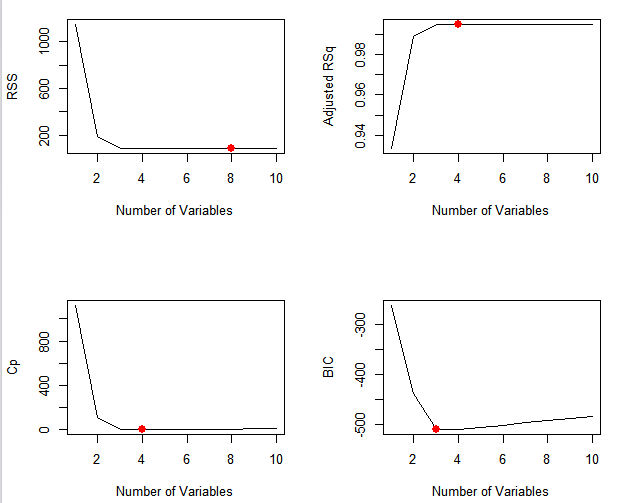
8. In this exercise, we will generate simulated data, and will then use this data to perform best subset selection.

1. Use the rnorm() function to generate a predictor X of length n = 100, as well as a noise vector of length n = 100.
2. Generate a response vector Y of length n = 100 according to the model Y = β0 + β1X + β2X2 + β3X3 +ε, where β0, β1, β2, and β3 are constants of your choice.

I set the β0, β1, β2, and β3 as 1,2,3,4 respectively.

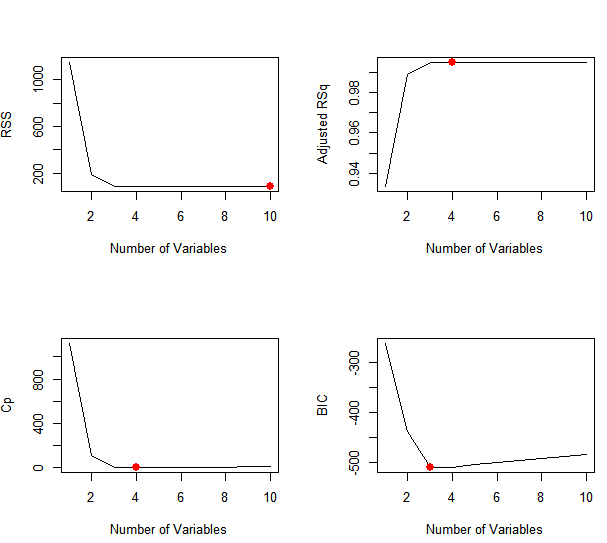
1. ****Use the regsubsets() function to perform best subset selection in order to choose the best model containing the predictors X, X2 , . . . , X10. What is the best model obtained according to Cp, BIC, and adjusted R2 ? Show some plots to provide evidence for your answer, and report the coefficients of the best model obtained. Note you will need to use the data.frame() function to create a single data set containing both X and Y .

The best model according to Cp and adjusted R squared, the model’s coefficient is

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (Intercept) | X1 | X2 | X3 | X5 |
| 1.072008 | 2.387456 | 2.845756 | 3.557974 | 0.080723 |

And if select the best model with BIC the model’s coefficient is

|  |  |  |  |
| --- | --- | --- | --- |
| (Intercept) | X1 | X2 | X3 |
| 1.061507 | 1.97528 | 2.876209 | 4.017639 |

