HW2-Jiayuan Guo

Chapter 6, Problem 1

(a) Best subset selection with k predictors has the smallest training RSS. Because for the two stepwise selection,the model with k predictors is chosen dependent on the path (forward or backward) and may not be the model with smallest RSS in all the models. Only in best subset selection, the model with k predictors is the one with smallest RSS among all the models.

(b) It's hard to answer without enough information. Best subset selection takes more models into consideration, but stepwise selection may begin with the predictor and get a model with smaller test RSS by luck.

(c) i. True ii. True iii. False iv.False v. False

Chapter 6, Problem 3

(a) steadily decrease As s increase, the model complexity increase, so this model fits training dataset better and get smaller RSS

(b) ii. is right. As s increase, the model complexity increase, which initially better fit the model and get good performance on test dataset, but later cause overfitting and test RSS increase in the U shape.

(c) steadily increase is right. The more complex the model is, the larger the variance becomes.

(d) steadily decrease is right. The more complex the model is, the smaller the bias becomes

(e)remain constant is right. The irreducible error is independent of the model, and stay the same as the s changes.

Chapter 6, Problem 8

(c) We find that with C_p, BIC and Adjusted R^2 criteria, we all pick 3 variables in our models.
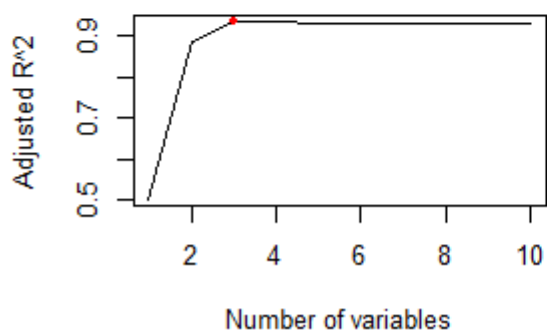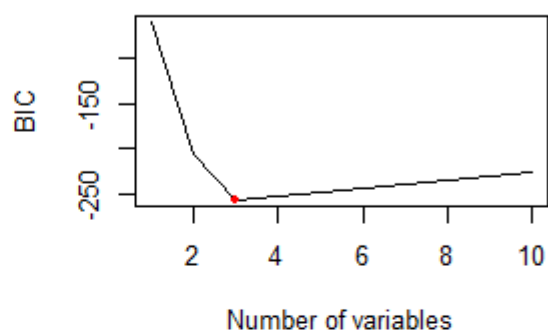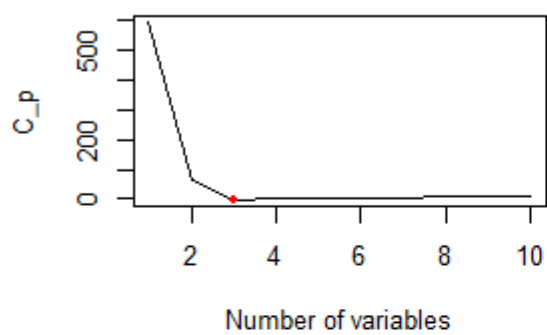
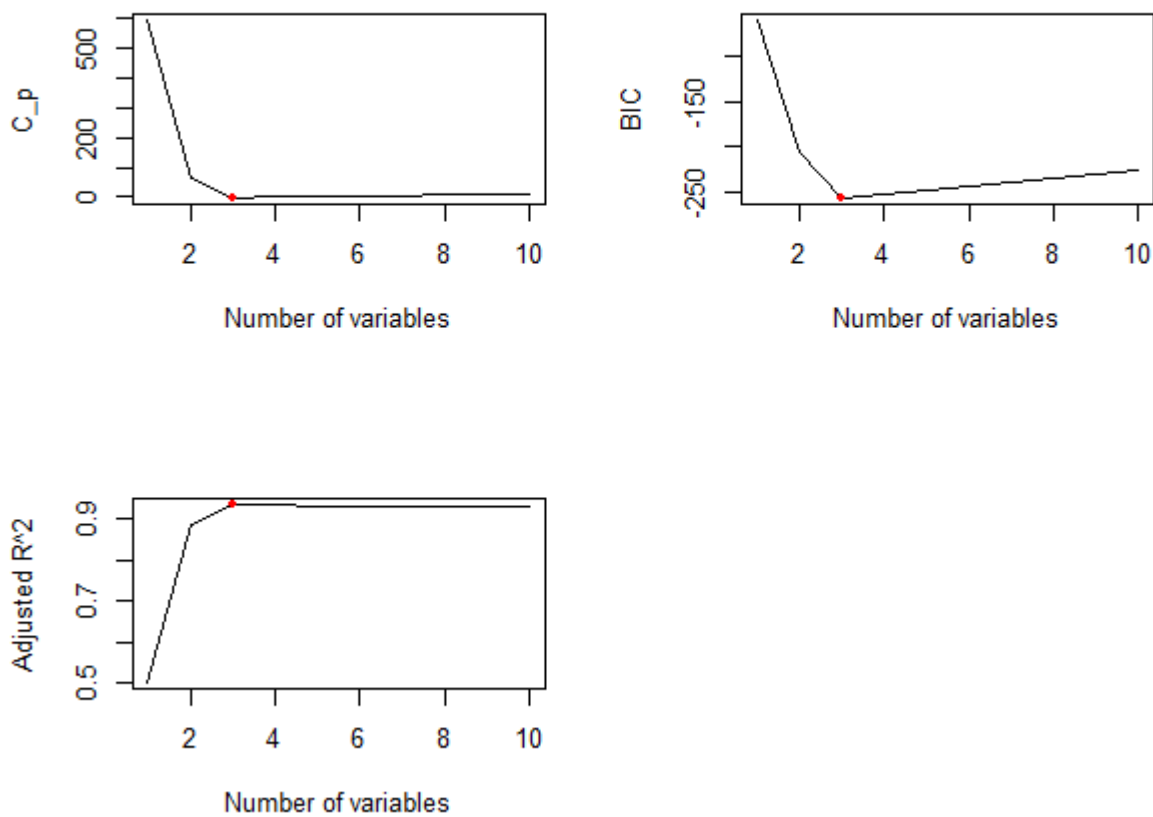(d) Forward stepwise, backward stepwise and best subset all select the three variables model with X, X^2, X^5.

