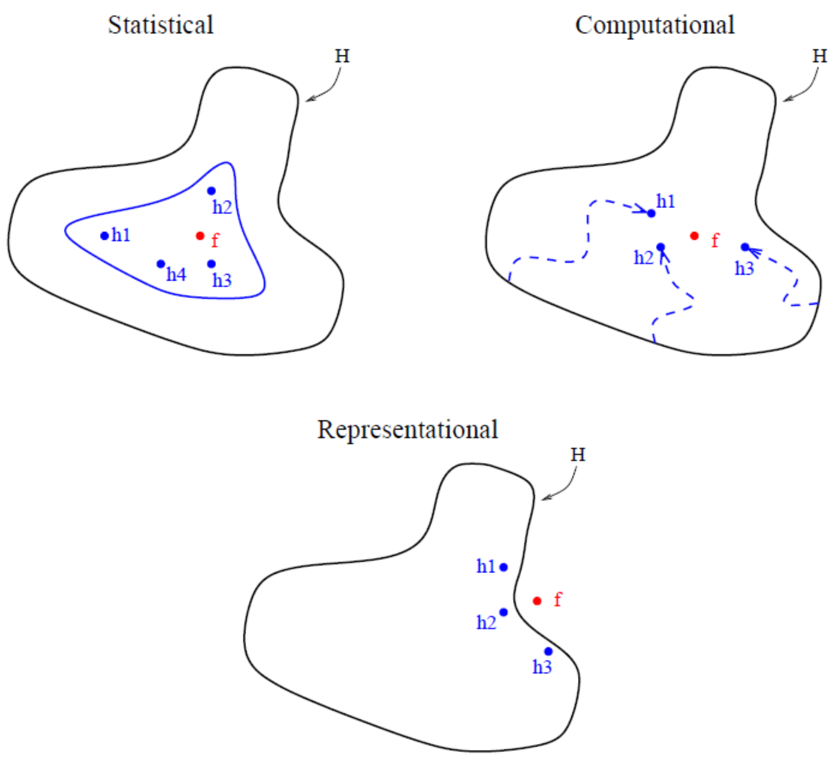

Figure 2: Three fundamental reasons why an ensemble may work better than a single classifier.