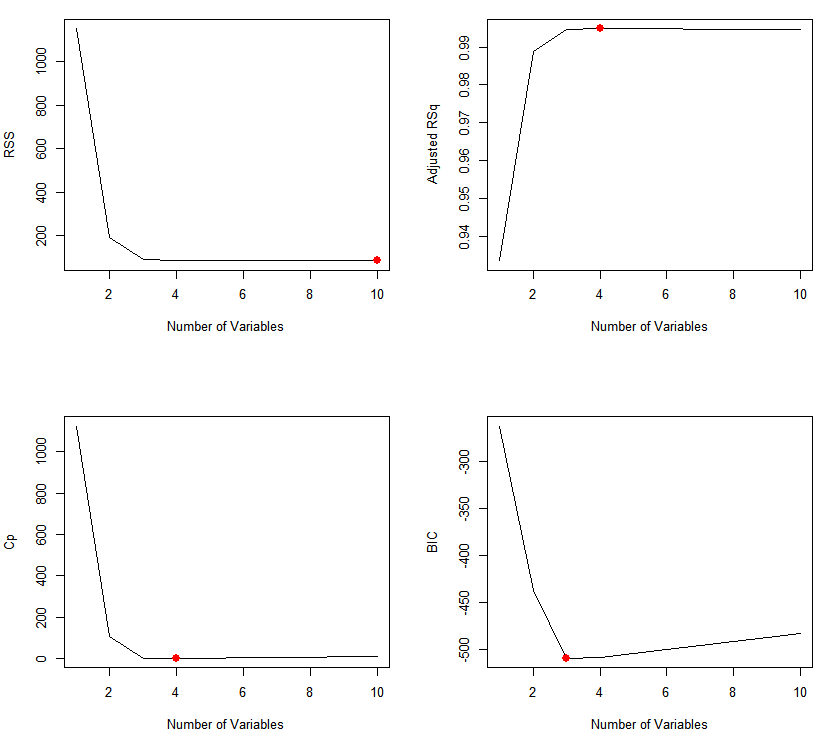
1. Repeat (c), using forward stepwise selection and also using backwards stepwise selection. How does your answer compare to the results in (c)?
   1. ****Forward selection

The best model according to Cp and adjusted R squared, the model’s coefficient is

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (Intercept) | V1 | V2 | V3 | V5 |
| 1.072008 | 2.387456 | 2.845756 | 3.557974 | 0.080723 |

And if select the best model with BIC the model’s coefficient is

|  |  |  |  |
| --- | --- | --- | --- |
| (Intercept) | X1 | X2 | X3 |
| 1.061507 | 1.97528 | 2.876209 | 4.017639 |

* 1. ****Backward selection  
     The best model according to Cp and adjusted R squared, the model’s coefficient is

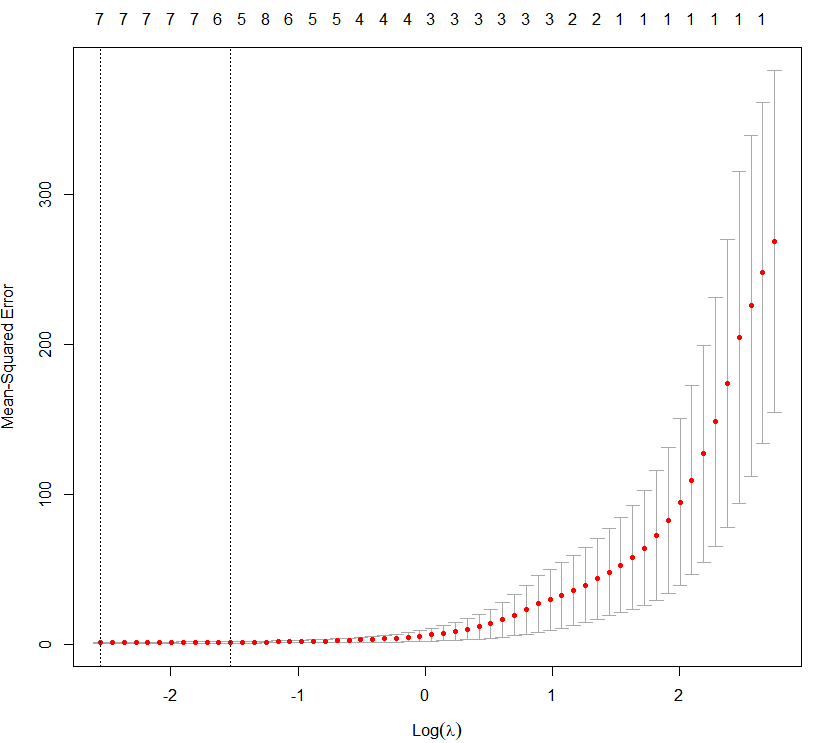
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (Intercept) | X1 | X2 | X3 | X9 |
| 1.079236 | 2.231906 | 2.833494 | 3.819556 | 0.001291 |

