Homework 8

Prakash Paudyal

Please do the following problems from the text book ISLR.

1. Question 6.8.4 pg 260

4.Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right) - \lambda\sum_{j=1}^{p}\beta_j^2$$

for a particular value of . For parts (a) through (e), indicate which of i. through v. is correct. Justify your answer.

(a)

As we increase λ from 0, the training RSS will: i. Increase initially, and then eventually start decreasing in an inverted U shape. ii. Decrease initially, and then eventually start increasing in a U shape. iii. Steadily increase. iv. Steadily decrease. v. Remain constant.

ANS (iii)Steadily increase. As tuning parameter λ increse from 0, the impact of the shrinkage penalty grows, and the ridge regression coefficient estimates will approach zero which eventually steadily increase training RSS

(b)

Repeat (a) for test RSS.

ANS

(ii.) Decrease initially, and then eventually start increasing in a U shape. When λ=0, all β's have their least square estimate values. In this case, the model tries to fit hard to training data and hence test RSS is high. As we increaseλ, beta's start reducing to zero and some of the overfitting is reduced. Thus, test RSS initially decreases. Eventually by inccreasing λ, as beta's approach 0, the model becomes too simple and test RSS increases, making a U shape.

(c)

Repeat (a) for variance.

ANS

(iv.) Steadily decrease. As λ increases, the flexibility of the ridge regression fit decreases,leading to decreased variance.As we increaseλ, βs start decreasing and model becomes simpler. In the limiting case ofλ approaching infinity, all betas reduce to zero and model predicts a constant and has no variance.

(d)

Repeat (a) for (squared) bias.

ANS

iii) Steadily increases: As λ increases, the flexibility of the ridge regression fit decreases,leading to increased bias. When λ=0,βs have their least-square estimate values and hence have the least bias. Asλ increases,βs start reducing towards zero, the model fits less accurately to training data and hence bias increases. In the limiting case ofλ approaching infinity, the model predicts a constant and hence bias is maximum

(e)

Repeat (a) for the irreducible error.

ANS

v) Remains constant: By definition, irreducible error is model independent and hence irrespective of the choice of λ, remains constant.

2. Question 6.8.9 pg 263

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

```
## 'data.frame':    777 obs. of  18 variables:
##  $ Private    : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Apps       : num  1660 2186 1428 417 193 ...
##  $ Accept     : num  1232 1924 1097 349 146 ...
##  $ Enroll     : num  721 512 336 137 55 158 103 489 227 172 ...
##  $ Top10perc  : num  23 16 22 60 16 38 17 37 30 21 ...
##  $ Top25perc  : num  52 29 50 89 44 62 45 68 63 44 ...
##  $ F.Undergrad: num  2885 2683 1036 510 249 ...
##  $ P.Undergrad: num  537 1227 99 63 869 ...
##  $ Outstate   : num  7440 12280 11250 12960 7560 ...
##  $ Room.Board : num  3300 6450 3750 5450 4120 ...
##  $ Books      : num  450 750 400 450 800 500 500 450 300 660 ...
##  $ Personal   : num  2200 1500 1165 875 1500 ...
##  $ PhD        : num  70 29 53 92 76 67 90 89 79 40 ...
##  $ Terminal   : num  78 30 66 97 72 73 93 100 84 41 ...
##  $ S.F.Ratio  : num  18.1 12.2 12.9 7.7 11.9 9.4 11.5 13.7 11.3 11.5 ...
##  $ perc.alumni: num  12 16 30 37 2 11 26 37 23 15 ...
##  $ Expend     : num  7041 10527 8735 19016 10922 ...
##  $ Grad.Rate  : num  60 56 54 59 15 55 63 73 80 52 ...
```

	Private	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad
Abilene Christian University	Yes	1660	1232	721	23	52	2885	537
Adelphi University	Yes	2186	1924	512	16	29	2683	1227
Adrian College	Yes	1428	1097	336	22	50	1036	99
Agnes Scott College	Yes	417	349	137	60	89	510	63
Alaska Pacific University	Yes	193	146	55	16	44	249	869
Albertson College	Yes	587	479	158	38	62	678	41

(a)

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

```
set.seed(11)
smp_size <- floor(0.50 * nrow(College))
train_ind <- sample(seq_len(nrow(College)), size = smp_size)
train.college<-College[train_ind, ]
test.college <- College[-train_ind, ]
```

I splitied data into 50%

(b)

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

Linear model coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-629.1837259	540.8363590	-1.1633532	0.2454359
PrivateYes	-339.6436264	186.4995422	-1.8211499	0.0693915
Accept	1.3064090	0.0734337	17.7903311	0.0000000
Enroll	-0.7034167	0.2830397	-2.4852226	0.0133881
Top10perc	38.5265483	8.5603834	4.5005634	0.0000091
Top25perc	-10.8204149	6.6861712	-1.6183275	0.1064441
F.Undergrad	0.1470792	0.0501894	2.9304823	0.0035944
P.Undergrad	0.0262238	0.0536316	0.4889625	0.6251579
Outstate	-0.0735270	0.0281935	-2.6079427	0.0094777
Room.Board	0.0346124	0.0680944	0.5082999	0.6115460
Books	-0.2317152	0.2960712	-0.7826333	0.4343431
Personal	0.0164631	0.0916039	0.1797210	0.8574700
PhD	-3.8361732	6.2555431	-0.6132438	0.5400919
Terminal	-6.1273085	6.8059521	-0.9002867	0.3685533
S.F.Ratio	15.9083233	17.3925701	0.9146620	0.3609648
perc.alumni	-8.8346041	5.8426037	-1.5121005	0.1313619
Expend	0.1633978	0.0232310	7.0336135	0.0000000
Grad.Rate	9.9569947	4.2052075	2.3677773	0.0184089

Test Error for Linear model

```
## [1] 1538442
```

The mean squared error,TEST MSE is $ lm_{MSE}$=1538442, which is extremely large.

