Lecturer: Prof. Homrighausen Scribe: Aaron Nielsen

1 AdaBoost

AdaBoost is a boosting algorithm developed by Yoav Freund and Robert Schapire in 1996.

1.1 Basics of AdaBoost

- The goal of boosting is to use a set of weak learners to create a single strong learner.
- Schapire (1990) introduced simple boosting.
- Freund (1995) built on Schapire (1995) to show that it is possible to simultaneously combine many weak learners.
- AdaBoost was invented in 1996 and the outline of the algorithm was outlined in Freund and Schapire (1997).
- AdaBoost is adaptive as weak learners are adjusted for the instances where misclassification occurred by previous classifiers
- As long as the individual classifiers are better than pure guessing, AdaBoost can provide a strong learner

1.2 AdaBoost Algorithm

1. Initialize $w_i \equiv 1/n$
2. For $m = 1, \ldots, M$
   (a) Fit $g_m(x)$ on $\mathcal{D}$, weighted by w_i
   (b) Compute $R_m = \frac{\sum_{i=1}^n w_i \mathbf{1}(Y_i \neq g_m(X_i))}{\sum_{i=1}^n w_i}$
   (c) Find $\beta_m = \log((1 - R_m)/R_m)$
   (d) Set $w_i \leftarrow w_i \exp\{\beta_m \mathbf{1}(Y_i \neq g_m(X_i))\}$
3. Output: $g(x) = \operatorname{sgn}\left(\sum_{m=1}^M \beta_m g_m(x)\right)$

The above algorithm is also known as Discrete AdaBoost since it is using discrete labels.

1.3 Simulation Results

- To illustrate the merits of AdaBoost, we will consider a particularly tricky classification example.
- A diagonal boundary between two groups is difficult for a small classification tree to handle.
- AdaBoost can combine a series of small trees to create a stronger classifier.
- In this example, we will look at "depth 2-stumps" which have no more than four terminal nodes.

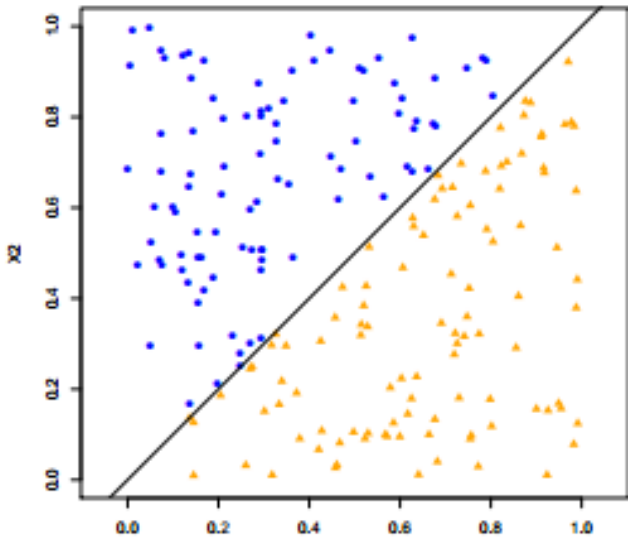

Figure 1: Diagonal Decision boundary is typically difficult for classification trees to handle

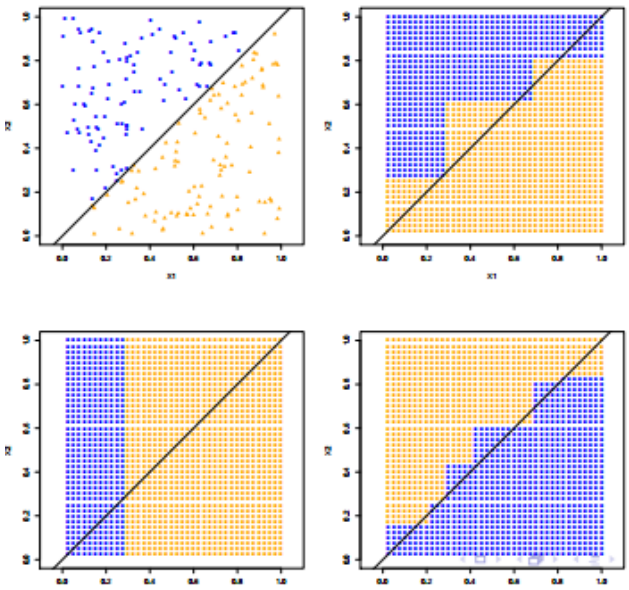

Figure 2: AdaBoost refining the classification method by adapting to misclassifications