And if select the best model with BIC the model’s coefficient is

|  |  |  |  |
| --- | --- | --- | --- |
| (Intercept) | X1 | X2 | X3 |
| 1.061507 | 1.97528 | 2.876209 | 4.017639 |

If we choose the model according to BIC three selection method give the same result, but using the Cp and adjusted R squared as criteria the subset selection and forward selection having same result (X1 X2 X3 X5), and backward selection suggest the best model is using (X1 X2 X3 X9) as predictor.

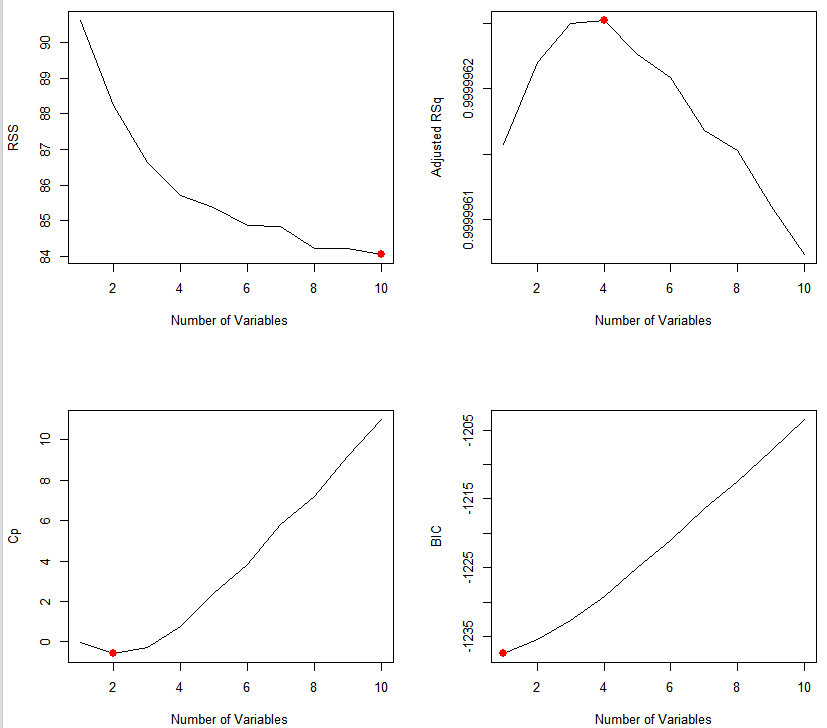
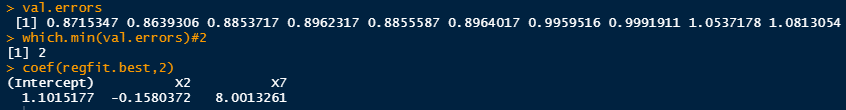
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | V1 | V2 | V3 | V4 | V5 | V7 |
| 1.18508 | 2.141599 | 2.618907 | 3.794647 | 0.043349 | 0.021557 | 0.002729 |

1. Now fit a lasso model to the simulated data, again using X, X2 , . . . , X10 as predictors. Use cross-validation to select the optimal value of λ. Create plots of the cross-validation error as a function of λ. Report the resulting coefficient estimates, and discuss the results obtained.  
   The best value of λ is 0.07787299. the test mean square error is 0.8683319.  
   

