

# Statistical Machine Learning Homework 4

Credit to Yuling Yao

## Problem 1

The following figures show the corresponding partition and the tree in part a and b.

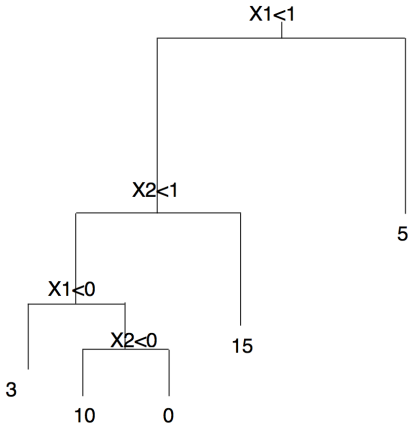


Figure 1: A Tree corresponding to part a.

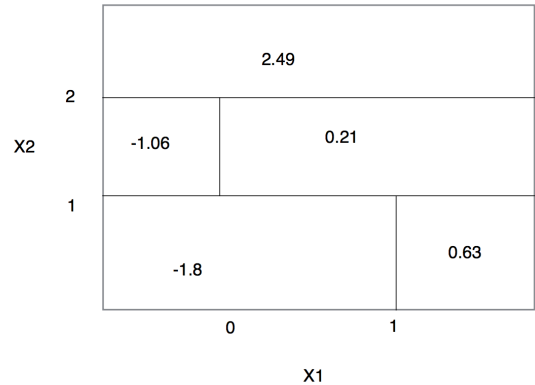


Figure 2: A partition of the feature space corresponding to part b.

## Problem 2

1. Based on majority vote

$$\sum_{k=1}^{10} I(p_k > 0.5) = 6$$

, so the final classification is Red.

2. Based on the average probability

$$\hat{p} = \frac{1}{10} \sum_{k=1}^{10} p_k = 0.45$$

, so the final classification is Green.

## Problem 3

We implement the AdaBoost algorithm as described in the question. The code can be found in the attachment. The following figure shows the training error and the testing error in 100 interation. As expected, when there are more and more iterations, the training error eventually goes to 0, while testing error won't.

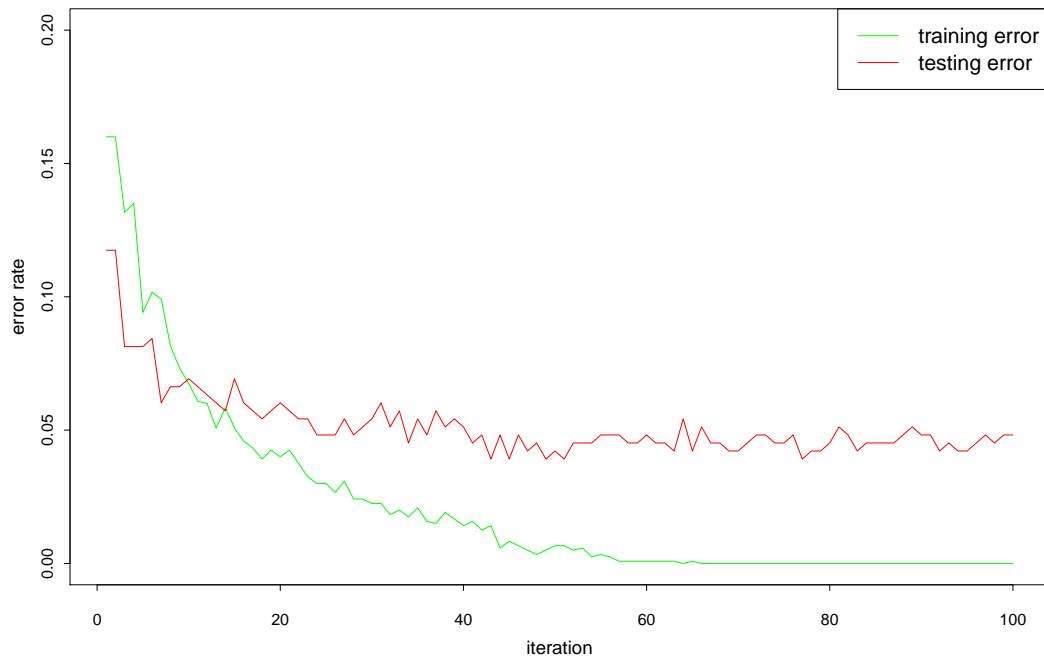


Figure 3: Training and Testing error of the AdaBoost algorithm on the ZIP code data.