Assignment 8
Shaheed Shihan
Modern Applied Statistics II

4.

a. As we increase λ from 0, the the training RSS will also increase gradually. It will eventually start increasing in a U shape. Because we are restricting the β_j coefficients, the model becomes less flexible which then tends to increase the training RSS.

b.

The training RSS will decrease initially and then start increasing in a U shape. Once again, the β_j coefficients are restricted, and the model becomes less flexible which then tends to increase the training RSS.

c.

The training RSS will steadily decrease. Same argument applies because as we increase λ from 0, we are restricting the β_j coefficients and the model becomes less flexible.

d.

Steadily increase. Same argument applies because as we increase λ from 0, we are restricting the β_j coefficients and the model becomes less flexible.

e.

Remain constant.

The irreducible error by definition is independent of the model and therefore independent of λ.

9.

(a) Split the data set into a training set and a test set.

(b) Fit a linear model using least squares on the training set, and report the test error obtained.

```
> mean((pred.lm-C.test$Apps)^2)
[1] 1423848
```

(c) Fit a ridge regression model on the training set, with λ chosen by cross-validation. Report the test error obtained.

```
> mean((ridge.pred - C.test$Apps)^2)
[1] 1535810
```

(d) 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.

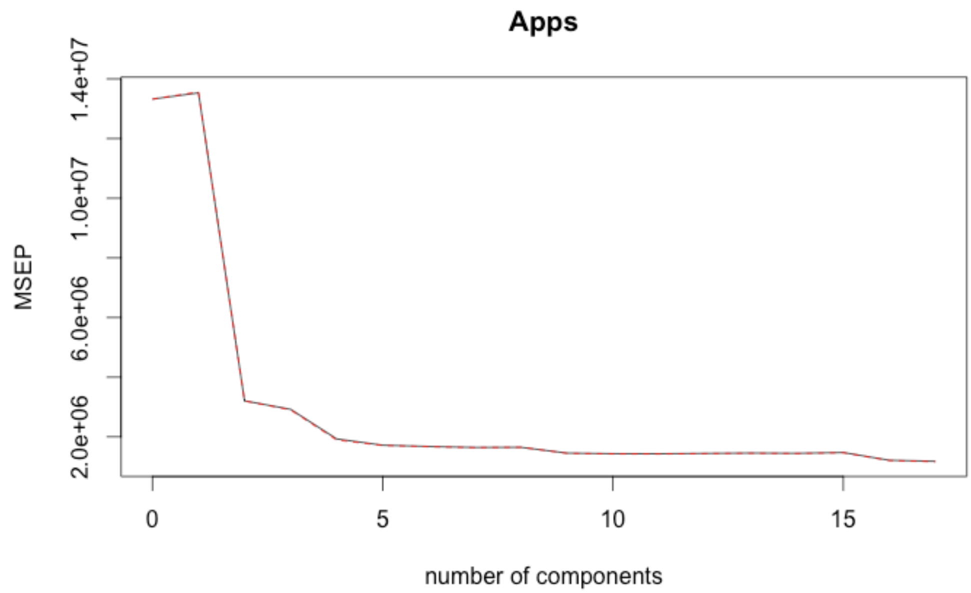
```
> mean((lasso.pred - C.test$Apps)^2)
[1] 1528837

> predict(lasso.C, s = b.lasso, type = "coefficients")
19 x 1 sparse Matrix of class "dgCMatrix"
                        1
(Intercept) -6.527397e+02
(Intercept)  .
PrivateYes  -4.961485e+02
Accept       1.251596e+00
Enroll       .
Top10perc    4.930955e+01
Top25perc   -1.264461e+01
F.Undergrad  3.938493e-02
P.Undergrad  1.610523e-03
Outstate    -4.034062e-02
Room.Board   1.789183e-01
Books        6.604007e-02
Personal     .
PhD         -7.242539e+00
Terminal     .
S.F.Ratio    2.345341e+00
perc.alumni -9.518836e+00
Expend       4.877833e-02
Grad.Rate    7.455284e+00
```

(e) Fit a PCR model on the training set, with M chosen by cross validation. Report the test error obtained, along with the value.