(c)

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

```
## Loading required package: Matrix
```

```
## Loading required package: foreach
```

```
## Loaded glmnet 2.0-13
```

Best chosen by cross-validation is

```
## [1] 18.73817
```

Test error for ridge regression model

```
## [1] 1608859
```

The mean squared error,TEST MSE is $ ridge_{MSE}$=1608859 which is slightly higher than test error of linear model.

(d)

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

Best chosen by cross-validation is

```
## [1] 14.17474
```

Test error for lasso model

```
## [1] 1626477
```

Number of non-zero coefficient estimates are 15. LASSO reduced the coefficent of F.Undergrad and Books to zero.

```
## 19 x 1 sparse Matrix of class "dgCMatrix"
##                           1
## (Intercept) -556.01427783
## (Intercept)    .
## PrivateYes  -458.50058120
## Accept         1.49947647
## Enroll        -0.34285887
## Top10perc     39.10796866
## Top25perc     -6.17240648
## F.Undergrad    .
## P.Undergrad    0.03388810
## Outstate      -0.06812801
## Room.Board     0.13486936
## Books          .
## Personal       0.01173185
## PhD           -6.71936715
## Terminal      -3.15826314
## S.F.Ratio      8.61896397
## perc.alumni   -0.72038115
## Expend         0.07249557
## Grad.Rate      6.28499717
```

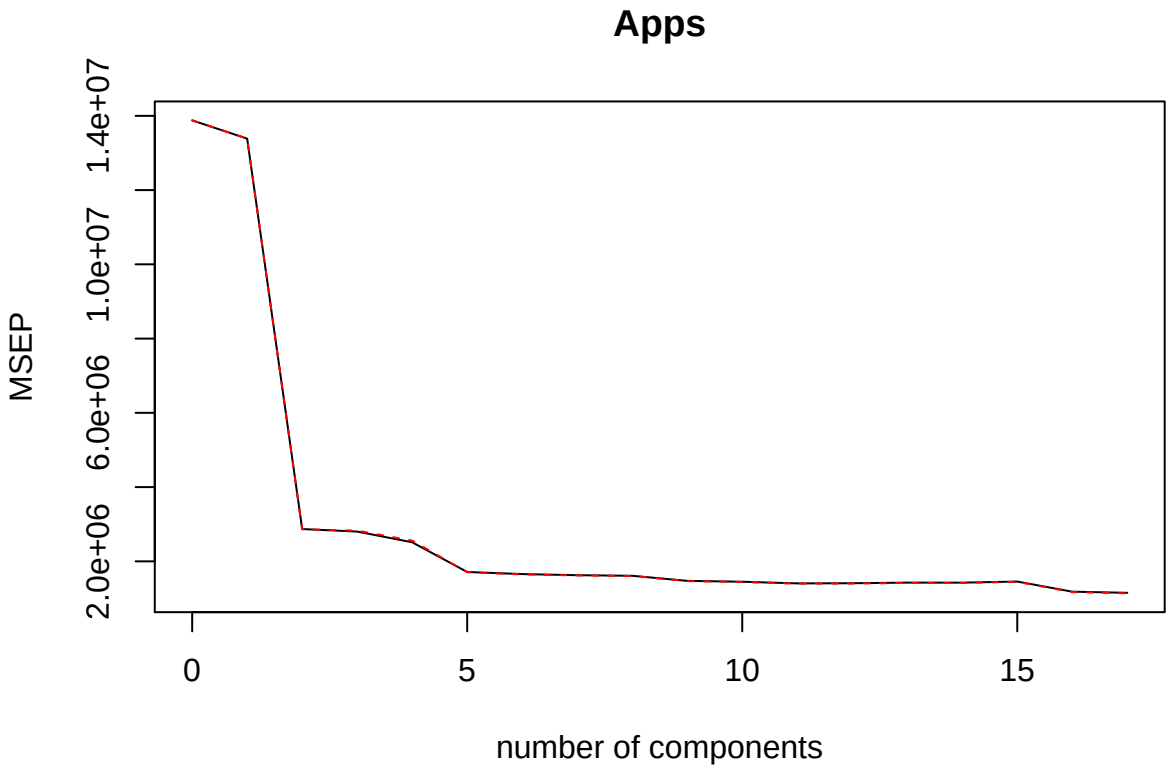
(e)

Fit a PCR 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.

Value of M selected by cross-validation.

```
##
## Attaching package: 'pls'
```

```
## The following object is masked from 'package:stats':
##
##     loadings
```




```
## Data:    X dimension: 388 17
##  Y dimension: 388 1
## Fit method: svdpc
## Number of components considered: 17
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## CV            3725     3659     1694     1674     1585     1309     1287
## adjCV         3725     3659     1693     1681     1599     1306     1284
##        7 comps  8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## CV        1275     1269     1214      1204      1186      1188      1194
## adjCV     1273     1267     1211      1201      1183      1184      1191
##        14 comps  15 comps  16 comps  17 comps
## CV         1194      1207      1087      1074
## adjCV      1191      1204      1082      1068
##
## TRAINING: % variance explained
##       1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
## X      33.709    58.65    65.49    71.54    77.23    82.09    85.71
## Apps    4.235    79.52    80.66    82.15    88.32    88.72    89.07
##       8 comps  9 comps  10 comps  11 comps  12 comps  13 comps  14 comps
## X       88.91    91.71     93.89     95.96     97.42     98.36     99.17
## Apps    89.13    90.13     90.43     90.62     90.77     90.78     90.81
##       15 comps  16 comps  17 comps
## X        99.58     99.88    100.00
```

```
## Apps     90.89     92.72     93.18
```

We see that the smallest cross-validation error occurs when M = 17 components are used Hence ,M selected by cross-validation is 17, which amounts to simply performing least squares, because when all of the components are used in PCR no dimension reduction occurs.

