WU #15 - Bagging

Math 154 - Jo Hardin

Thursday, November 4, 2021

Name: _____

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(Class is Red$|X$):

$$0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, 0.75$$

There are two common ways to combine these results together into a single class prediction. One is the majority vote. 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?

Solution:

Majority vote:

In 6 out of the 10 trees, the value of X was predicted to be Red. The predicted value is **RED**.

Average Probability:

```
mean(c(0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, 0.75))
```

```
## [1] 0.45
```

The average predicted probability is 0.45, so the predicted value of X is **GREEN**.