Assignment 8

Question 11.1

Using the crime data set uscrime.txt from Questions 8.2, 9.1, and 10.1, build a regression model using: 1. Stepwise regression 2. Lasso 3. Elastic net For Parts 2 and 3, remember to scale the data first – otherwise, the regression coefficients will be on different scales and the constraint won't have the desired effect.

Answer 11.1

1. Stepwise Regression

```
library(MASS)
library(caret)
library(glmnet)
```

```
#Getting the data
crime_data<-read.table('uscrime.txt', sep = "", header = TRUE)

#Set seed
set.seed(42)

#Fitting the full model
lmr<-lm(Crime~., data = crime_data)
summary(lmr)
```

```
##
## Call:
## lm(formula = Crime ~ ., data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -395.74  -98.09   -6.69  112.99  512.67
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.984e+03  1.628e+03  -3.675 0.000893 ***
## M            8.783e+01  4.171e+01   2.106 0.043443 *
## So          -3.803e+00  1.488e+02  -0.026 0.979765
## Ed           1.883e+02  6.209e+01   3.033 0.004861 **
## Po1          1.928e+02  1.061e+02   1.817 0.078892 .
## Po2         -1.094e+02  1.175e+02  -0.931 0.358830
## LF          -6.638e+02  1.470e+03  -0.452 0.654654
## M.F          1.741e+01  2.035e+01   0.855 0.398995
## Pop         -7.330e-01  1.290e+00  -0.568 0.573845
```

```
## NW           4.204e+00  6.481e+00   0.649 0.521279
## U1          -5.827e+03  4.210e+03  -1.384 0.176238
## U2           1.678e+02  8.234e+01   2.038 0.050161 .
## Wealth       9.617e-02  1.037e-01   0.928 0.360754
## Ineq         7.067e+01  2.272e+01   3.111 0.003983 **
## Prob        -4.855e+03  2.272e+03  -2.137 0.040627 *
## Time        -3.479e+00  7.165e+00  -0.486 0.630708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209.1 on 31 degrees of freedom
## Multiple R-squared:  0.8031, Adjusted R-squared:  0.7078
## F-statistic: 8.429 on 15 and 31 DF,  p-value: 3.539e-07
```

```r
#Using Stepwise regression and evaluating the results
step<-stepAIC(lmr, scope = list(lower = Crime~1, upper = Crime~.),
              direction = 'both',trace = 1)
```

```
## Start:  AIC=514.65
## Crime ~ M + So + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 +
##     U2 + Wealth + Ineq + Prob + Time
##
##          Df Sum of Sq     RSS    AIC
## - So      1        29 1354974 512.65
## - LF      1      8917 1363862 512.96
## - Time    1     10304 1365250 513.00
## - Pop     1     14122 1369068 513.14
## - NW      1     18395 1373341 513.28
## - M.F     1     31967 1386913 513.74
## - Wealth  1     37613 1392558 513.94
## - Po2     1     37919 1392865 513.95
## <none>                1354946 514.65
## - U1      1     83722 1438668 515.47
## - Po1     1    144306 1499252 517.41
## - U2      1    181536 1536482 518.56
## - M       1    193770 1548716 518.93
## - Prob    1    199538 1554484 519.11
## - Ed      1    402117 1757063 524.86
## - Ineq    1    423031 1777977 525.42
##
## Step:  AIC=512.65
## Crime ~ M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 +
##     Wealth + Ineq + Prob + Time
##
##          Df Sum of Sq     RSS    AIC
## - Time    1     10341 1365315 511.01
## - LF      1     10878 1365852 511.03
## - Pop     1     14127 1369101 511.14
## - NW      1     21626 1376600 511.39
## - M.F     1     32449 1387423 511.76
## - Po2     1     37954 1392929 511.95
## - Wealth  1     39223 1394197 511.99
## <none>                1354974 512.65
## - U1      1     96420 1451395 513.88
```

```
## + So      1        29 1354946 514.65
## - Po1     1    144302 1499277 515.41
## - U2      1    189859 1544834 516.81
## - M       1    195084 1550059 516.97
## - Prob    1    204463 1559437 517.26
## - Ed      1    403140 1758114 522.89
## - Ineq    1    488834 1843808 525.13
##
## Step:  AIC=511.01
## Crime ~ M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 +
##     Wealth + Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## - LF      1     10533 1375848 509.37
## - NW      1     15482 1380797 509.54
## - Pop     1     21846 1387161 509.75
## - Po2     1     28932 1394247 509.99
## - Wealth  1     36070 1401385 510.23