Test error

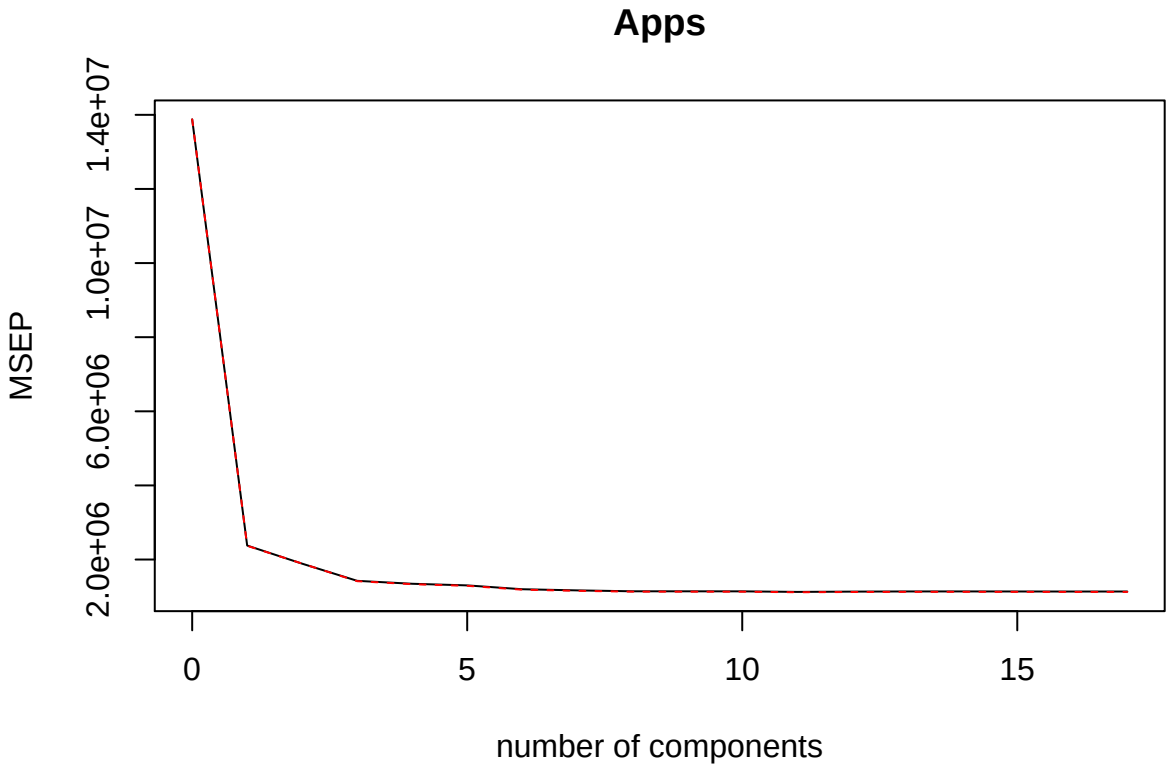
```
## [1] 1538442
```

The mean squared error,TEST MSE is $ pcr_{MSE}$=1538442.

(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.

```
## Data:    X dimension: 388 17
##  Y dimension: 388 1
## Fit method: kernelpls
## Number of components considered: 17
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## CV            3725     1542     1374     1194     1160     1141     1095
## adjCV         3725     1540     1375     1190     1155     1136     1089
##        7 comps  8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## CV        1080     1069     1069      1069      1061      1065      1067
## adjCV     1075     1064     1064      1064      1057      1060      1062
##        14 comps  15 comps  16 comps  17 comps
## CV         1068      1067      1066      1066
## adjCV      1062      1062      1061      1061
##
## TRAINING: % variance explained
##       1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
## X       25.63    51.06    64.29    67.91    72.61    74.91    78.28
## Apps    83.28    87.16    90.67    91.47    91.94    92.72    92.97
##       8 comps  9 comps  10 comps  11 comps  12 comps  13 comps  14 comps
## X       81.92    85.51      88.8     91.13     92.61     94.08     96.47
## Apps    93.05    93.08      93.1     93.13     93.16     93.17     93.17
##       15 comps  16 comps  17 comps
## X        97.79     99.11    100.00
## Apps     93.17     93.18     93.18
```



We see that the smallest cross-validation error occurs when M = 11 componentsare used Hence ,M selected by cross-validation is 11.

Test Error PLS model

```
## [1] 1494427
```

The mean squared error,TEST MSE is $ pls_{MSE}$=1494427 which smaller than the test error of pcr model.

(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?

Disscussion

Coefficient of determination, often referred to as R^2, represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. In our case, R^2 for all models are close to 0.9, hence, Ordinary least squares, PLS regression, lasso, and PCR regression predicted with high accuracy and almost equally. Lasso reduces the F.Undergrad and Books variables to zero and shrinks coefficients of other variables but its accuracy was poor than other models. Eventhough, PCR regression used all variables which means no dimension reduction occurs, the predection acuracy was high.

The follwing table and graphs shows the R^2 for all fitted models.

