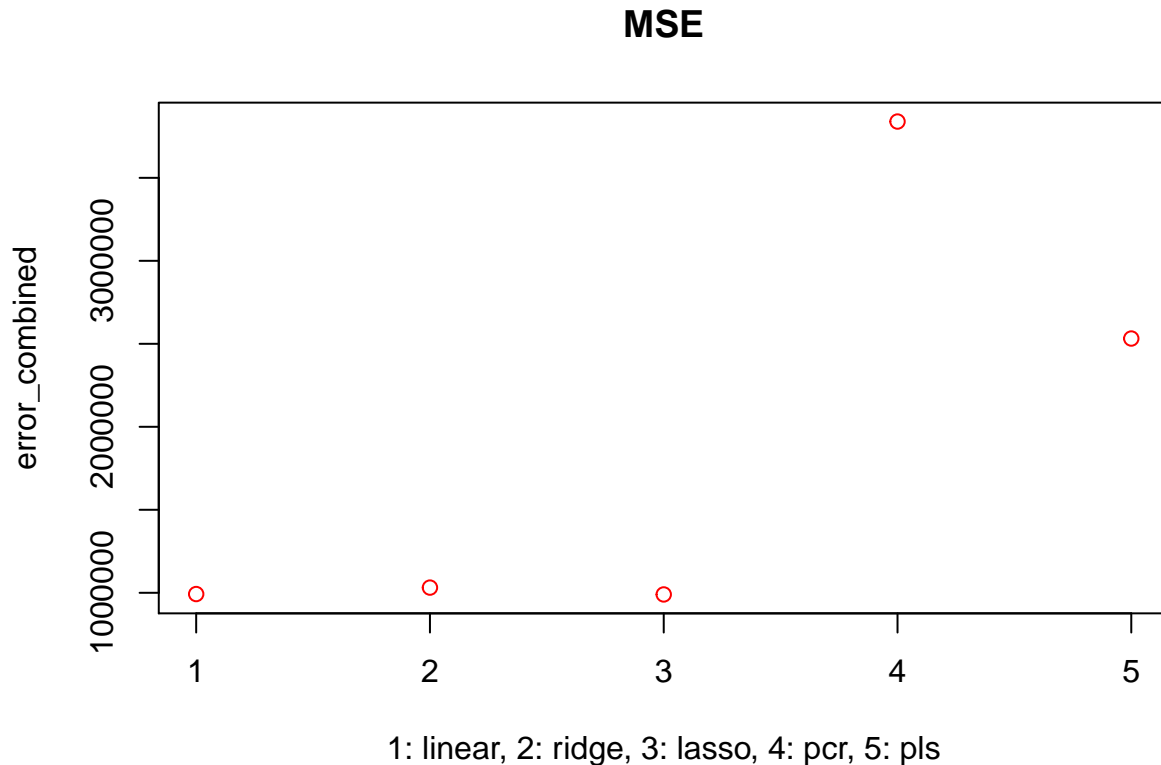


Write-up for Homework2

Hyungkyu Lim

Question 1

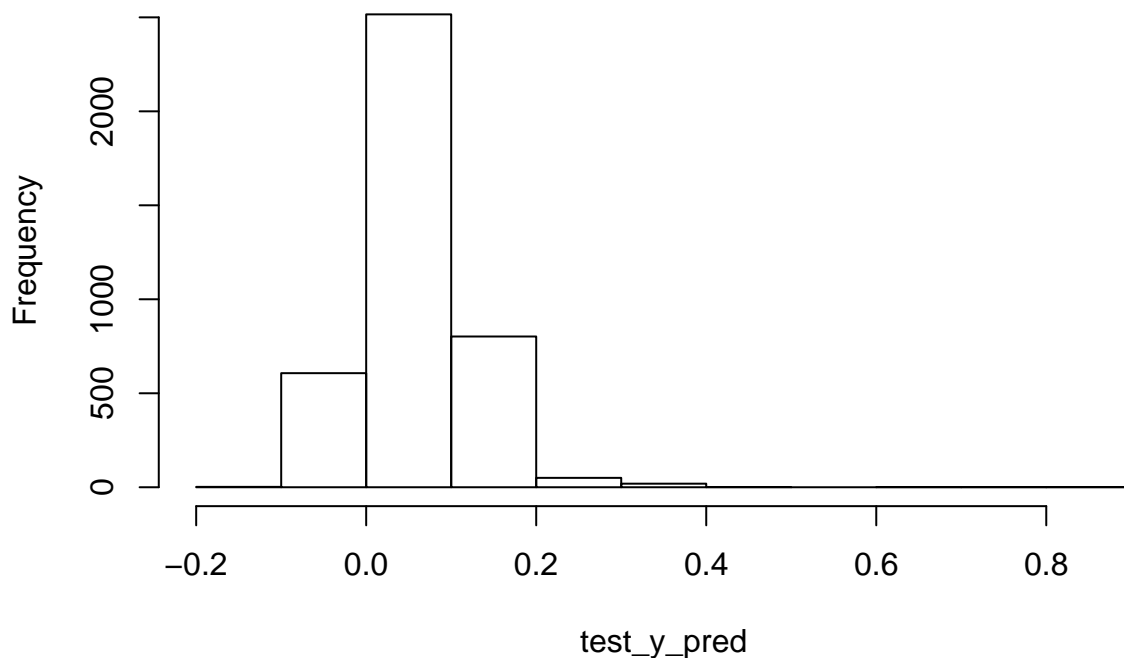
The test errors among 5 methods are not much different. Based on the plot, PCR has the highest test error. It means that PCR is not appropriate for fitting procedure for this College dataset; It is clear then that the direction with most variance of predictors is not strongly related to the predictors. With this, lasso has the smallest test error value. As far as I guess, lasso method works similar with the subset selection. So it will give us a better prediction rather than other methods.



