ER 131 - Additional Exam Practice Questions

1) For each of parts (a) through (d), indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify your answer.
   a) The sample size n is extremely large, and the number of predictors p is small.
   b) The number of predictors p is extremely large, and the number of observations n is small.
   c) The relationship between the predictors and response is highly non-linear.
   d) The variance of the error terms, i.e. $\sigma^2 = \text{Var}(\epsilon)$, is extremely high.

2) Explain the differences between the KNN classifier and KNN nearest neighbours methods.

3) Explain how k-fold cross-validation is implemented.

4) What are the advantages and disadvantages of k-fold cross-validation relative to:
   a) The randomized train and test set approach?
   b) Leave-one-out cross-validation?

5) Indicate which of (a) through (d) is correct. Lasso, relative to least squares, is:
   a) More flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.
   b) More flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.
   c) Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.
   d) Less flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.

6) This question relates to the plots in Figure 8.12, from ISLR.
   a) Sketch the tree corresponding to the partition of the predictor space illustrated in the left-hand panel of Figure 8.12. 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 8.12, 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.

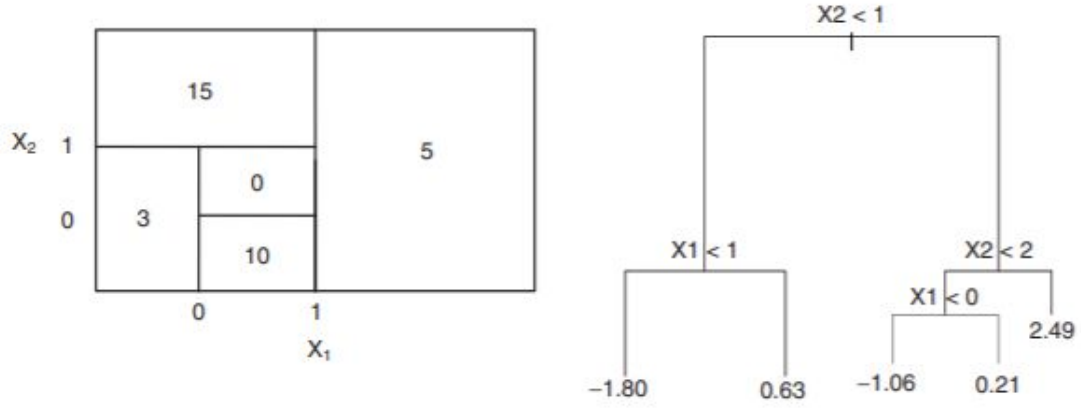

FIGURE 8.12. Left: *A partition of the predictor space corresponding to Exercise 4a.* Right: *A tree corresponding to Exercise 4b.*