Statistical Learning - Homework 7 (Conceptual)

James Hahn

Chapter 8 - Exercise 2

We first make the assumption that $\hat{f}(x) = 0$ and let $\hat{f}^1(x) = c_1 I(x_1 < t_1) + c_1' = \frac{1}{\lambda} f_1(x_1)$ be the first step of the boosting algorithm. Then, $\hat{f}(x) = \lambda \hat{f}^1(x)$ and $r_i = y_i - \lambda \hat{f}^1(x_i) \quad \forall i$.

Next, we have $\hat{f}^2(x) = c_2 I(x_2 < t_2) + c_2' = \frac{1}{\lambda} f_2(x_2)$ for the second step of the boosting algorithm.

In order to maximize the fit to the residuals, a new, unique/distinct stump must be generated and fit. So, $\hat{f}(x) = \lambda \hat{f}^1(x) + \lambda \hat{f}^2(x)$ and $r_i = y_i - \lambda \hat{f}^1(x_i) - \lambda \hat{f}^2(x_i) \quad \forall i$. So, finally, we have

$$\hat{f}(x) = \sum_{j=1}^{p} f_j(x_j)$$

This is the additive model discussed in the question.

Chapter 8 - Exercise 4

  a) The tree can be seen below:

  b)

```
par(xpd = NA)
plot(NA, NA, type = "n", xlim = c(-2, 2), ylim = c(-3, 3), xlab = "X1", ylab = "X2")

# X2 < 1
lines(x = c(-2, 2), y = c(1, 1))

# X1 < 1 with X2 < 1
lines(x = c(1, 1), y = c(-3, 1))
text(x = (-2 + 1)/2, y = -1, labels = c(-1.8))
text(x = 1.5, y = -1, labels = c(0.63))
```
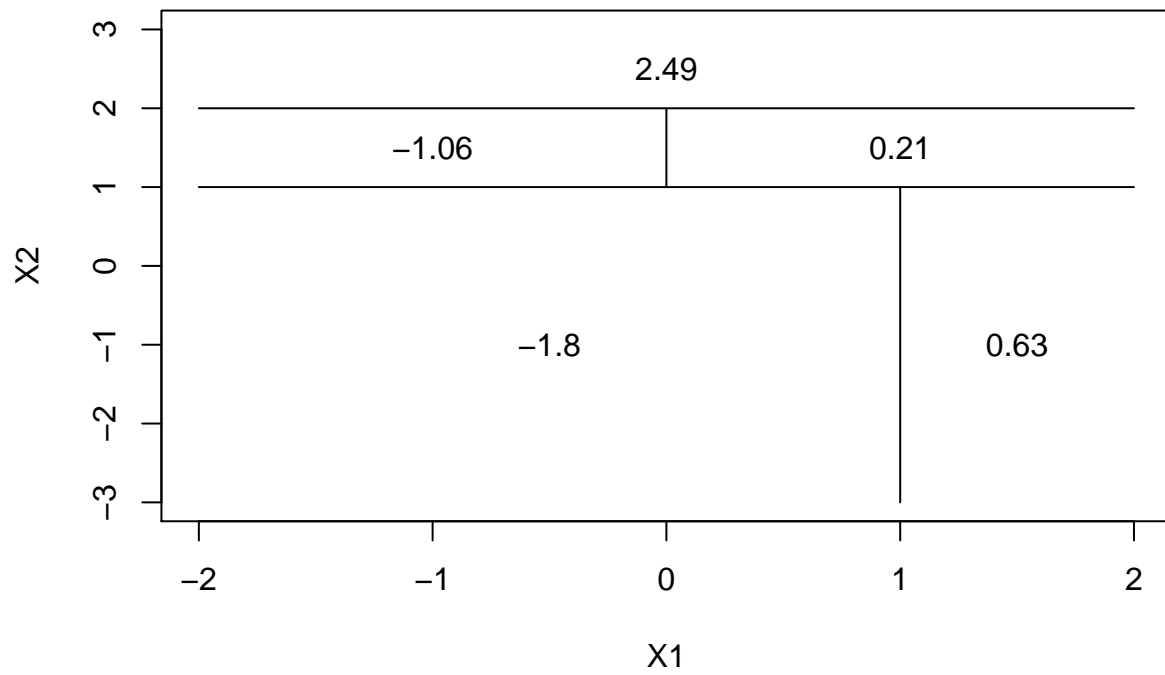
```
# X2 < 2 with X2 >= 1
lines(x = c(-2, 2), y = c(2, 2))
text(x = 0, y = 2.5, labels = c(2.49))

# X1 < 0 with X2<2 and X2>=1
lines(x = c(0, 0), y = c(1, 2))
text(x = -1, y = 1.5, labels = c(-1.06))
text(x = 1, y = 1.5, labels = c(0.21))
```


Chapter 8 - Exercise 5

With majority vote, we classify X as red since it occurs most often among all 10 predictions (6 red and 4 green). With average probability, we classify X as green since the average of the 10 probabilities is 0.45.