Question 2

When I did OLS with response data; categorical data which has 0 and 1. The range of predicted value of \hat{y} is about $[-1, 1]$. Based on the response dataset, there are only 0 and 1 categorical values so actually the \hat{y} values such as -0.001 and 0.04 do not make sense to estimate the RSS between \hat{y} values and response values. To solve this problem, we have to round the \hat{y} values off for getting 0 or 1.

Histogram of test_y_pred



When I tried to look the \hat{y} value which is greater than 0.5 for rounding off, there were only 3 persons. Based on this result, I can say that only 3 persons will buy the caravan policy. And the max value is However, basically \hat{y} values have negative values, it does not make sense at all. As a result, we cannot do OLS method for the response value which has categorical value.

```
## [1] 0.8082358
```

```
## 576 2863 3139
```

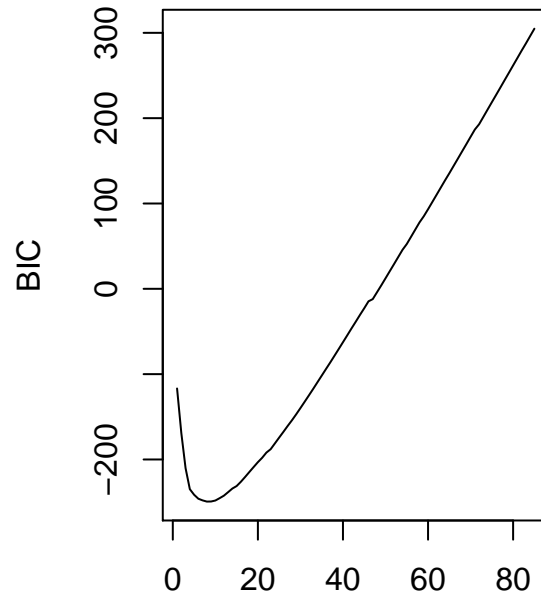
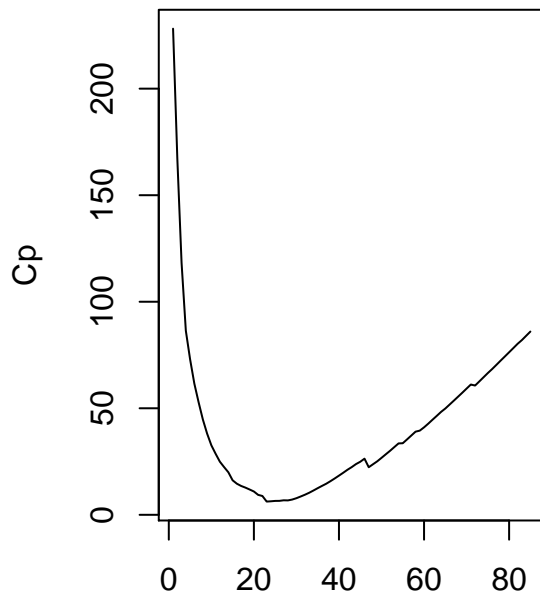
```
## 576 2863 3139
```

Here are the train and test error of OLS.

```
## [1] 0.05210329
```

```
## [1] 0.053985
```

When I did Forward method, I generated the Cp and BIC plot. Among these two, I chose BIC. The lowest BIC is the eight variable models; V10: Married, V18: Lower level education, V43: Purchasing power class, V44: Contribution private third party insurance, V47: Contribution car policies, V59: Contribution fire policies, V82: Number of boat policies, V85: Number of social security insurance policies.