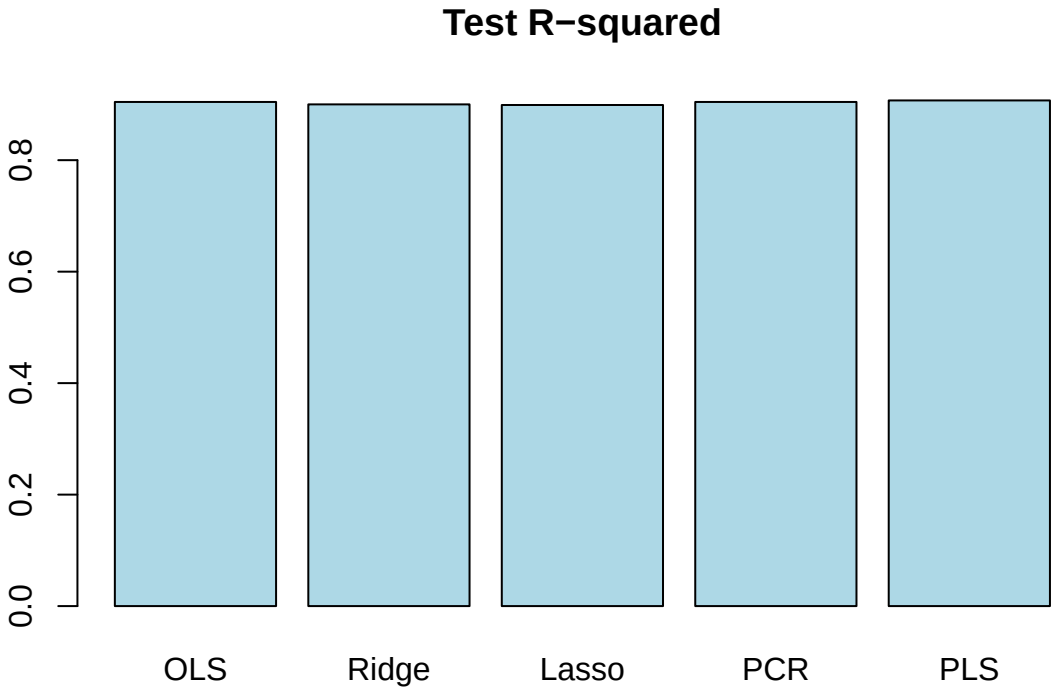
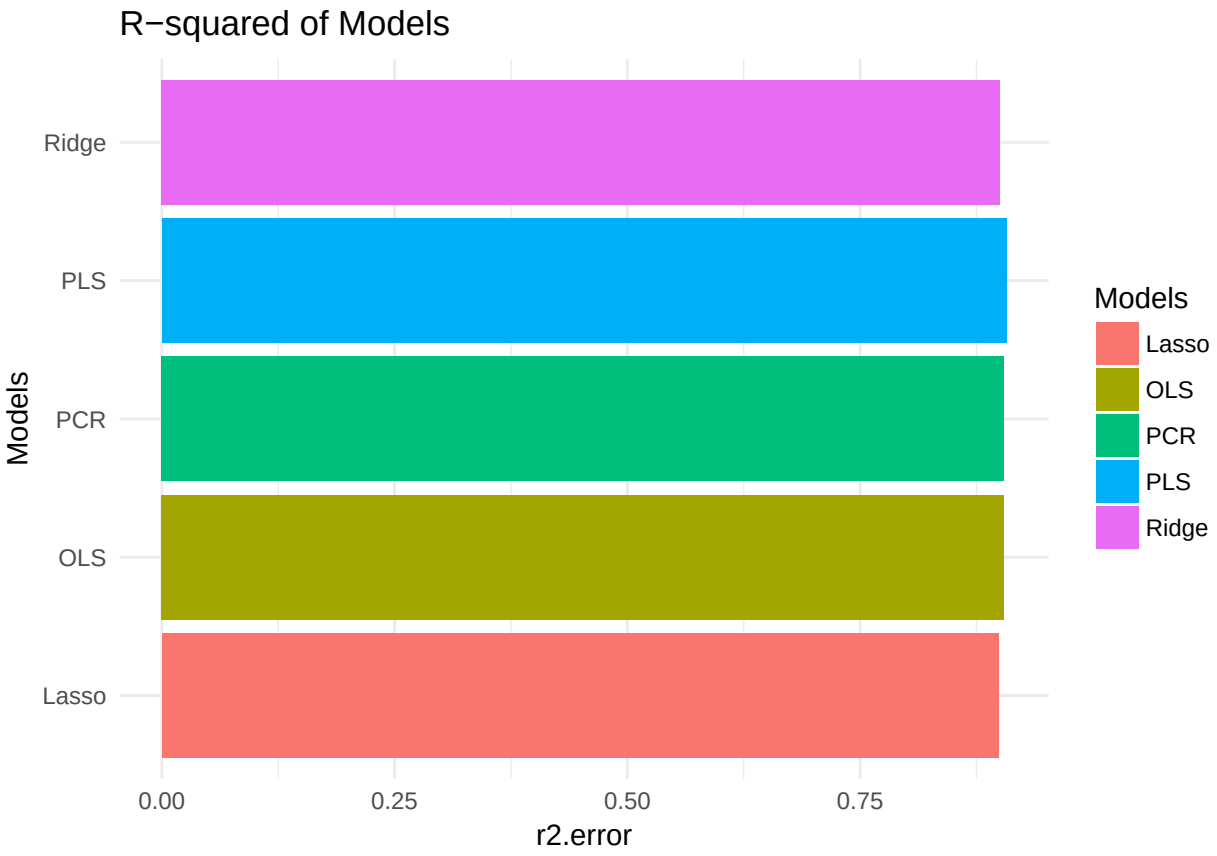


Table 3: Test R-squared

Models	r2.error
OLS	0.9044281
Ridge	0.9000536
Lasso	0.8989591
PCR	0.9044281
PLS	0.9071624

ggplot



3. Question 6.8.11 pg 26

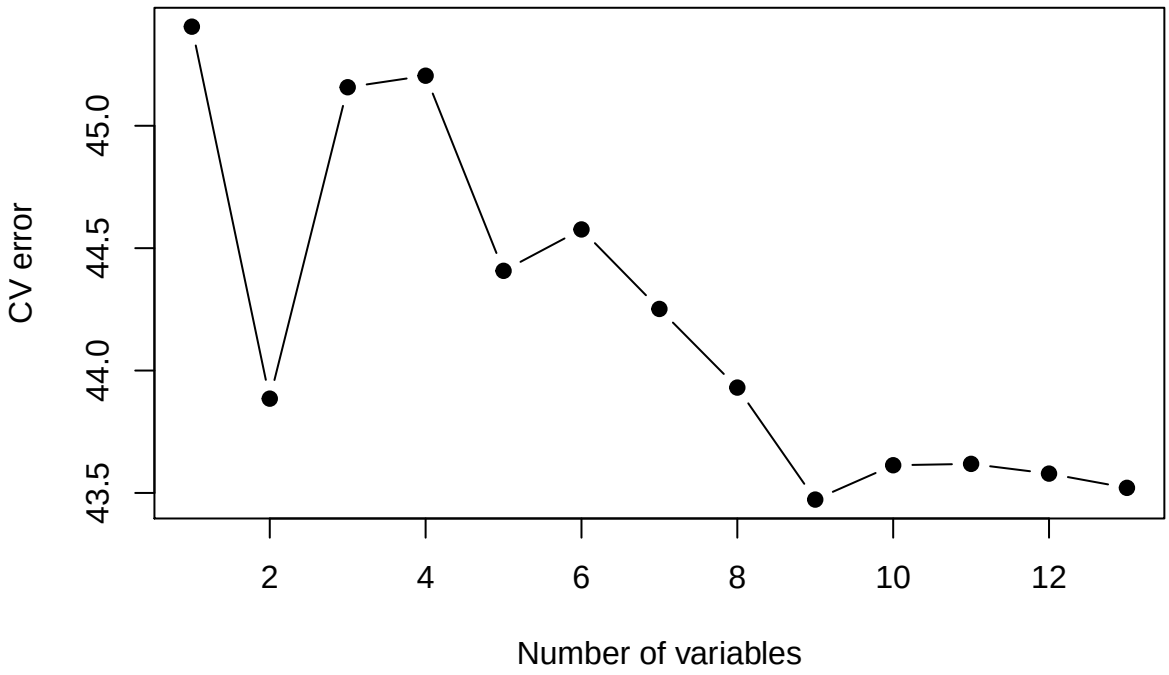
11. We will now try to predict per capita crime rate in the Boston dataset.

