

5. Suppose we produce ten bootstrapped samples from a data set 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  $P(\text{Class is Red}|\text{X})$ :

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

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?

**Answer:**

Approach #1- The majority vote approach involves choosing the most commonly occurring class among the 10 estimates. If we assume that the class assigned is green if  $P(\text{Class is Red}|\text{X}) < 0.5$  and the class assigned is red if  $P(\text{Class is Red}|\text{X}) \geq 0.5$ , then there are 4 green predictions and 6 red predictions. Therefore, based on the majority vote approach the final classification would be **red**.

Approach #2- If we classify based on the average probability, we sum up all the estimates of  $P(\text{Class is Red}|\text{X})$  then divide by 10:  $\frac{0.1 + 0.15 + 0.2 + 0.2 + 0.55 + 0.6 + 0.6 + 0.65 + 0.7 + 0.75}{10} = 0.45$ . Since the average probability of  $0.45 < 0.50$ , and the class assigned is green if  $P(\text{Class is Red}|\text{X}) < 0.5$ , the final classification would be **green**.