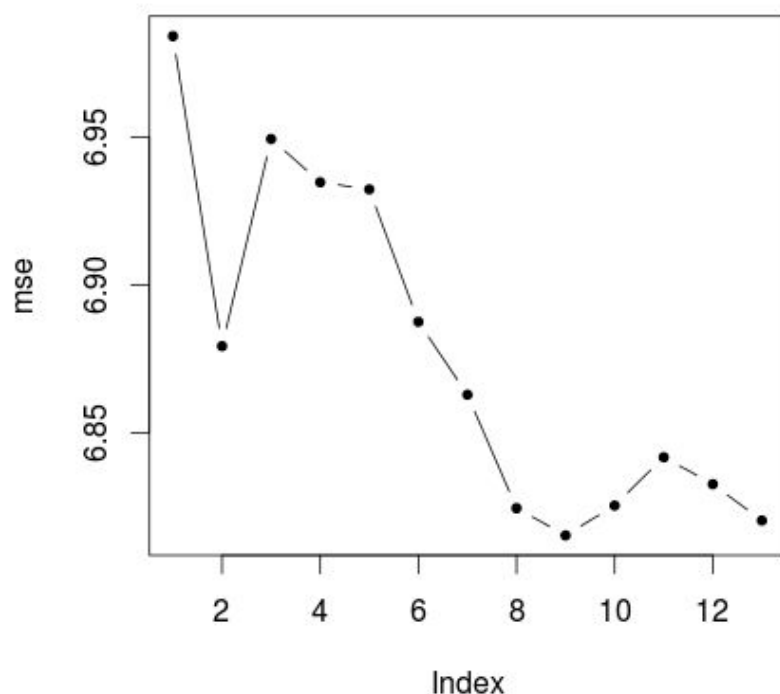


11. We will now try to predict per capita crime rate in the Boston data set.

(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



How many variables were in the best subset? → 9

```
which.min(mse)
```

```
# [1] 9
```

What is the test error for that subset?

```
mse[which.min(mse)]
```

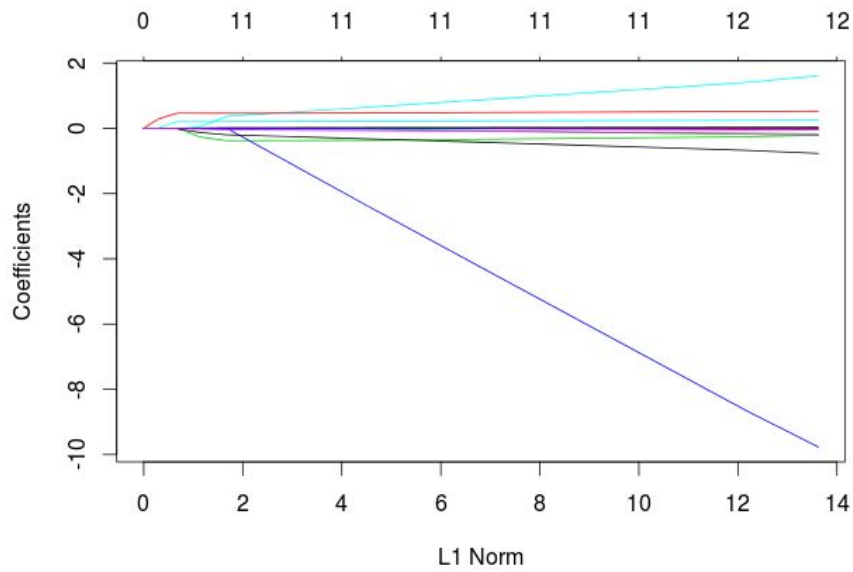
```
# [1] 6.81526
```

What are the chosen variables and their coefficients?