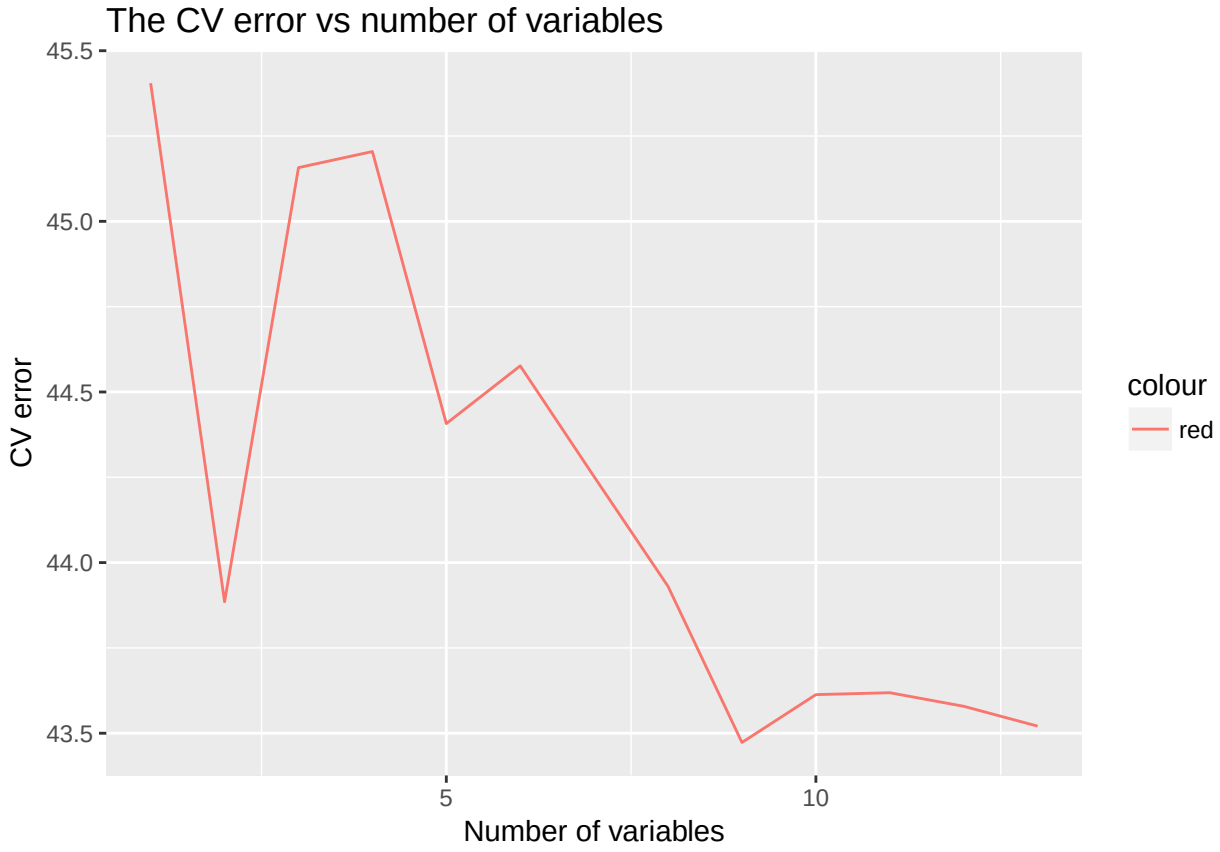
(a)

Try out some of the regression methods explored in this chapter,such as best subset selection, the lasso, ridge regression, and PCR. Present and discuss results for the approaches that you consider.

Best subset selection

crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
0.00632	18	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
0.02731	0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
0.02729	0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
0.03237	0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
0.06905	0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2
0.02985	0	2.18	0	0.458	6.430	58.7	6.0622	3	222	18.7	394.12	5.21	28.7





