

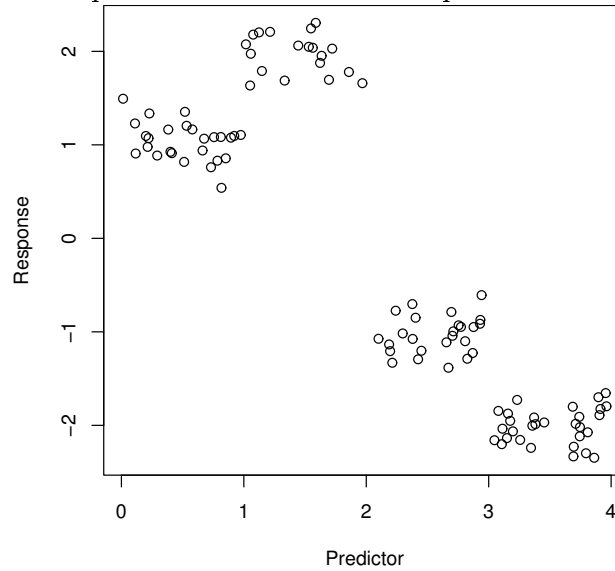
# 36-350: Data Mining

## Homework 8

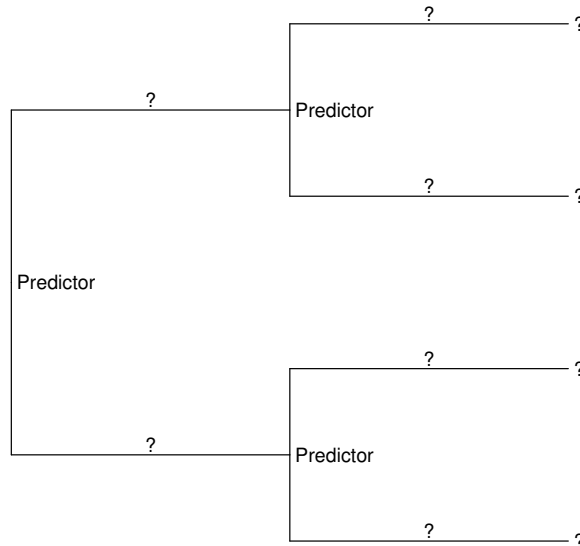
Date: October 13, 2003

Due: start of class October 20, 2003

1. Below is a scatterplot of a predictor variable and a response variable.

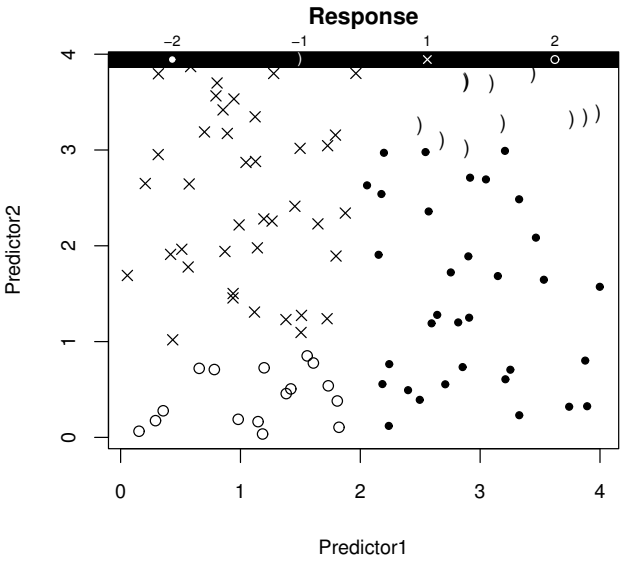


Draw a regression tree which divides this data into four subgroups with different mean responses. It should have the following shape:

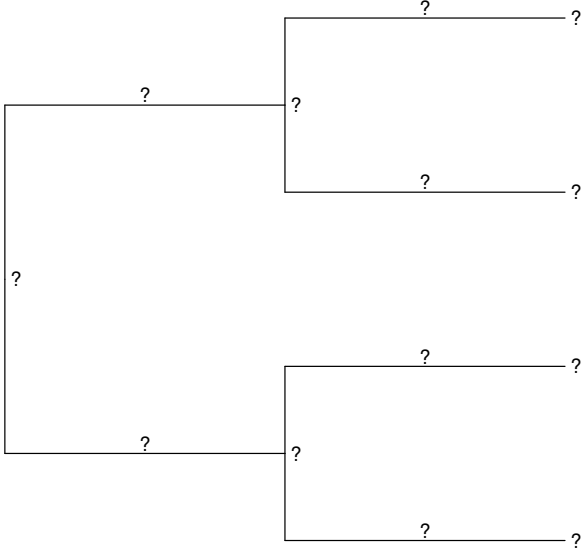


Replace the ?s with the appropriate branch conditions and group means. (See problem 3 for an example.) The splits should minimize the variance within groups, and should be ordered so that the most important split is at the top of the tree.

2. Below is a symbol plot of a response variable versus two predictors. (This is a different kind of plot than in the previous problem.)