## - M.F     1     41784 1407099 510.42
## <none>                1365315 511.01
## - U1      1     91420 1456735 512.05
## + Time    1     10341 1354974 512.65
## + So      1        65 1365250 513.00
## - Po1     1    134137 1499452 513.41
## - U2      1    184143 1549458 514.95
## - M       1    186110 1551425 515.01
## - Prob    1    237493 1602808 516.54
## - Ed      1    409448 1774763 521.33
## - Ineq    1    502909 1868224 523.75
##
## Step:  AIC=509.37
## Crime ~ M + Ed + Po1 + Po2 + M.F + Pop + NW + U1 + U2 + Wealth +
##     Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## - NW      1     11675 1387523 507.77
## - Po2     1     21418 1397266 508.09
## - Pop     1     27803 1403651 508.31
## - M.F     1     31252 1407100 508.42
## - Wealth  1     35035 1410883 508.55
## <none>                1375848 509.37
## - U1      1     80954 1456802 510.06
## + LF      1     10533 1365315 511.01
## + Time    1      9996 1365852 511.03
## + So      1      3046 1372802 511.26
## - Po1     1    123896 1499744 511.42
## - U2      1    190746 1566594 513.47
## - M       1    217716 1593564 514.27
## - Prob    1    226971 1602819 514.54
## - Ed      1    413254 1789103 519.71
## - Ineq    1    500944 1876792 521.96
##
## Step:  AIC=507.77
## Crime ~ M + Ed + Po1 + Po2 + M.F + Pop + U1 + U2 + Wealth + Ineq +
```

```
##     Prob
##
##          Df Sum of Sq     RSS    AIC
## - Po2     1     16706 1404229 506.33
## - Pop     1     25793 1413315 506.63
## - M.F     1     26785 1414308 506.66
## - Wealth  1     31551 1419073 506.82
## <none>                1387523 507.77
## - U1      1     83881 1471404 508.52
## + NW      1     11675 1375848 509.37
## + So      1      7207 1380316 509.52
## + LF      1      6726 1380797 509.54
## + Time    1      4534 1382989 509.61
## - Po1     1    118348 1505871 509.61
## - U2      1    201453 1588976 512.14
## - Prob    1    216760 1604282 512.59
## - M       1    309214 1696737 515.22
## - Ed      1    402754 1790276 517.74
## - Ineq    1    589736 1977259 522.41
##
## Step:  AIC=506.33
## Crime ~ M + Ed + Po1 + M.F + Pop + U1 + U2 + Wealth + Ineq +
##     Prob
##
##          Df Sum of Sq     RSS    AIC
## - Pop     1     22345 1426575 505.07
## - Wealth  1     32142 1436371 505.39
## - M.F     1     36808 1441037 505.54
## <none>                1404229 506.33
## - U1      1     86373 1490602 507.13
## + Po2     1     16706 1387523 507.77
## + NW      1      6963 1397266 508.09
## + So      1      3807 1400422 508.20
## + LF      1      1986 1402243 508.26
## + Time    1       575 1403654 508.31
## - U2      1    205814 1610043 510.76
## - Prob    1    218607 1622836 511.13
## - M       1    307001 1711230 513.62
## - Ed      1    389502 1793731 515.83
## - Ineq    1    608627 2012856 521.25
## - Po1     1   1050202 2454432 530.57
##
## Step:  AIC=505.07
## Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Wealth + Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## - Wealth  1     26493 1453068 503.93
## <none>                1426575 505.07
## - M.F     1     84491 1511065 505.77
## - U1      1     99463 1526037 506.24
## + Pop     1     22345 1404229 506.33
## + Po2     1     13259 1413315 506.63
## + NW      1      5927 1420648 506.87
## + So      1      5724 1420851 506.88
```

```
## + LF      1      5176 1421398 506.90
## + Time    1      3913 1422661 506.94
## - Prob    1    198571 1625145 509.20
## - U2      1    208880 1635455 509.49
## - M       1    320926 1747501 512.61
## - Ed      1    386773 1813348 514.35
## - Ineq    1    594779 2021354 519.45
## - Po1     1   1127277 2553852 530.44
##
## Step:  AIC=503.93
## Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob
##
##          Df Sum of Sq     RSS    AIC
## <none>                1453068 503.93
## + Wealth  1     26493 1426575 505.07
## - M.F     1    103159 1556227 505.16
## + Pop     1     16697 1436371 505.39
## + Po2     1     14148 1438919 505.47
## + So      1      9329 1443739 505.63
## + LF      1      4374 1448694 505.79
## + NW      1      3799 1449269 505.81
## + Time    1      2293 1450775 505.86
## - U1      1    127044 1580112 505.87
## - Prob    1    247978 1701046 509.34
## - U2      1    255443 1708511 509.55
## - M       1    296790 1749858 510.67
## - Ed      1    445788 1898855 514.51
## - Ineq    1    738244 2191312 521.24
## - Po1     1   1672038 3125105 537.93
```

