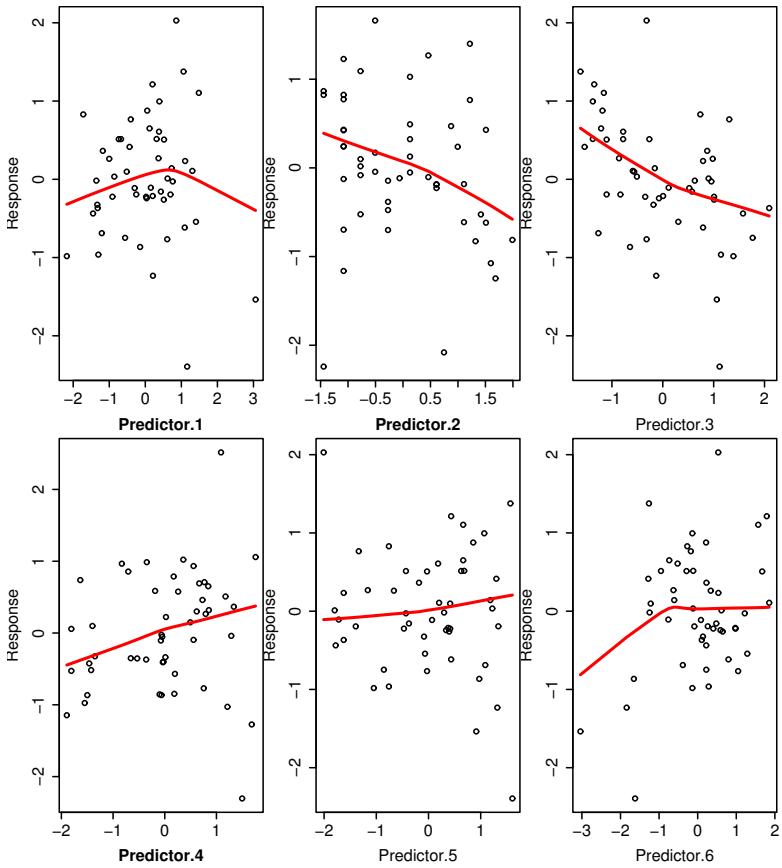
36-350: Data Mining

Homework 9
Date: October 20, 2003 **Due: start of class October 27, 2003**

---

1. Suppose you've constructed a linear model to predict a response from a few predictors. To decide if more or fewer predictors should be used, you make a plot of partial residuals, which turns out as follows:



Predictors which are not in the model are plotted versus the residuals of the current model. Predictors which are already in the model (1,2, and 4, written in bold) are plotted versus the residuals of a smaller model with that predictor removed.

(a) Name one predictor that should be added to the current model.

(b) Will this addition improve the R^2 value of the fit?

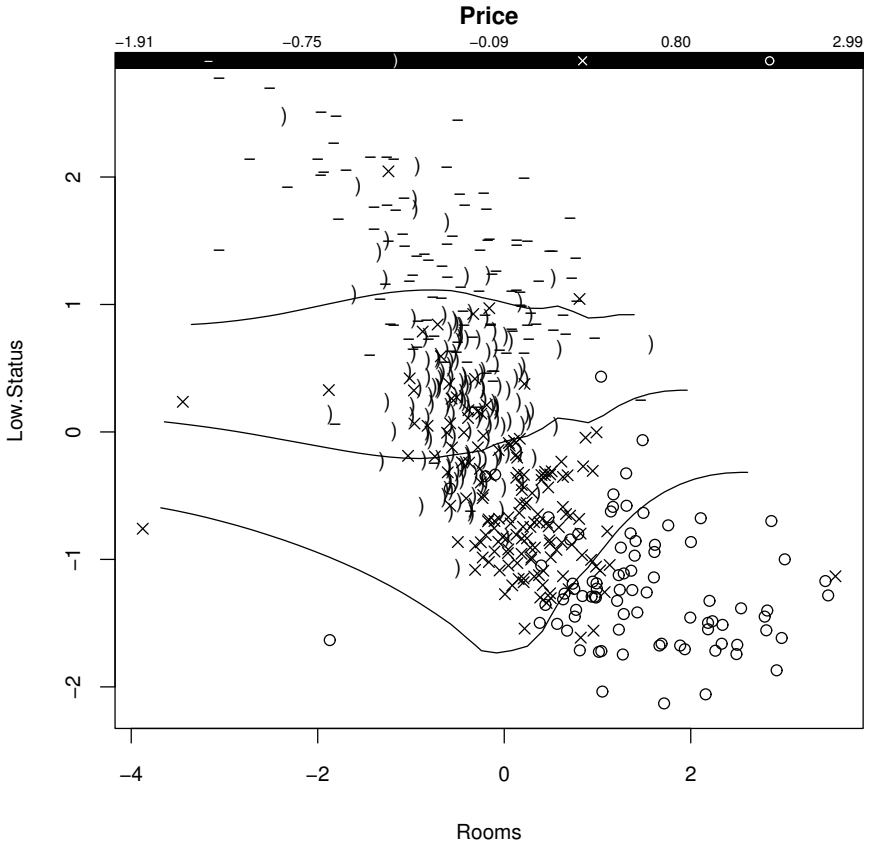
(c) Name one predictor that should be removed from the model.

(d) Will this removal improve the R^2 value of the fit?

(e) If a new predictor is added to the model, could this make other predictors in the model less important? Could this make other predictors not in the model more important?

(f) Is it possible that, after a series of additions and deletions using the above strategy, the best model could involve the opposite set of predictors (`Predictor.3`, `Predictor.5`, and `Predictor.6`)?

2. Below is a contour plot of `Price` versus `Rooms` and `Low.Status` for the neighborhood data of lab 7:



The variables have been standardized to have zero mean and unit standard deviation. Suppose you fit a linear model to predict `Price` via least squares. It turns out to have the following coefficients:

```
(Intercept)       Rooms  Low.Status
       0.00        0.32       -0.58
```

(a) Suppose `Low.Status` is 0.5 and `Rooms` is -2. What is the predicted `Price` (in standardized units)?

(b) Now suppose `Low.Status` is 0.5 and `Rooms` is 0. What is the predicted `Price` (in standardized units)?

(c) Explain why a linear model is misleading for this dataset.

3. In the computer lab, you constructed a linear model to predict the sales of one product as a function of the prices of other products.

(a) In the model found by stepwise selection, which two predictors are most important?

(b) In the model found by stepwise selection, which predictor is least important (and perhaps should be removed)?

(c) Explain what the linear model tells you in terms of competition and cooperation between these products in the store.