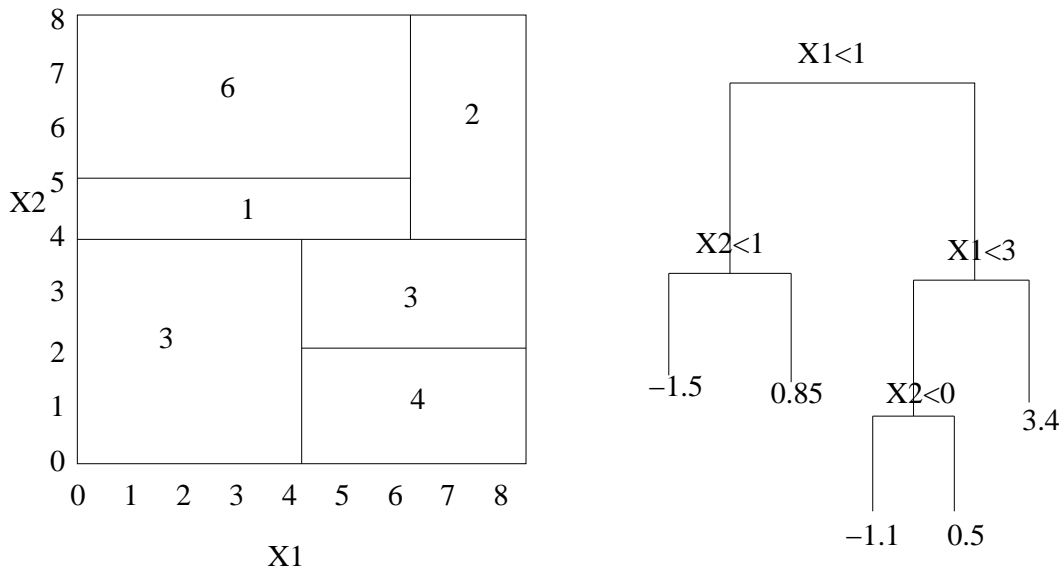


**CS 5565, HW7(Tree-Based Methods) 40pts.**

Name \_\_\_\_\_

1. (10 points total) This question relates to the plots in the figure below.
  - (a) (5 points) Sketch the tree corresponding to the partition of the predictor space illustrated in the left-hand panel of the Figure. The numbers inside the boxes indicate the mean of  $Y$  within each region.
  - (b) (5 points) Create a diagram similar to the left-hand panel of the figure, 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.



2. (10 points total) Consider the Gini index, classification error, and 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. (Note: you may use R to make this plot.)
3. (10 points total) 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  $P(\text{Class is Red}|X)$ :

0.25, 0.3, 0.35, 0.4, 0.4, 0.45, 0.8, 0.8, 0.85, and 0.9.

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?

4. (10 points) A company is trying to determine how to market two new products based on the age of the consumer. Consumers are asked to indicate Y or N if they would be willing to buy product A or B. The following table is given:

ID	Age	Product A	Product B
1	12	N	Y
2	13	N	Y
3	14	N	Y
4	16	Y	N
5	20	N	Y
6	24	Y	N
7	32	N	N
8	45	Y	N
9	65	Y	Y
10	73	Y	N
11	75	Y	N

Using the squared error, calculate the decision trees for determining the mean ages for those who would (or would not) buy product A and those who would (or would not) buy product B.