Plots of C_p, BIC and adjusted R^2 for forward stepwise selection
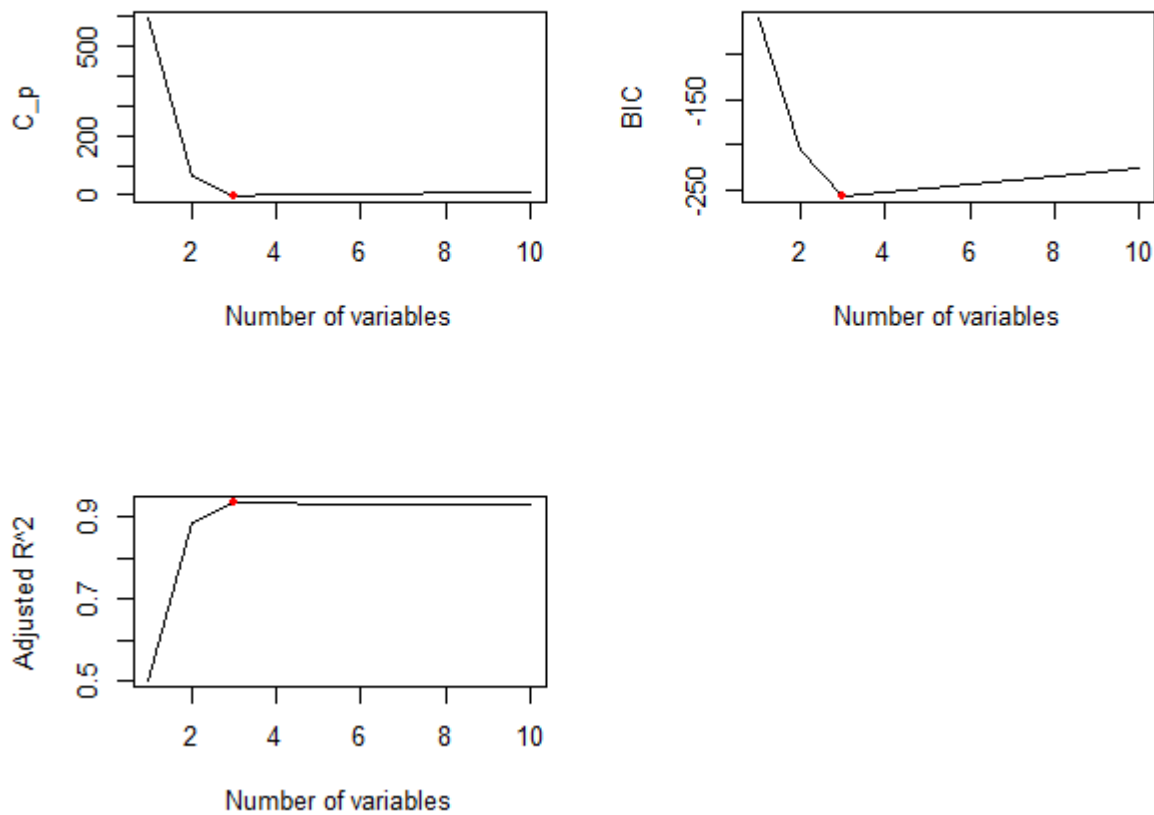

```
Forward Stepwise Output:
(Intercept) poly(x, 10, raw = T)1 poly(x, 10, raw = T)2
1.07219472             2.44514720             -2.15676236
poly(x, 10, raw = T)5
0.09022577
```

