Chapter 6 HW

Fabiani Rafael

Conceptual Questions

1. We perform best subset, forward stepwise, and backward stepwise selection on a single data set. For each approach, we obtain $p+1$ models, containing $0, 1, 2, \ldots, p$predictors. Explain your answers:

(a) Which of the three models with k predictors has the smallest *training* RSS?

(b) Which of the three models with k predictors has the smallest test RSS?

(c) True or False:

i. The predictors in the k-variable model identified by forward stepwise are a subset of the predictors in the $(k+1)$ variable model identified by forward stepwise selection.

ii. The predictors in the k-variable model identified by backward stepwise are a subset of the predictors in the $(k+1)$ variable model identified by backward stepwise selection.

iii. The predictors in the k-variable model identified by backward stepwise are a subset of the predictors in the (k + 1) variable model identified by forward stepwise selection.

iv. The predictors in the k-variable model identified by forward stepwise are a subset of the predictors in the $(k+1)$ variable model identified by backward stepwise selection.

v. The predictors in the k-variable model identified by best subset are a subset of the predictors in the $(k+1)$ variable model identified by best subset selection.

For parts (a) through (c), indicate which of i. through iv. is correct. Justify your answer.

(a) The lasso, relative to least squares, is:

i. More flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.

ii. More flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.

iii. Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.

iv. Less flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.

(b) Repeat (a) for ridge regression relative to least squares.

(c) Repeat (a) for non-linear methods relative to least squares.

Applied Questions

9. In this exercise, we will predict the number of applications received using the other variables in the College data set.

(a) Split the data set into a training set and a test set.

(b) Fit a linear model using least squares on the training set, and report the test error obtained.

(c) Fit a ridge regression model on the training set, with λ chosen by cross-validation. Report the test error obtained.

(d) Fit a lasso model on the training set, with λ chosen by cross validation. Report the test error obtained, along with the number of non-zero coefficient estimates.

(e) Fit a PCR model on the training set, with M chosen by cross validation. Report the test error obtained, along with the value of M selected by cross-validation.

(f) Fit a PLS model on the training set, with M chosen by cross validation. Report the test error obtained, along with the value of M selected by cross-validation.

(g) Comment on the results obtained. How accurately can we predict the number of college applications received? Is there much difference among the test errors resulting from these five approaches?