MATH 4385/5385 - Statistical Learning - Fall 2023
Department of Mathematics, University of New Orleans
Instructor: Xueyan "Sherry" Liu

Homework 4

MATH 4385/5385:

Problem 1: Chapter 8: (ISLR 2nd edition, Page 362) # 4.

4. This question relates to the plots in Figure 8.14.

   (a) Sketch the tree corresponding to the partition of the predictor space illustrated in the left-hand panel of Figure 8.14. The numbers inside the boxes indicate the mean of Y within each region.

   (b) Create a diagram similar to the left-hand panel of Figure 8.14, 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.

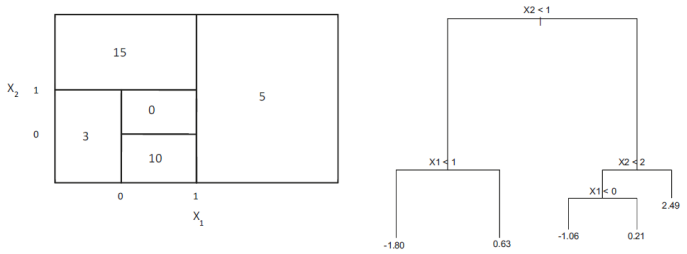


FIGURE 8.14. Left: *A partition of the predictor space corresponding to Exercise 4a.* Right: *A tree corresponding to Exercise 4b.*

Problem 2: Chapter 8: (ISLR 2nd edition, Page 362) # 5.

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 $P(\text{Class is Red}|X)$:

$$0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, \text{ and } 0.75.$$

   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?

Problem 3: **Chapter 8**: (ISLR 2nd edition, Page 362) # 6.

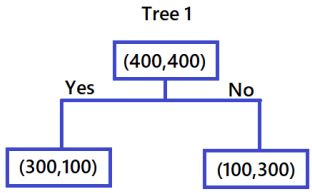
6. Provide a detailed explanation of the algorithm that is used to fit a regression tree.

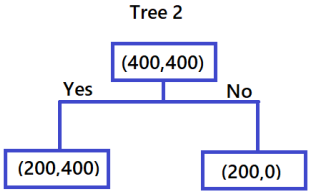
Note: The question is not clearly stated. On page 333 of ISLR 2, there is the Algorithm 8.1 on Building a Regression Tree, which is on a pruned tree and contains a few steps. Step 1 is about

growing a raw tree. From my point of view, I would suggest you to just describe how to grow a raw regression tree. But if you would like to include the K-fold CV approach for the pruning process, that is definitely fine.

When you describe how to grow a basic tree, you basically summarize materials of the book from Pg 328 to pg 331, or the notes Math_StatLearning_8_part1_annotated.pdf from page 1 to page 5. Make sure to include the key information, such as, recursive binary splitting (top-down and greedy search), prediction, splitting criteria, disjoint non-overlapping sub-regions, stopping criterion, etc.

Problem 4: And the following question on classification trees:
Consider a dataset with 400 observations in each of the two classes $(+1, -1)$ and the following 2 trees with a single split for each:





Note: in each pair of parentheses, we can think of the first component as the number of observations in the positive class $+1$ and the second component as the number of observations in the negative class -1.

(a) Calculate the information gain for each tree using misclassification error rate and Gini index, respectively.
(b) If we use the misclassification error rate, which tree is preferred? What if we use Gini index?
(c) (Bonus) What if using cross-entropy?