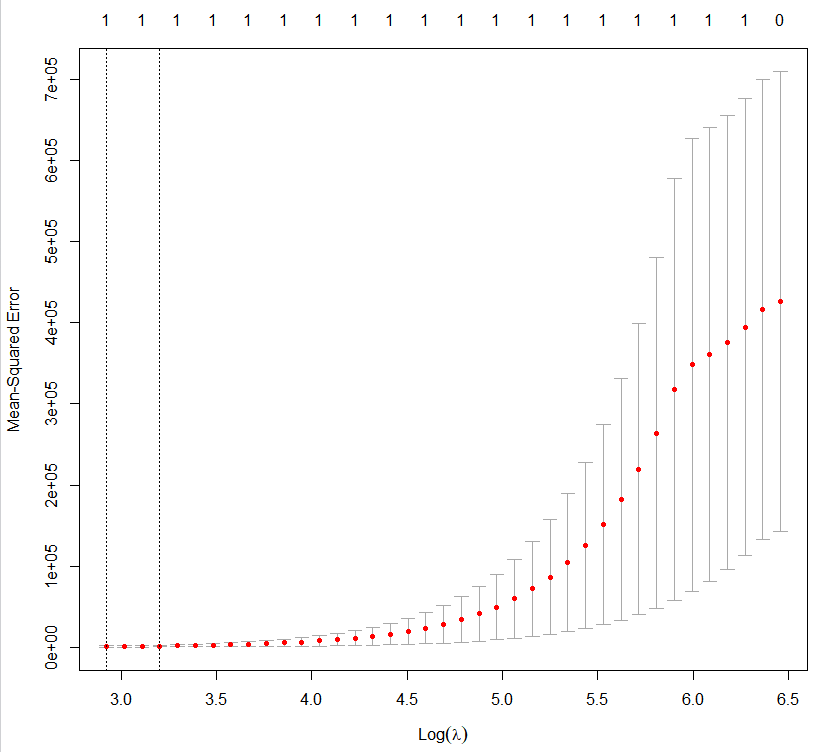
The lasso’s best model have six predictor as result, and this is much more than the true function(with only beta0-3).  
一張含有 文字 的圖片

自動產生的描述  
the testing mean square error in best subset method is 0.8654413, and lasso method’s testing mean square error is 0.8683319 both are very close.

1. Now generate a response vector Y according to the model Y = β0 + β7X7 + ε, and perform best subset selection and the lasso. Discuss the results obtained.
   1. Best subset

The Adj.R^2, Cp, BIC suggest different model. Using cross validation, it suggest to use the model with two predictors. The mean testing error is 0.8639306

* 1. Lasso

The best value of λ is 18.59963. the test mean square error is 103.892.  
the coefficient of lasso’s best model is

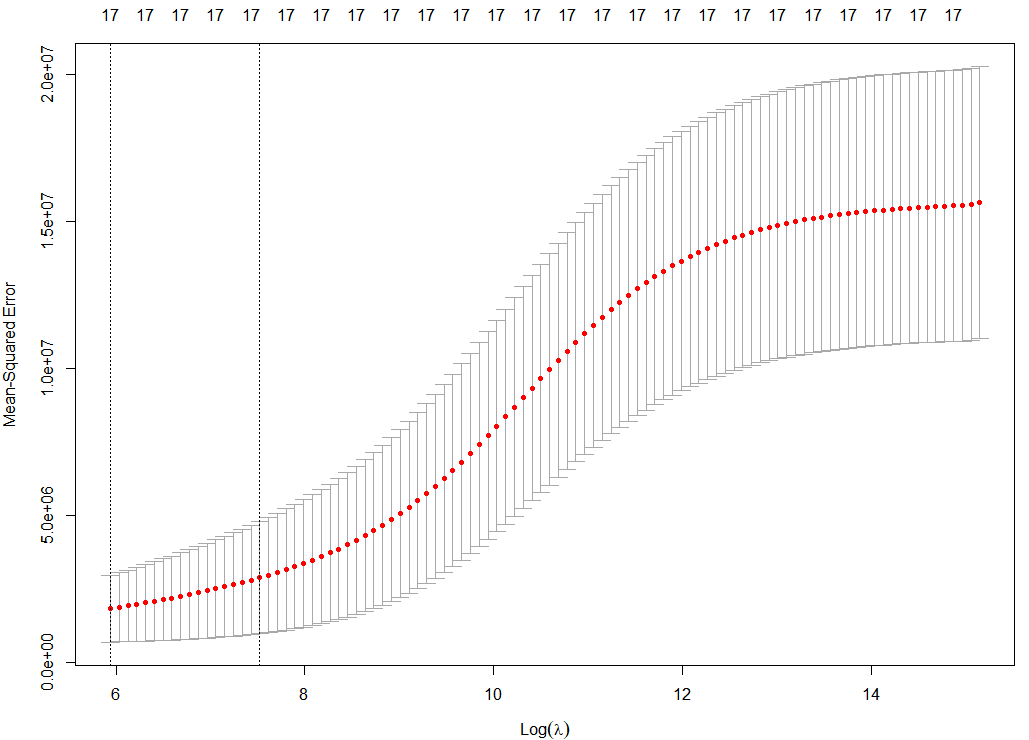
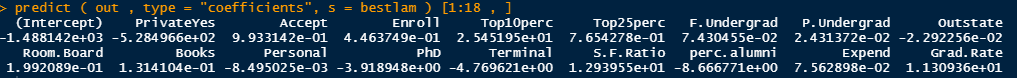
|  |  |
| --- | --- |
| (Intercept) | V7 |
| 2.254081 | 7.693891 |

