Introduction to Statistical Learning, Spring 2016

Homework exercise 3

Due date: 24 May 2016 in class

1. Recall that we defined in class the *Fixed-X* setting as one where the prediction is done at the same x values as in training. So for prediction we dimply draw a new vector Y^{new} at the same x values as the training vector Y. We showed that in this setting:

$$\mathbb{E}\frac{1}{n}\sum_{i=1}^{n}(y_i^{new}-\hat{y}_i)^2 - \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \frac{2}{n}\sum_{i=1}^{n}Cov(y_i,\hat{y}_i).$$

   Accordingly, we termed the expression on the right *optimism*.

   Now assume we are given a modeling problem with a given data size n, and we have two modelers proposing modeling approaches:

   - Modeler A: I will calculate the average of the training vector Y and that will be my prediction: $\hat{Y} = \bar{Y}$.
   - Modeler B: I will take the first response from the training data, and that will be my prediction for all future observations: $\hat{Y} = y_1$.

   In both cases, all coordinates of $\hat{Y}$ will be identical and equal to the choice described.

   (a) Which modeler will have lower training RSS? Explain.

   (b) Calculate the optimism of each modeling approach, and compare their optimism.

   (c) Which modeler will have lower (Fixed-X) expected prediction error? Explain.

2. Assume we have a problem with a single (scalar) predictor variable x. We turn it into a problem with two predictor variables by replicating x. In other words, in our training set we will have $x_{i1} = x_{i2}$ for all i. Now we fit a linear regression model with no intercept, using ridge or lasso penalty, which minimize RSS + $\lambda\cdot$ penalty. We want to understand what the solution will look like in each case.

   (a) Prove that for ridge regression, where the penalty is $\sum_j(\beta_j)^2$, the optimal solution will always have $\hat{\beta}_1 = \hat{\beta}_2$.
   Hint: You can prove it by negation, by showing that if the solution is different you can improve its penalized loss.

   (b) For lasso, where the penalty is $\sum_j|\beta_j|$, prove that for a given value β, every solution for which $\beta_1+\beta_2=\beta$ and β_1,β_2 have the same sign (positive or negative), has the same penalized loss. Hence conclude that this lasso problem has many different optimal solutions (explain why!).

3. Assume we have a 10-variable regression problem. We run all-subsets, forward stepwise and backward stepwise regression on it, and keep the chosen models with 5 and 7 variables from each method (total 6 models).

(a) Which 5-variable model will have the lowest training RSS?

(b) Which 5-variable model will have the lowest prediction (test) RSS?

(c) For which of the methods are we guaranteed that the 5-variables in the 5-variable model are a subset of the 7 variables in the 7-variable model?

4. ISLR 6.8 (p. 262): examining the performance of subset selection and lasso on simulated data.