Plots of C_p, BIC and adjusted R^2 for backward stepwise selection


```
Backward Stepwise Output:
(Intercept) poly(x, 10, raw = T)1 poly(x, 10, raw = T)2
1.07219472              2.44514720              -2.15676236
poly(x, 10, raw = T)5
0.09022577
```
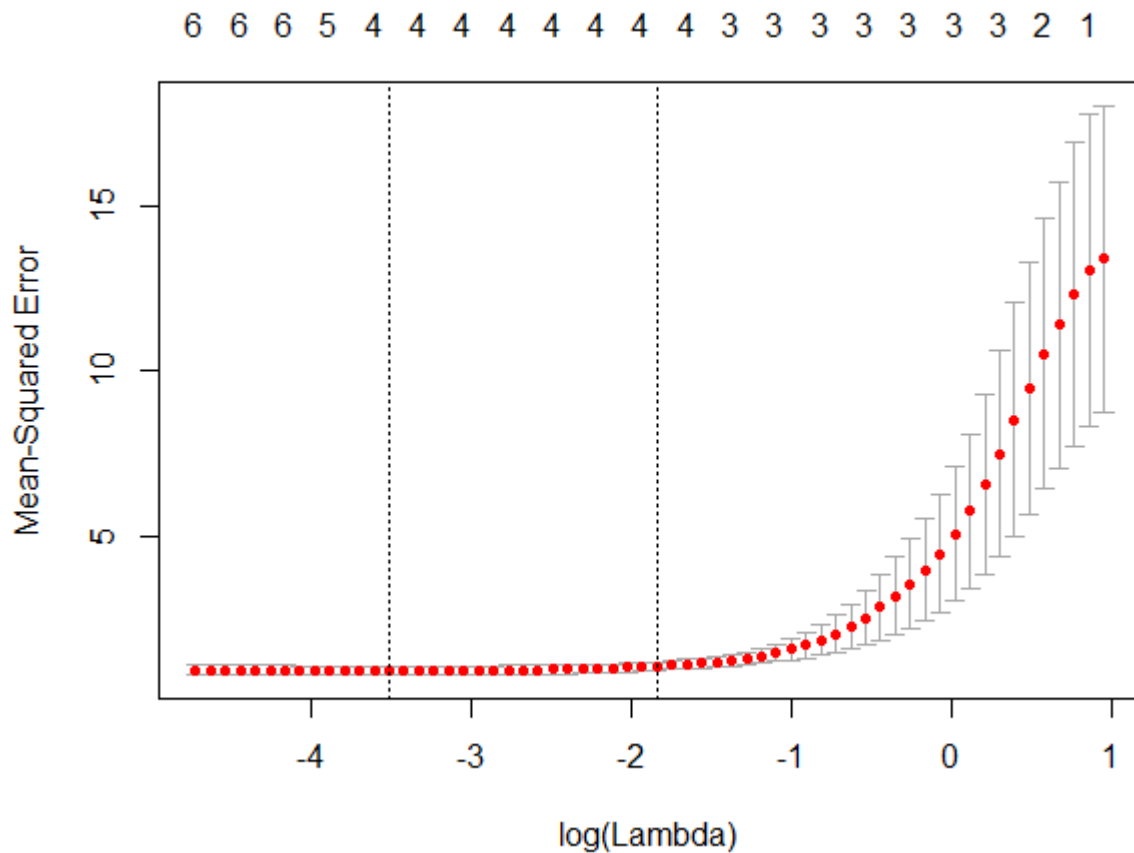
(e)The optimal value of lamda is 0.02980709. The lasso method picks X, X^2, X^3, X^5 as variables for the model.