ggplot

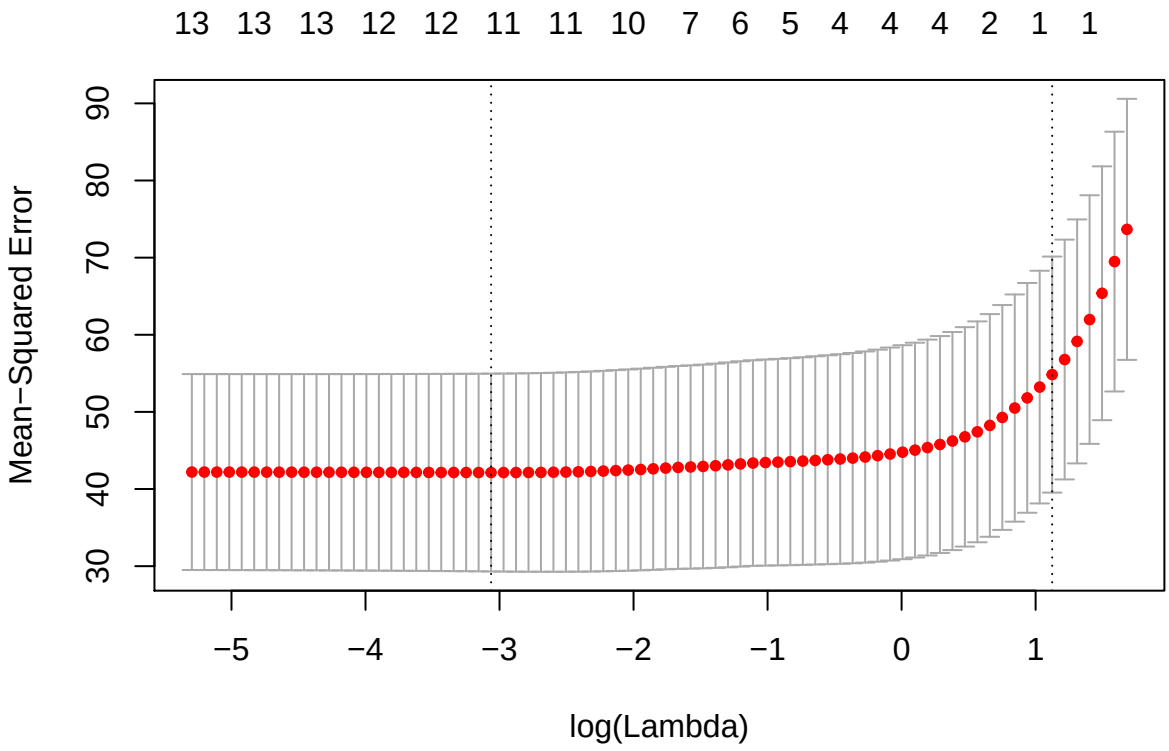
Test error and number of variables selected

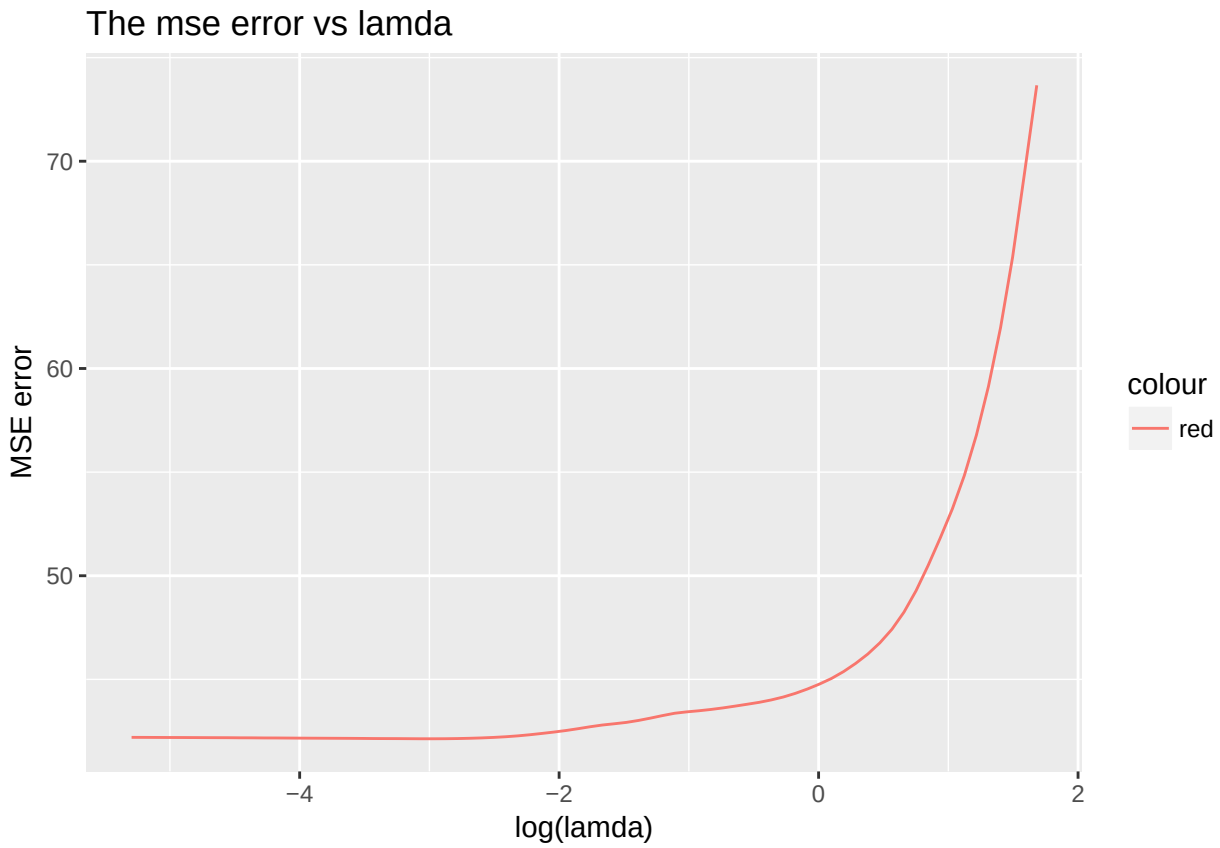
```
## [1] 9
```

```
## [1] 43.47287
```

Test error for best subset selection, MSE= 43.47287 with 9 variables.

