A Brief Introduction to Boosting

Boosting is an algorithm used mainly for reducing bias and variance in a supervised learning technique. It is a meta-algorithm that is used by those machine learning algorithms that converts a weak learner to a strong one. In this paper we get to understand one of the boosting algorithms called AdaBoost. The paper explains the underlying concepts of boosting and the reasons why in boosting we are not worried about the overfitting.

AdaBoost solves the previous difficulties seen in the boosting algorithms. The AdaBoost algorithm takes the samples without labels from one set and labels from another set for training. Weak learners are called repeatedly in turns based on a specified condition.

The set of weights distribution is maintained over the training set. This forms the basic principle of the AdaBoost which enables the algorithm to avoid being overfit.

The paper concludes by discussing a few experiments that are being performed. One of the advantages of the AdaBoost is that it is fast and simple to implement. It does not require the prior knowledge of the weak learner although it’s not ideal to use when there’s noisy data or insufficient data. The performance of boosting is dependent on the data and the weak learner.

The algorithm does a good job in identifying the outliers and can be used to train the mislabeled data which otherwise is difficult to categorize.

Boosting outperforms the other tested methods in the experiments talked about in the paper.