Spring 2020

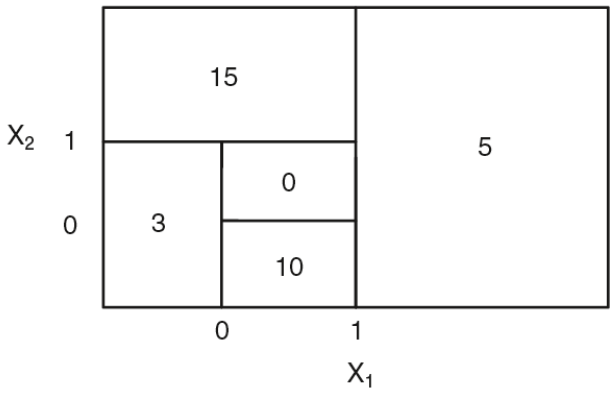
Quantitative Analysis
Chung-Ming Kuan

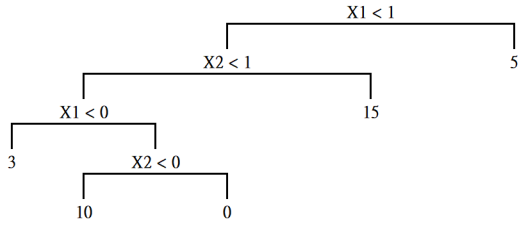
Problem Set 10

Due: 5/25

Part One: Hand-Written Exercise

1. Sketch the tree corresponding to the partition of the predictor space illustrated in the following figure. The numbers inside the boxes indicate the mean of Y within each region.





2. Suppose we have five equal sized data set containing red and green classes. We then apply a classification tree to each data set and, for a specific value of X, produce 5 estimates of $Pr(\text{Class is Red}|X)$: 0.1, 0.2, 0.55, 0.6, and 0.75. There are two common ways to "combine" these results together into a single class prediction. One is the majority vote approach. 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?

   (a) Majority Vote: because the number of predicted red class is 3 while green class is 2, the final classification is red class.

   (b) Average Probability: $\frac{0.1+0.2+0.55+0.6+0.75}{5} = 0.44 < 0.5$. The final classification is green class.