Lasso





ggplot

coefficents, best lamda and Error Rate

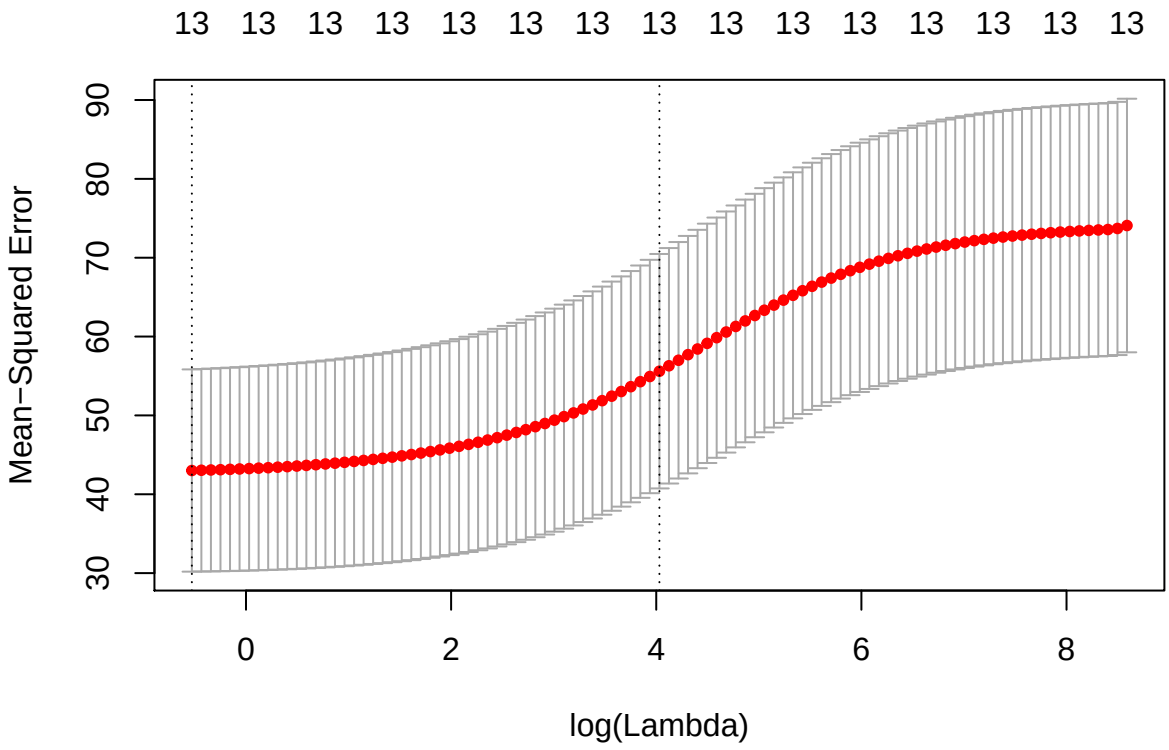
```
## 14 x 1 sparse Matrix of class "dgCMatrix"
##                      1
## (Intercept) 1.0894283
## zn          .
## indus       .
## chas        .
## nox         .
## rm          .
## age         .
## dis         .
## rad         0.2643196
## tax         .
## ptratio     .
## black       .
## lstat       .
## medv        .
```

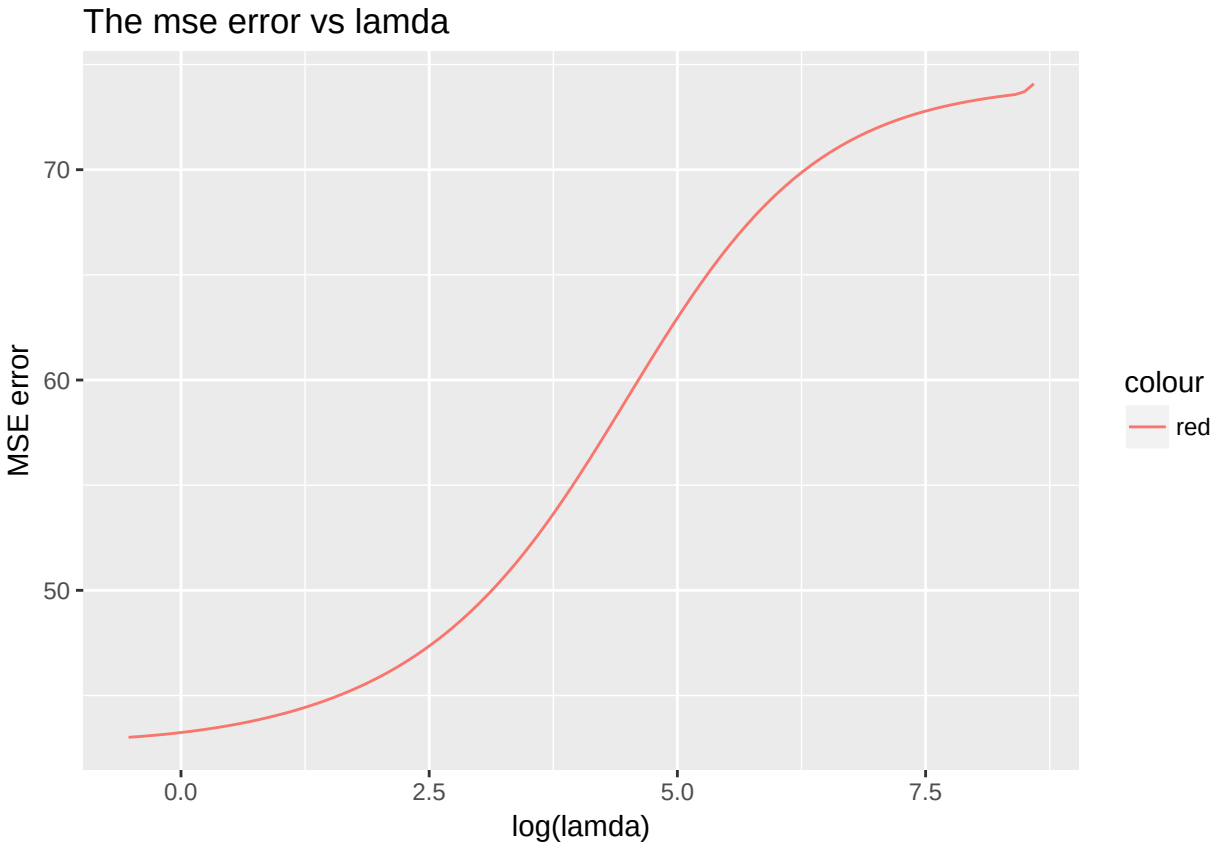
```
## [1] 0.04674894
```

```
## [1] 54.83663
```