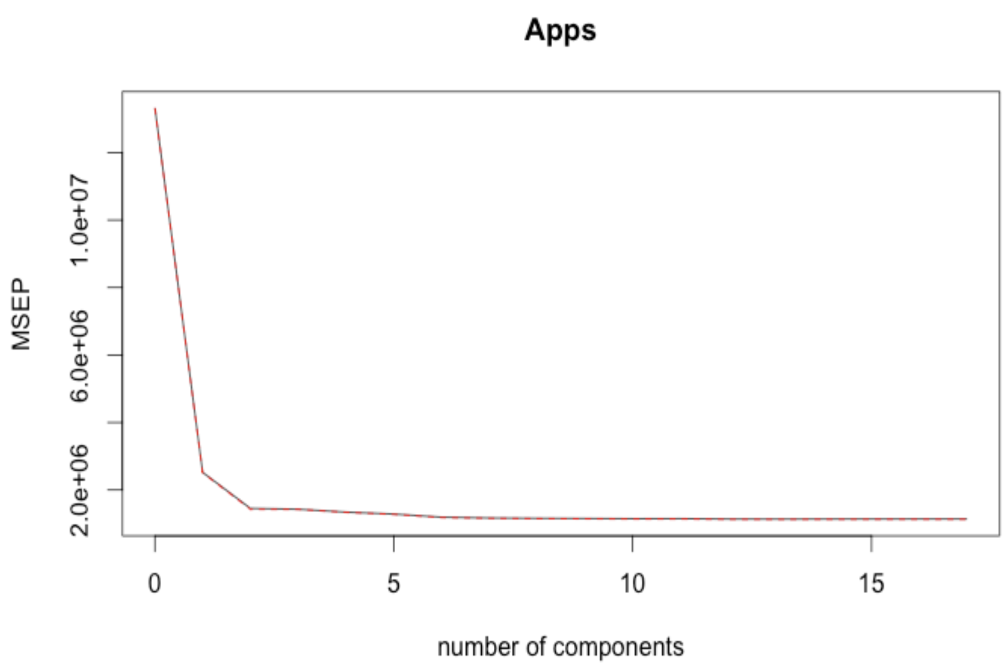
```
> mean((pcr.pred - C.test$Apps)^2)
[1] 3644125
```

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



(f) 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.

```
> mean((pls.pred - C.test$Apps)^2)
[1] 1455808
```

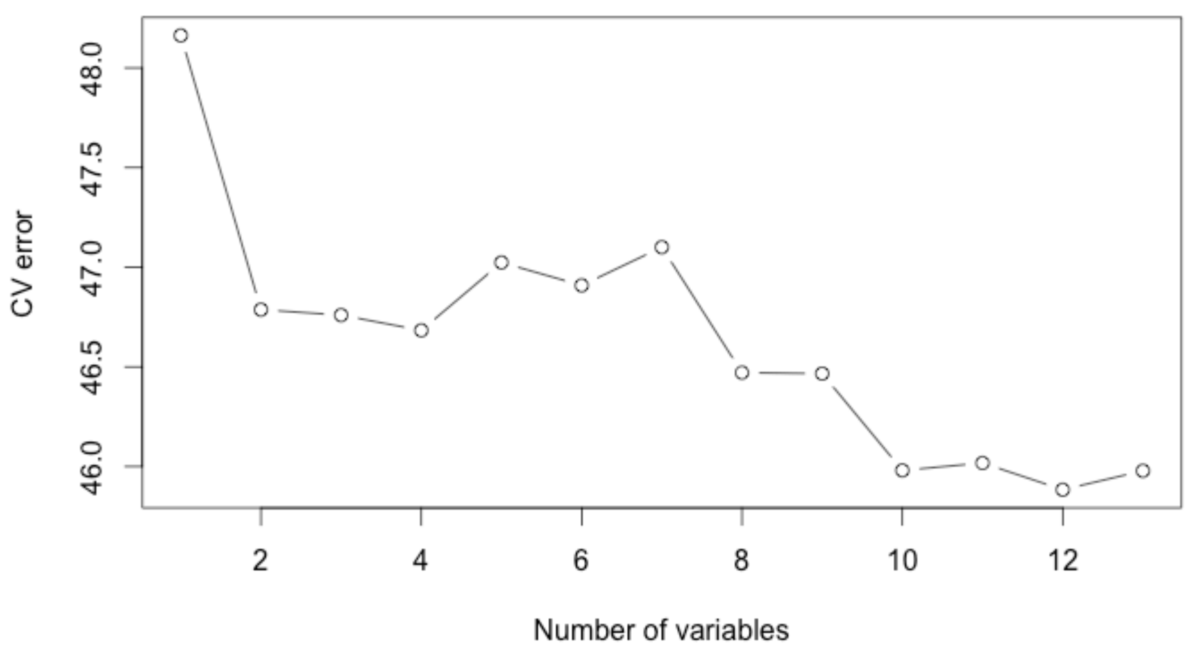


(g) Comment on the results obtained. How accurately can we predict the number of college applications received? Is there much difference among the test errors resulting from these five approaches?

