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, …, p predictors. Explain your answers:

1. Which of the three models with k predictors has the smallest training RSS?

**Answer**: Best subset because this method will fit many more models than the other methods, and therefore will find the model best suited to fit the training data.

1. Which of the three models with k predictors has the smallest test RSS?

**Answer**: Best subset will most likely have the lowest test RSS because this method contains the variables that are most explanatory to the response. However, by random, it is possible that the other methods have a better RSS.

1. True or False
   1. 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

**Answer**: True, forward selection adds one variable to the previous model, and selects the model with the lowest error

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

**Answer**: True, backwards selection removes the least important variable from the previous model.

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

**Answer**: False, there is no relationship between these two methods. Collinearity and variable interaction have influence over which variables are added/removed in each step.

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

**Answer**: False, see above

* 1. 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

**Answer**: False, best subset selects the best k+1 variables, and may not select all variables from the model using k variables.

2) For the next parts, identify which of i through iv are correct and justify.

1. The lasso, relative to least squares
   1. More flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance
   2. More flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias
   3. Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance
   4. Less flexible and hence will give improved prediction accuracy when its increase is less than its decrease in bias

**Answer**: iii

1. Repeat for ridge regression relative to least squares

**Answer**: iii, (same as the lasso)

1. Repeat for non linear methods relative to least squares

**Answer**: ii

3) Suppose we estimate the regression coefficients in a linear regression model by minimizing

for a particular value of s. For the following parts, indicate which of i through v is correct and justify.

1. As we increase s from 0, the training RSS will
   1. Increase initially, and then eventually start decreasing in an inverted U shape
   2. Decrease initially, and then eventually start increasing in a U shape
   3. Steadily increase
   4. Steadily decrease
   5. Remain constant

**Answer**: iv, s increasing implies more variables or more flexibility, which in turn implies a better accuracy

1. Repeat for rest RSS

**Answer**: ii, s increasing implies more flexibility, which will increase accuracy, however, as s gets too large, the model will overfit the training data, and the test RSS will begin to increase again

1. Repeat for variance

**Answer**: iii, s increasing implies the coefficient estimates will be more and more flexible, which will imply a larger variance of the estimates

1. Repeat for (squared) bias

**Answer**: iv, s increasing implies more flexibility, which implies a lower bias

1. Repeat for the irreducible error

**Answer**: v, the irreducible error is always constant and the model has no effect on it

4) Repeat 3, but suppose we estimate the regression coefficients in a linear regression model by minimizing

for a particular value of ƛ.

**Answer**: As lambda increases, flexibility decreases, therefore

1. Training RSS

**Answer**: iii

1. Test RSS

**Answer**: ii

1. Variance

**Answer**: iv

1. (squared) bias

**Answer**: iii

1. Irreducible error

**Answer**: v