Test error is 54.83663 with best lamda 0.04674894

Ridge regression





ggplot

coefficents, best lamda and Error Rate

```
## 14 x 1 sparse Matrix of class "dgCMatrix"
##                        1
## (Intercept)  1.017516864
## zn          -0.002805664
## indus        0.034405928
## chas        -0.225250602
## nox          2.249887499
## rm          -0.162546004
## age          0.007343331
## dis         -0.114928730
## rad          0.059813844
## tax          0.002659110
## ptratio      0.086423005
## black       -0.003342067
## lstat        0.044495213
## medv        -0.029124577
```

```
## [1] 0.5899047
```

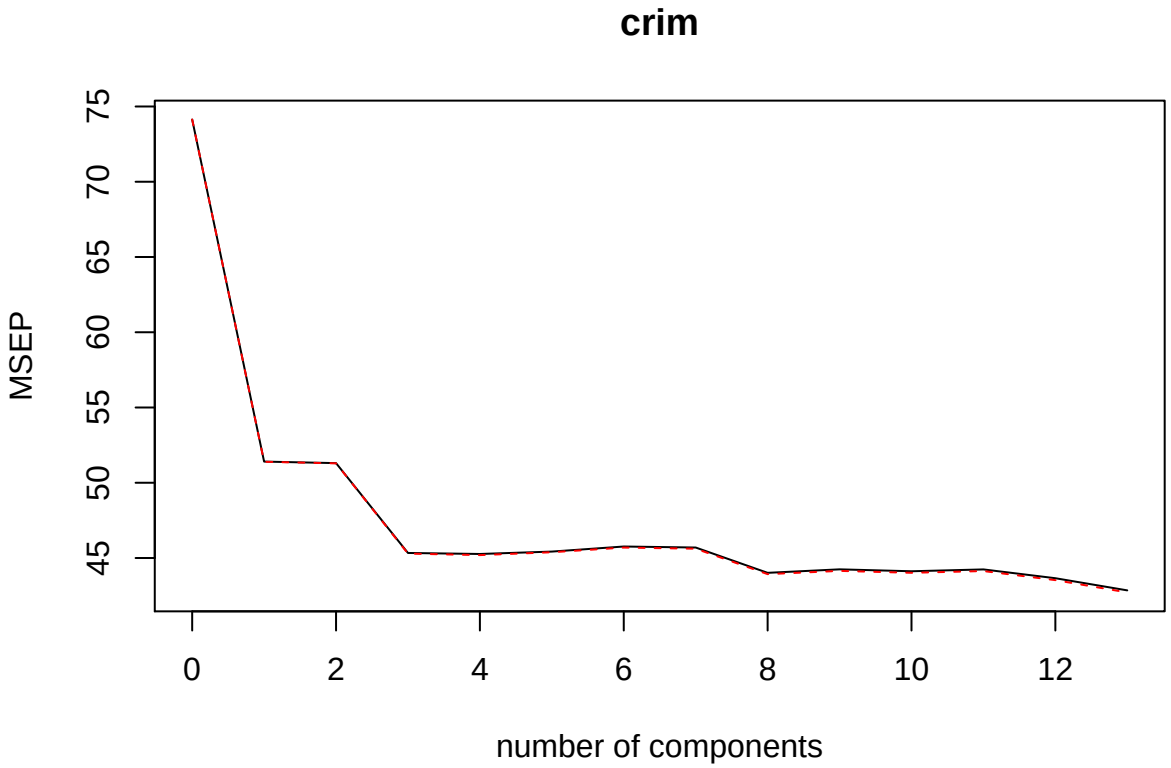
```
## [1] 55.60604
```

Test error is 54.83705 with best lamda 0.5899047

PCR

```
## 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
## CV            8.61    7.170    7.163    6.733    6.728    6.740    6.765
## adjCV         8.61    7.169    7.162    6.730    6.723    6.737    6.760
##        7 comps  8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## CV       6.760    6.634    6.652     6.642     6.652     6.607     6.546
## adjCV    6.754    6.628    6.644     6.635     6.643     6.598     6.536
##
## TRAINING: % variance explained
##       1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
## X       47.70    60.36    69.67    76.45    82.99    88.00    91.14
## crim    30.69    30.87    39.27    39.61    39.61    39.86    40.14
##       8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## X       93.45    95.40     97.04     98.46     99.52     100.0
## crim    42.47    42.55     42.78     43.04     44.13      45.4
```



We see smallest cross validation error at M=13, which is root mean squre value , to calculate MSE =(6.594)^2=43.480836

Disscussion

Using a best subset selection approach, the cross-validation methods selects 9 variable model and its CV estimate for test MSE is 43.47287. The lasso model selects minimum lambda value of 0.04674894 whith the model CV estimate for test MSE is 54.83663. The ridge regression selects minimum lambda value of 0.5899047 and the model CV estimate for test MSE is 54.83705. The pcr model shows that 13 variables,indicates no dimension reduction occur and CV estimate for its test MSE is 43.480836.

(b)

Propose a model (or set of models) that seem to perform well onthis data set, and justify your answer. Make sure that you are evaluating model performance using validation set error, crossvalidation,or some other reasonable alternative, as opposed to using training error.

Ans By comaring the cross validation error, best subset selection method with 9 variables is the best model for this data.(with lowest CV ERROR)

(c)

Does your chosen model involve all of the features in the dataset? Why or why not?

Ans No, subset selection approach uses 9 variables.

Q4.

In the past couple of homework assignments you have used different classification methods to analyze the dataset you chose. For this homework, please write a summary report including but not limited to i) Introduction to the dataset - (response, predictor variables, number of observations, and number of predictors) ii) The question you are trying to address iii) Initial cleaning of the data performed iv) Initial descriptive (numerical summary and graphical - only relevant ones) analysis done v) Classification methods used vi) Choice of the model - test error/cross validation vii) Conclusion and discussion (refer back to the question you are trying to address) viii) Write the report neatly!

ANS I HAVE SUBMITED SEPRATE PDF FILE. I could not include image.