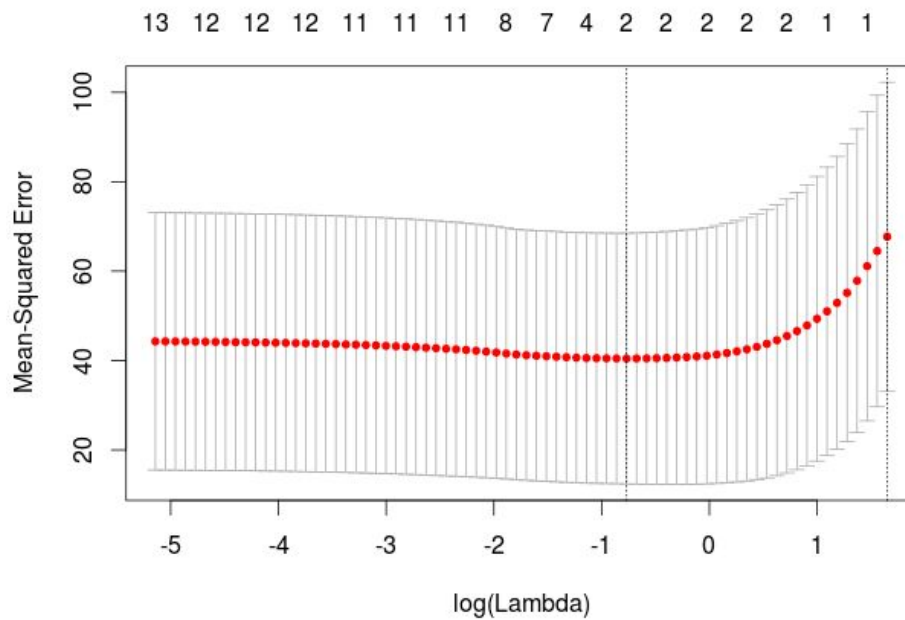
(Intercept)	zn	indus	nox	dis	rad	ptratio	black	lstat	medv
19.12	0.042	-0.09	-10.46	-1.00	0.53	-0.27	-0.00	0.11	-0.18

- Lasso

Coefficients plot:



Lambdas:



Best lambda:

```
bestlam <- cv.out$lambda.min
```

```
bestlam
```

```
# [1] 0.4634552
```

MSE associated with it:

```
# [1] 48.59338
```

Coefficients estimates after fitting the model to the full data set:

(Intercept) 1.148570770

zn .

indus .

chas .

nox .

rm .

age .

dis -0.045523304

rad 0.459539406

tax .

ptratio .