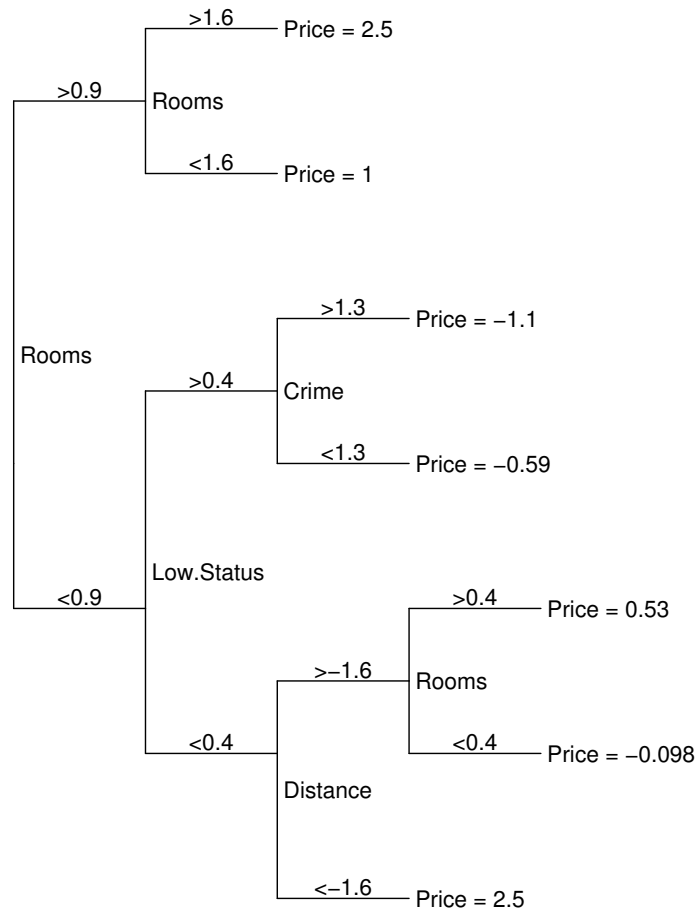


Draw a regression tree which divides this data into four subgroups with different mean responses. It should have the following shape:



Replace the ?s with the appropriate predictor names, branch conditions, and group means.

3. Below is a regression tree to predict Price in the neighborhood data from lab 7.

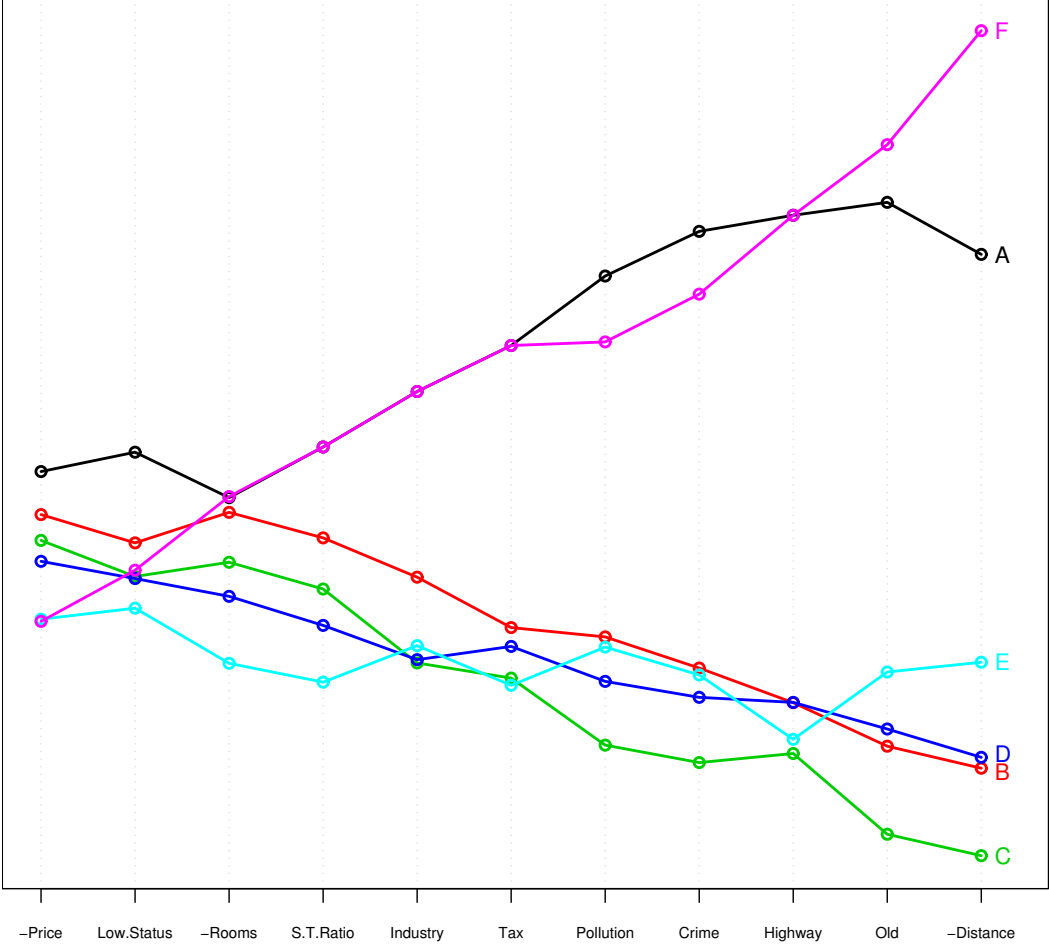


(a) According to the tree, what is the expected house price for the following neighborhood?

Crime	-1.3
Industry	-0.6
Pollution	-0.7
Rooms	0.2
Old	0.4
Distance	0.8
Highway	-0.9
Tax	-1.0
Student.Teacher.Ratio	-0.3
Low.Status	-0.4

(b) Data collection usually costs time and money. Explain why predicting house prices via the tree is more economical than using linear regression, i.e. taking a weighted sum of the neighborhood variables.

4. The regression tree in the previous problem defines seven neighborhood groups. Below is a parallel-coordinate plot of the group prototypes:



Note that three of the variables have been negated, including Price, so that down generally means “better”.

- (a) One of the group prototypes is missing from the plot. Which one? (Give its expected price.)
- (b) Regression trees are good at finding groups with unusual response values. One of the prototypes in the plot can be described as having overpriced homes, while another has underpriced homes. Find them and explain your choice.