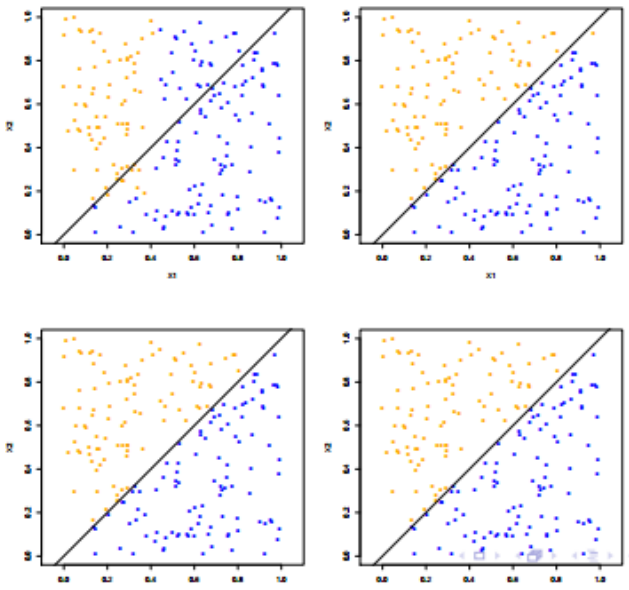

Figure 3: The misclassification rate decreases as the algorithm progresses

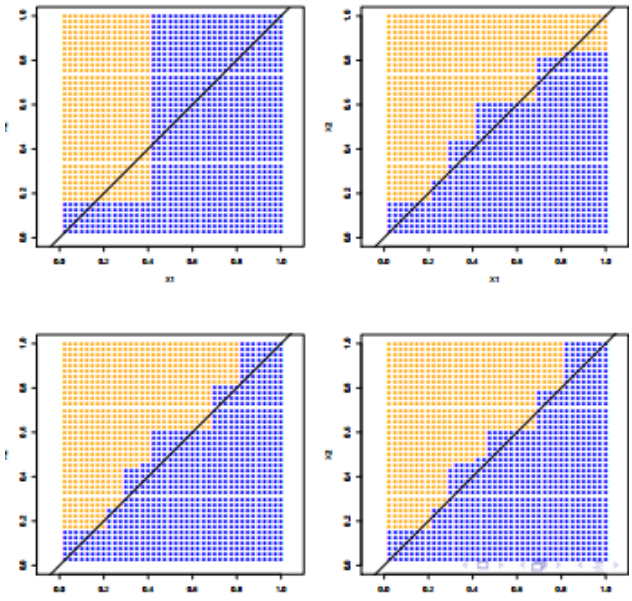

Figure 4: A lower misclassification rate is achieved

1.4 Real AdaBoost Algorithm

The AdaBoost algorithm can be modified to estimate the contribution of the m^{th} classifier, rather than the estimated label.

1. Initialize $w_i \equiv 1/n$
2. For $m = 1, \ldots, M$
   (a) Fit the classifier on $\mathcal{D}$, weighted by w_i and produce $p_m(x) = \hat{P}_w(Y = 1|x)$
   (b) Set $g_m(x) \leftarrow \frac{1}{2}\log(p_m/(1 - p_m(x)))$
   (c) Set $w_i \leftarrow w_i \exp\{-Y_i g_m(X_i)\}$
3. Output: $g(x) = \text{sgn}\left(\sum_{m=1}^{M} g_m(x)\right)$

2 Additive Models

2.1 Introduction

- Additive Models are a nonparametric regression method.
- We have previously modeled nonlinearity using constrained linear regression.
- Nonparametric methods form different types of local averages of the Y values of points near each other
- This can work well when p is small
- When p is large, we have to deal with the curse of dimensionality

The Curse of Dimensionality and methods to counter the issue will be examined in the next set of notes.