In conclusion though the best subset have better estimation, the lasso have the right number of predictor.

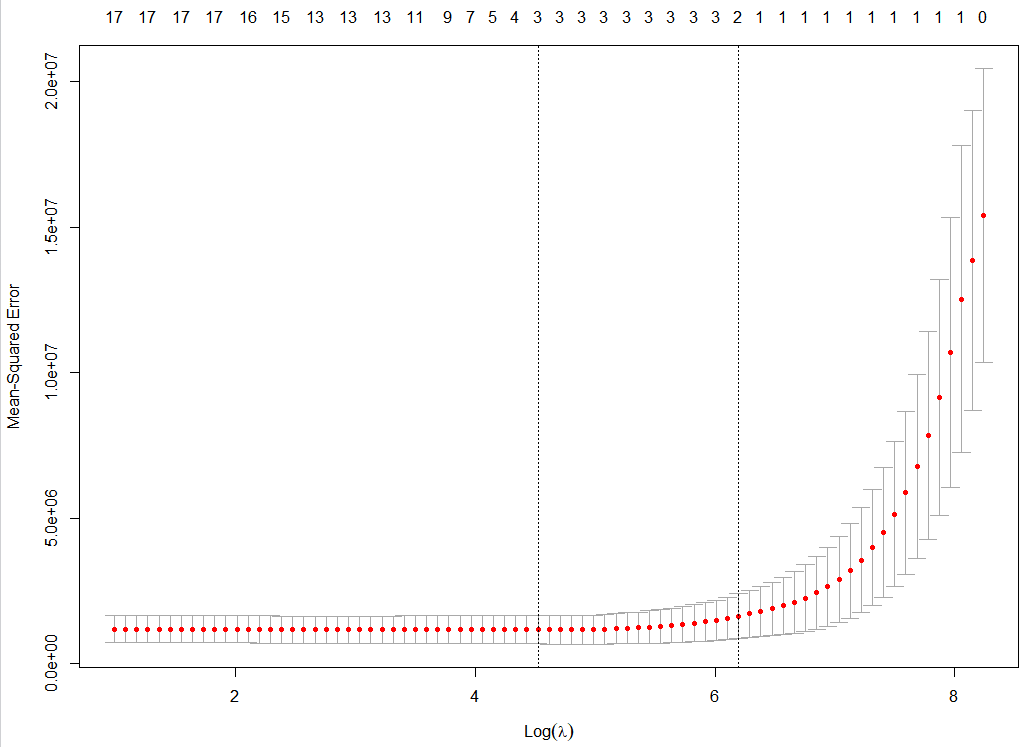
9. In this exercise, we will predict the number of applications received using the other variables in the College data set.

