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

the model with K predictors that have the smallest training RSS is the best subset selection because the model with K predictors is the one with smallest RSS among all the Cpk models.

On the other hand, In forward stepwise selection and backward stepwise selection the model with k predictors is the one with smallest RSS among the p-k models, the first increases the predictors by one and the other contains all of the predictors in Mk+1 except for one

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

Since the best subset accounts for more models, it has higher chance at getting the best RSS

1. True or False:
2. 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.

True- it is the increase of the K predictors in the model by 1

1. 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. 6.6 Exercises 283

True- we’re are removing one predictor from the model

1. 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.

False- there is no relation between the models generated in forward and backward selection

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 backward stepwise selection.

False

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 selection.

False

2- (a) The lasso, relative to least squares, is: iii. Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.

The lasso performs better than the least squares because it is a restrictive model so it doesn’t allow high bias therefore reduces overfitting and variance in predictions

(b) for ridge regression relative to least squares, is: iii. Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance.

The ridge regression is more restrictive than the least squares, therefore it doesn’t have as much bias to overfit the model

(c)for non-linear methods relative to least squares, is: ii. More flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias.

Non linear methods are more flexible than the least squares, as they try to fit the underlying data, therefore they have high variance and perform better when the bias is low