Number of Variables

Number of Variables

```
## [1] 23
```

```
## [1] 8
```

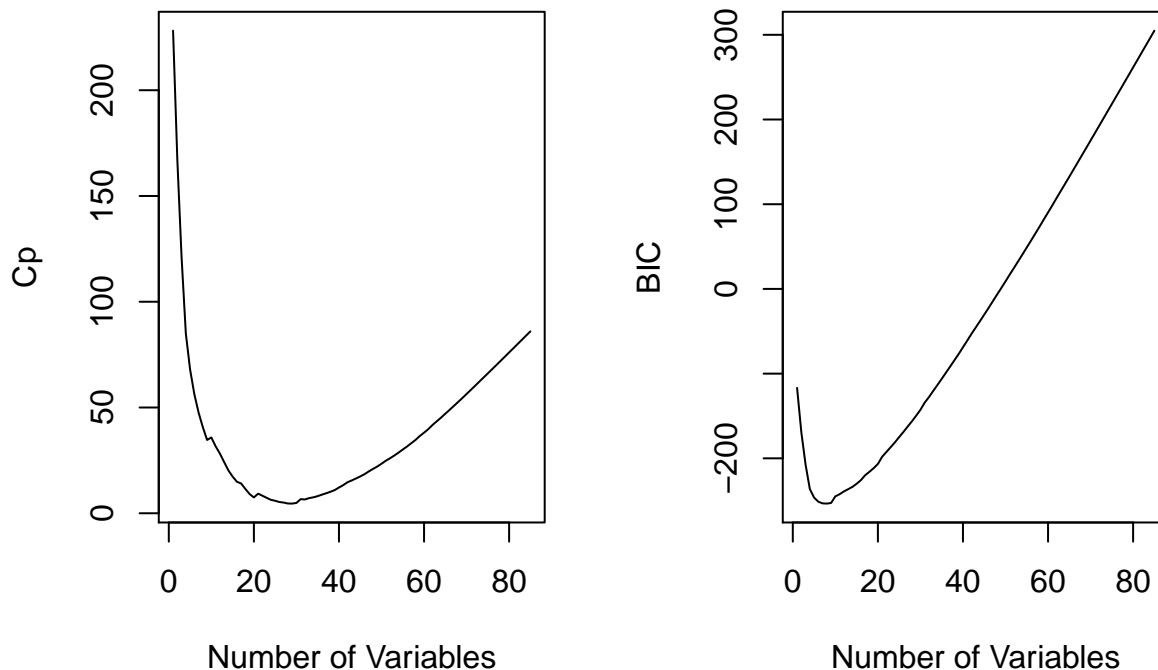
There are coefficients for best eight variabls, train and test errors for Forward selection method.

```
## (Intercept)      V10      V18      V43      V44
## -0.022141589  0.005775927 -0.006291720  0.004805098  0.012538290
##      V47      V59      V82      V85
##  0.010545078  0.005886931  0.284878668  0.080443574
```

```
## [1] 0.05977327
```

```
## [1] 0.05421567
```

When I did Backward method, I generated the Cp and BIC plot same with Forward method. Among these two, I chose BIC. The lowest BIC is the eight variable models; V10: Married, V18: Lower level education, V10: Married, V18: Lower level education ,V21: Farmer, V46: Contribution third party insuran, V47: Contribution car policies, V59: Contribution fire policies, V82: Number of boat policies, V85: Number of social security insurance policies.



```
## [1] 29
```

```
## [1] 8
```

There are coefficients for best eight variabls, train and test errors for Backward selection method.

```
## (Intercept)      V10      V18      V21      V46
## 0.001850234 0.006879012 -0.007523787 -0.008752079 -0.019827878
##      V47      V59      V82      V85
## 0.011057523 0.010985109 0.283583028 0.080852868
```

```
## [1] 0.05309043
```

```
## [1] 0.05411259
```

When I did ridge regression, the best lambda is 419.8756. And it means that when we have the lambda 419.8756, it is good tuning lambda for Ridge regression and also gives us the minimum test error. Here are train and test error for Ridge regression.

```
## [1] 0.1018902
```

```
## [1] 0.05263956
```

```
## [1] 0.05369624
```

When I did Lasso regression, the best lambda is 2.962139. And it means that when we have the lambda 2.962139, it is good tuning lambda for Lasso regression and gives us the minimum test error. Compare to Ridge regression, the Lasso regression has a substantial advantage over Ridge regression in that the eresulting coefficient estimate are sparse. Here are train and test error for Lasso regression.

```
## [1] 0.003495312
```

```
## [1] 0.05278685
```

```
## [1] 0.0537716
```

Question 3

For which model size does the test set MSE take on its minimum value? Comment on your results. How does the model at which the test set MSE is minimized compare to the true model used to generate the data? Comment on the coefficient values.

The model which has 13 variables for the MSE of test set is minimum. When I look at the true model used to generate data, some variables which have 0 are removed from the model which has minimum MSE. As I guess, variables which has 0 doesn't affect the prediction for the best subset selection method. With this, the graph is increasing after it passes the minimum point due to overfitting.

```
## [1] 13
```

```
## (Intercept)      x.1      x.3      x.6      x.8      x.9
## -0.01500769  0.55373531  0.27930116 -0.31218871 -0.58940853  0.17138201
##      x.11      x.12      x.13      x.14      x.17      x.18
##  0.23860860  0.68147272  1.34383548  1.25415414 -1.08183814 -0.48003340
##      x.19      x.20
## -0.33935993 -1.41067349
```