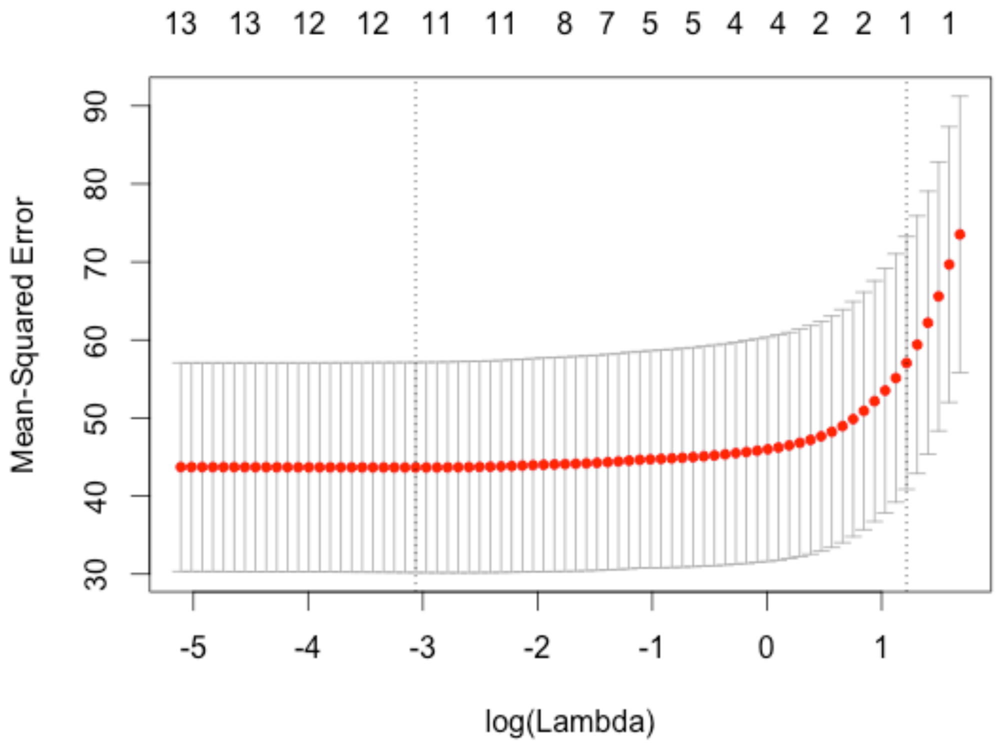
```
> data.frame(r.lm,r.ridge, r.lasso,r.pcr,r.pls)
    r.lm       r.ridge     r.lasso     r.pcr       r.pls
1 0.92234   0.9162333   0.9166137   0.8012409   0.9205968
```

Using the R-squared value we can say that besides the pcr model, all the others are very good predictors of the number of college applications that are received.

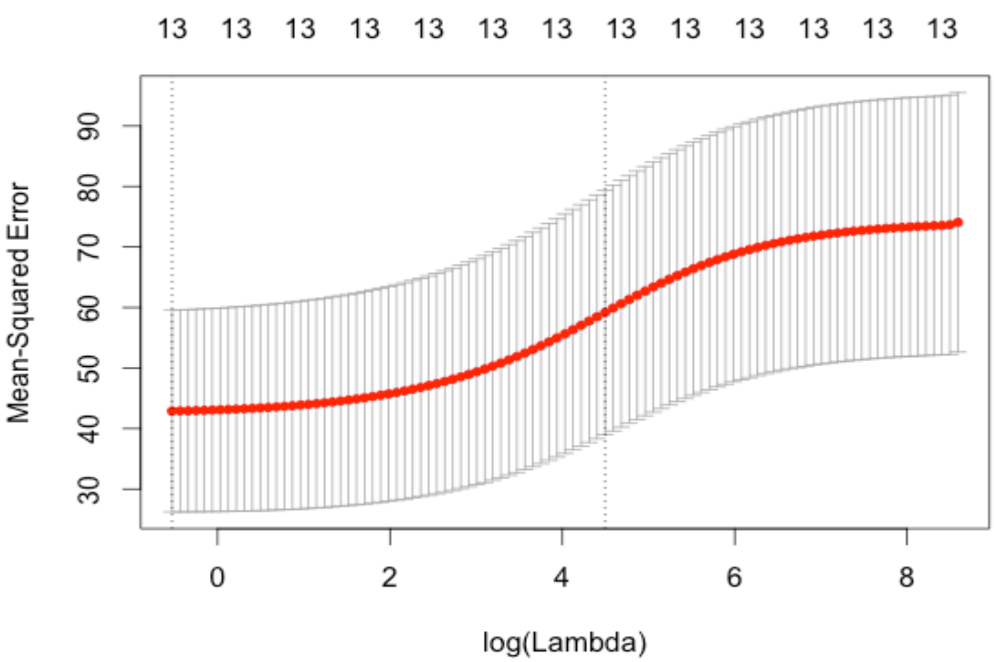
11.