1. Split the data set into a training set and a test set.
2. Fit a linear model using least squares on the training set, and report the test error obtained.   
   **一張含有 文字 的圖片

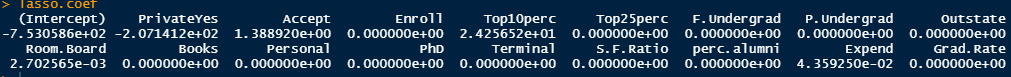
   自動產生的描述**  
   The result of multiple linear regression is shown above, and the test error(MSE)is 1,555,348.
3. Fit a ridge regression model on the training set, with λ chosen by cross-validation. Report the test error obtained.

  
The best λ is 380.0277 and the test error is 1,457,825  
****

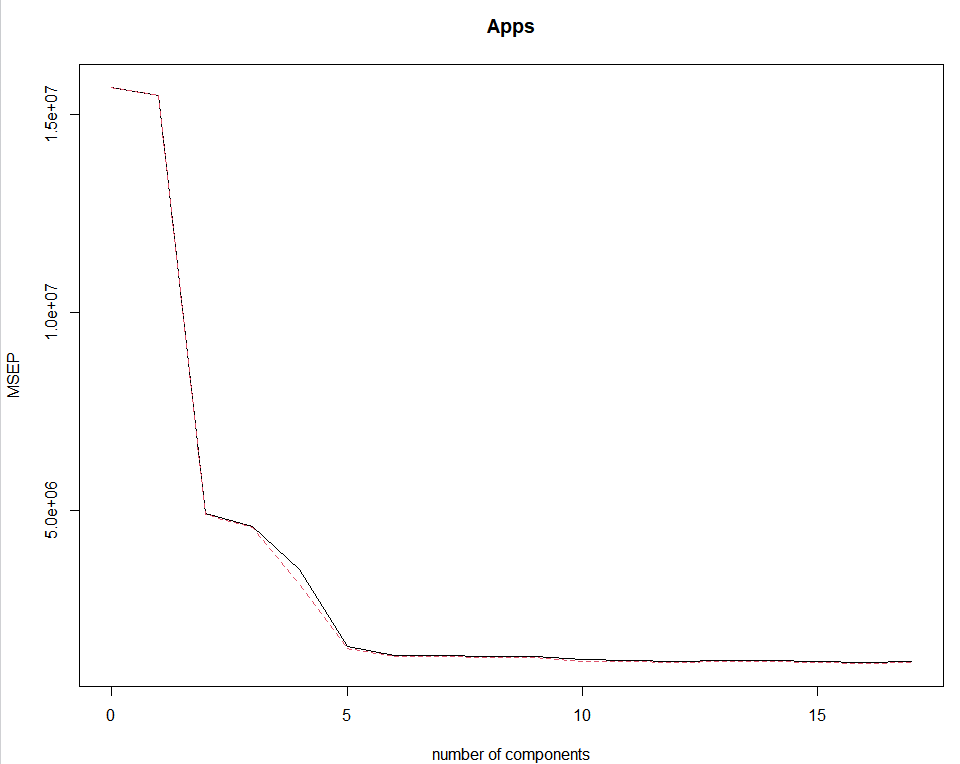
1. Fit a lasso model on the training set, with λ chosen by cross validation. Report the test error obtained, along with the number of non-zero coefficient estimates.

