1.

(a)

The model with k predictors that has the smallest training RSS is the one that includes all k predictors, which means that it has the smallest bias.

(b)

We cannot determine which of the three models with k predictors has the smallest test RSS without testing them on the test data. However, in general, the model selected by best subset selection is expected to have the smallest test RSS because it considers all possible subsets of predictors.

(c)

i. True. In forward stepwise selection, the k+1 variable model is obtained by adding one more predictor to the k-variable model that minimizes the RSS. Therefore, the predictors in the k-variable model are a subset of the predictors in the (k+1)-variable model.

ii. True. In backward stepwise selection, the k-variable model is obtained by removing one predictor from the (k+1)-variable model that minimizes the RSS. Therefore, the predictors in the k-variable model are a subset of the predictors in the (k+1)-variable model.

iii. False. In backward stepwise selection, the k-variable model is obtained by removing one predictor from the (k+1)-variable model that minimizes the RSS. In contrast, in forward stepwise selection, the k+1 variable model is obtained by adding one more predictor to the k-variable model that minimizes the RSS. Therefore, the predictors in the k-variable model obtained by backward stepwise are not necessarily a subset of the predictors in the (k+1)-variable model obtained by forward stepwise selection.

iv. False. The predictors in the k-variable model obtained by forward stepwise selection are not necessarily a subset of the predictors in the (k+1)-variable model obtained by backward stepwise selection because the two methods use different criteria for selecting the predictors.

v. True. In best subset selection, the k-variable model is a subset of the (k+1)-variable model because the former considers all possible subsets of predictors.

2.

(a) ii

(b) iii

(c)

too general to give a specific answer. Non-linear methods can have different degrees of flexibility and regularization compared to linear models like least squares, depending on the specific method and parameters used.

3.

(a) iv, Steadily decrease

(b) ii, Decrease initially, and then eventually start increasing in a U shape

(c) iii, Steadily increase

(d) iv, Steadily decrease

(e) v, Remain constant

4.

(a) iii. Steadily increase.

(b) in general, as the model becomes less flexible with increasing λ, the test RSS is expected to decrease initially and then increase again in an inverted U shape.

(c) the variance is expected to steadily decrease as λ increases.

(d) the squared bias is expected to steadily increase as λ increases.

(e) it is not affected by the choice of regularization parameter λ, and will remain constant as λ increases.