Lasso:


13 13 12 12 11 11 8 7 5 5 4 4 2 2 1 1
Mean-Squared Error
30 40 50 60 70 80 90
-5 -4 -3 -2 -1 0 1
log(Lambda)

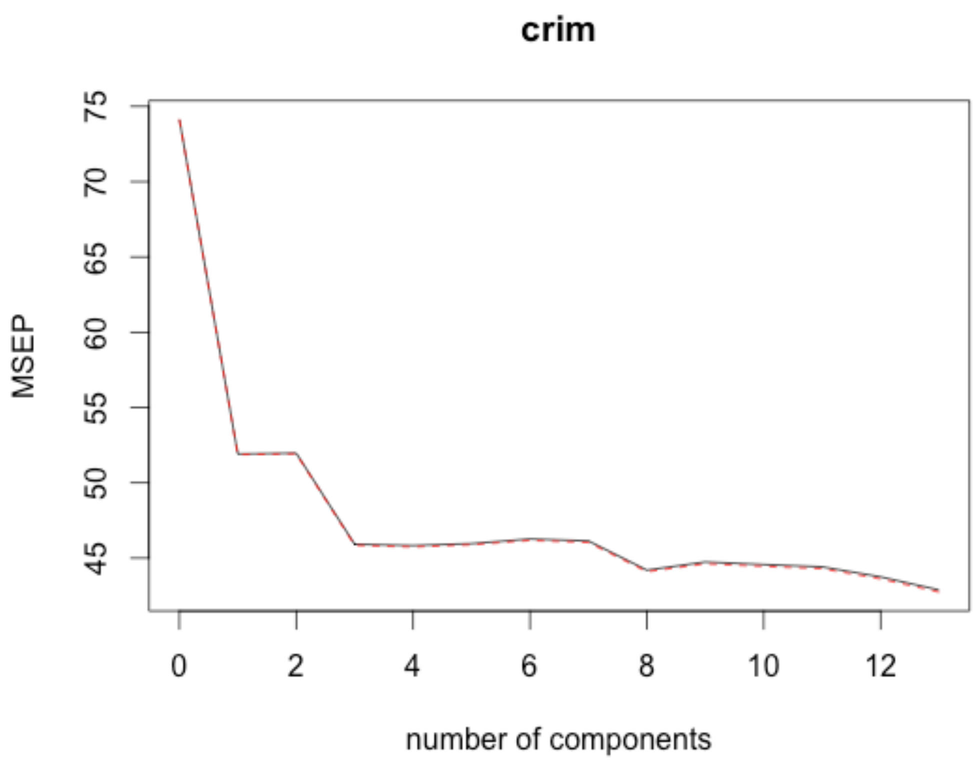
Ridge:


13 13 13 13 13 13 13 13 13 13 13 13 13
Mean-Squared Error
30 40 50 60 70 80 90
0 2 4 6 8
log(Lambda)

```
> summary(fitpcr)
Data:   X dimension: 506 13
        Y dimension: 506 1
Fit method: svdpc
Number of components considered: 13

VALIDATION: RMSEP
Cross-validated using 10 random segments.
     (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps  9
comps  10 comps  11 comps
CV          8.61    7.205    7.209    6.776    6.770    6.779    6.802    6.793    6.648    6.689     6.676
6.665
adjCV       8.61    7.202    7.206    6.771    6.763    6.774    6.795    6.785    6.641    6.680     6.667
6.655
     12 comps  13 comps
CV       6.615    6.548
adjCV    6.604    6.538

TRAINING: % variance explained
    1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps  9 comps  10
comps  11 comps  12 comps
X      47.70    60.36    69.67    76.45    82.99    88.00    91.14    93.45    95.40     97.04     98.46
99.52
crim   30.69    30.87    39.27    39.61    39.61    39.86    40.14    42.47    42.55     42.78     43.04
44.13
    13 comps
X       100.0
crim     45.4
```



The subset selection method had the lowest MSE even though this did not select all the variables. It only took 13 of the predictors.