(f) With C_p we pick the 2-variables model, with BIC we pick the model with the most accurate 1-variable, and with adjusted R^2 we pick the 4-variables model. Lasso also picks the best 1-variable model.

Chapter 6, Problem 9

(b)For a linear model using least squares, the test RSS is 1538442.

(c)For the ridge model, the optimal value of lambda is 18.73817, and test error is 1608859, which is higher than linear model with least square.

(d) For the ridge model, the test error is1635280, which is higher than ridge model

(e) For PCR, test RSS is about 3014496

(f) For PLS, test RSS is about 1508987

Test R-squared

(g)
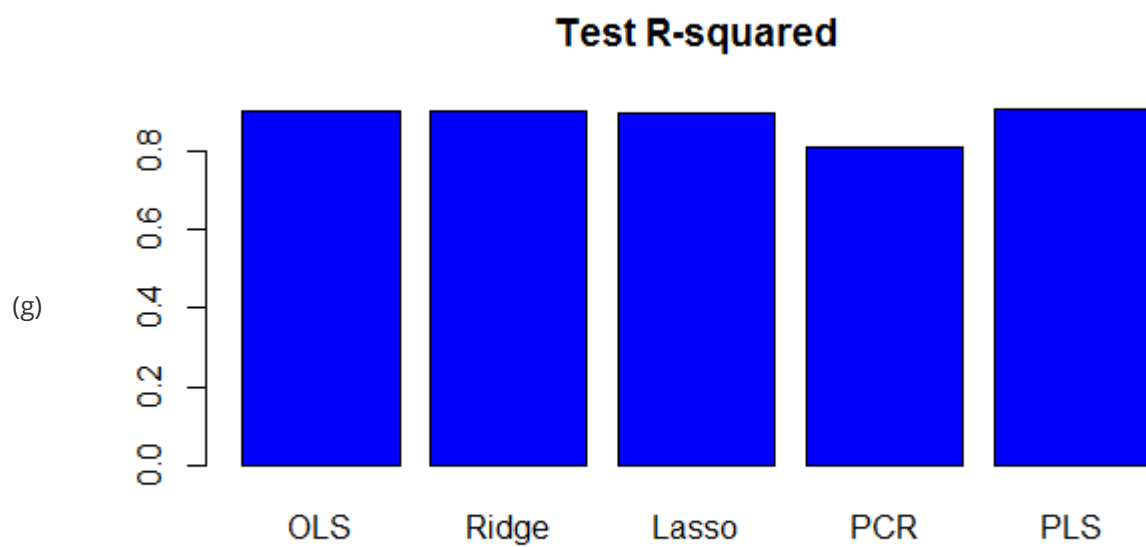
All the models have high R square.