

## Chapter 8

1. Draw an example (of your own invention) of a partition of two-dimensional feature space that could result from recursive binary splitting. Your example should contain at least six regions. Draw a decision tree corresponding to this partition. Be sure to label all aspects of your figures, including the regions  $R_1, R_2, \dots$ , the cutpoints  $t_1, t_2, \dots$ , and so forth.<sup>1</sup>

Please see the attached sheet.<sup>2</sup>

3. Consider the Gini index, classification error, and cross-entropy in a simple classification setting with two classes. Create a single plot that displays each of these quantities as a function of  $\hat{p}_{m1}$ . The x-axis should display  $\hat{p}_{m1}$ , ranging from 0 to 1, and the y-axis should display the value of the Gini index, classification error, and entropy.<sup>3</sup>

Please see the attached sheet.

4. This question relates to the plots in Figure 8.12.

- (a) Sketch the tree corresponding to the partition of the predictor space illustrated in the left-hand panel of Figure 8.12. The numbers inside the boxes indicate the mean of  $Y$  within each region.

Please see the attached sheet.

- (b) Create a diagram similar to the left-hand panel of Figure 8.12, using the tree illustrated in the right-hand panel of the same figure. You should divide up the predictor space into the correct regions, and indicate the mean for each region.

Please see the attached sheet.

5. Suppose we produce ten bootstrapped samples from a dataset containing red and green classes. We then apply a classification tree to each bootstrapped sample and, for a specific value of  $X$ , produce 10 estimates of  $P(\text{Class is Red} \mid X)$ :

0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, and 0.75.

---

<sup>1</sup>Hint: Your result should look something like Figures 8.1 and 8.2.

<sup>2</sup>I wish I had time to put this all into **R** and make it look beautiful (and easier for me to reference later) but I was just a little too busy with my thesis :(

<sup>3</sup>In a setting with two classes,  $\hat{p}_{m1} = 1 - \hat{p}_{m2}$ . You could make this plot by hand, but it will be much easier to make in **R**.

*There are two common ways to combine these results together into a single class prediction. One is the majority vote approach discussed in this chapter. The second approach is to classify based on the average probability. In this example, what is the final classification under each of these two approaches?*

**Average approach** You would take the mean of the above dataset and would get that  $\text{mean}(\text{dataset}) = 0.45$ . Since this is less than the 50%, the threshold, you would find that you would obtain **green** as the result for this approach.

**Majority approach** Here you get a different outcome. You would find if the number of observations that are greater than or equal to 0.5 are more common than the number of observations that are less than 0.5. You end up finding out that the number of observations that are greater than or equal to 0.5 are larger than those that are smaller than 0.5. Therefore you would classify this as **red**.