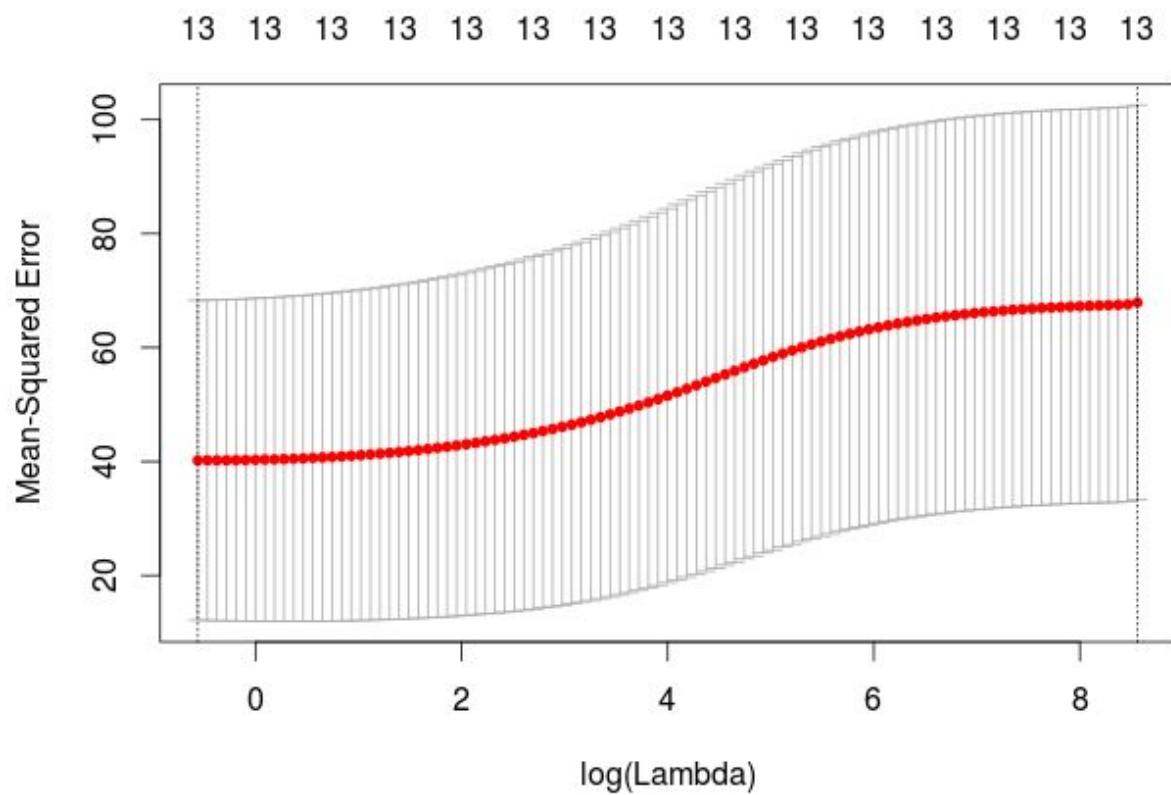
black -0.006068077

lstat 0.125501974

medv -0.052113886

- Ridge regression

First I find the best  $\lambda$ :



```
> bestlam <- cv.out$lambda.min
```

```
> bestlam
```

```
[1] 0.5713693
```

What is the MSE associated with this lambda?

```
> ridge.pred <- predict(ridge.mod, s = bestlam, newx = x[test,])
```

```
> mean((ridge.pred - y.test)^2)
```

```
[1] 47.51344
```

Coefficient estimates:

(Intercept) 8.775348879

zn 0.032582156

indus -0.081488862

chas -0.739172099

nox -5.200862298

rm 0.330930067

age 0.002035181

dis -0.690272029

rad 0.417258521

tax 0.003591263

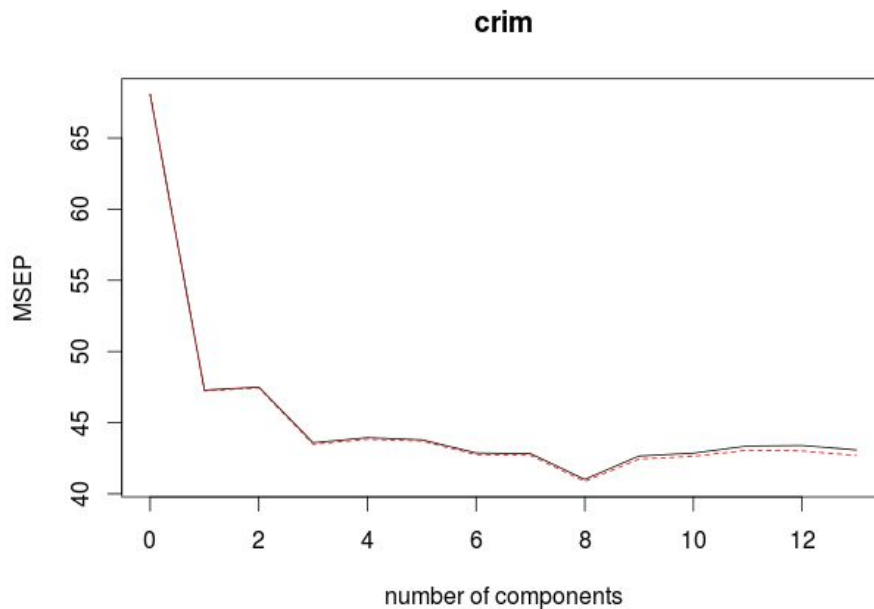
ptratio -0.130549272

black -0.008515279

lstat 0.142676269

medv -0.137473899

- PCR



It seems that the lowest error occurs when 8 components are used. The test error is:  
# [1] 48.45407

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

It looks like the lasso method achieved the lowest error value of 0.47. However, the results from ridge regression, best subset selection and pcr were very close to the result of the lasso.

**(c) Does your chosen model involve all of the features in the data set? Why or why not?**  
No, because 8 of the coefficients were determined to be exactly 0, thus only 5 predictors were used to create the model.