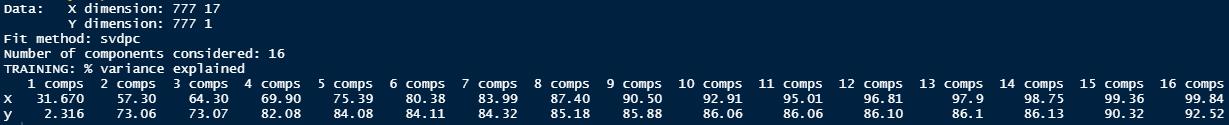
****

The best λ is 91.97157 and the test error is 1,160,570

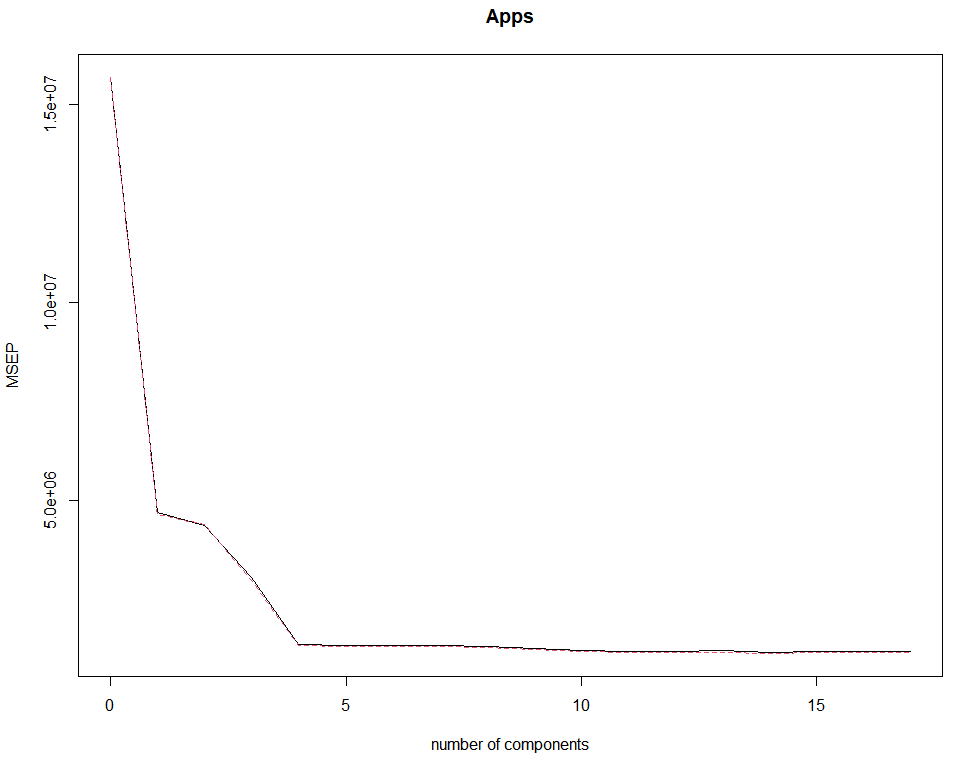
****

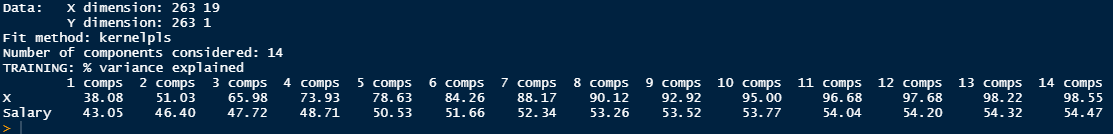
1. Fit a PCR model on the training set, with M chosen by cross validation. Report the test error obtained, along with the value of M selected by cross-validation.

****The lowest error occurs in M=16, and the test error is 1,588,261

The result by fitting all the data

1. Fit a PLS model on the training set, with M chosen by cross validation. Report the test error obtained, along with the value of M selected by cross-validation.

****

****The lowest error occurs in M=14, and the test error is 1,563,763  
The result by fitting all the data

1. Comment on the results obtained. How accurately can we predict the number of college applications received? Is there much difference among the test errors resulting from these five approaches?

Five approach’s test error is

|  |  |
| --- | --- |
| method | Test error |
| Least squares | 1,555,348. |
| Ridge regression | 1,457,825 |
| Lasso model | 1,160,570 |
| PCR | 1,588,261 |
| PLS | 1,563,763 |

Lasso provide the lowest test error so it might be a best model. And lasso reduce the number of predictors to 6. This can help us intercept the model.

The Least squares model doing slightly better than PCR and PLS, I believe if we apply the best subset on the Least squares method this might provide better performance than PCR and PLS.