```r
#Model summary
summary(step)
```

```
##
## Call:
## lm(formula = Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob,
##     data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -444.70 -111.07    3.03  122.15  483.30
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6426.10    1194.61  -5.379 4.04e-06 ***
## M              93.32      33.50   2.786  0.00828 **
## Ed            180.12      52.75   3.414  0.00153 **
## Po1           102.65      15.52   6.613 8.26e-08 ***
## M.F            22.34      13.60   1.642  0.10874
## U1          -6086.63    3339.27  -1.823  0.07622 .
## U2            187.35      72.48   2.585  0.01371 *
## Ineq           61.33      13.96   4.394 8.63e-05 ***
## Prob        -3796.03    1490.65  -2.547  0.01505 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 195.5 on 38 degrees of freedom
## Multiple R-squared:  0.7888, Adjusted R-squared:  0.7444
## F-statistic: 17.74 on 8 and 38 DF,  p-value: 1.159e-10
```

```
step$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Crime ~ M + So + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 +
##     U2 + Wealth + Ineq + Prob + Time
##
## Final Model:
## Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob
##
##
##       Step Df    Deviance Resid. Df Resid. Dev      AIC
## 1                                31    1354946 514.6488
## 2     - So  1    28.57405        32    1354974 512.6498
## 3   - Time  1 10340.66984        33    1365315 511.0072
## 4     - LF  1 10533.15902        34    1375848 509.3684
## 5     - NW  1 11674.63991        35    1387523 507.7655
## 6    - Po2  1 16706.34095        36    1404229 506.3280
## 7    - Pop  1 22345.36638        37    1426575 505.0700
## 8 - Wealth  1 26493.24677        38    1453068 503.9349
```

```
#We can see that the step model predicts 8 predictors for the best model.
#Now we will build a new regression model using just the 8 predictors and
#evaluate the model

#New model
new<-lm(Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob, data = crime_data)

#New model summary
summary(new)
```

```
##
## Call:
## lm(formula = Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob,
##     data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -444.70 -111.07    3.03  122.15  483.30
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6426.10    1194.61  -5.379 4.04e-06 ***
## M              93.32      33.50   2.786  0.00828 **
## Ed            180.12      52.75   3.414  0.00153 **
```

```
## Po1           102.65      15.52   6.613 8.26e-08 ***
## M.F            22.34      13.60   1.642  0.10874
## U1          -6086.63    3339.27  -1.823  0.07622 .
## U2            187.35      72.48   2.585  0.01371 *
## Ineq           61.33      13.96   4.394 8.63e-05 ***
## Prob        -3796.03    1490.65  -2.547  0.01505 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 195.5 on 38 degrees of freedom
## Multiple R-squared:  0.7888, Adjusted R-squared:  0.7444
## F-statistic: 17.74 on 8 and 38 DF,  p-value: 1.159e-10
```

```
#Now to further evaluate this model let use cross validation
#Linear regression model using cross validation
set.seed(123)
train.control <- trainControl(method = "cv", number = 5)
model<-train(Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob,
               data = crime_data, method='lm',  #Linear Model2 with CV
               trControl=train.control)
# CV Model summary
summary(model)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -444.70 -111.07    3.03  122.15  483.30
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6426.10    1194.61  -5.379 4.04e-06 ***
## M              93.32      33.50   2.786  0.00828 **
## Ed            180.12      52.75   3.414  0.00153 **
## Po1           102.65      15.52   6.613 8.26e-08 ***
## M.F            22.34      13.60   1.642  0.10874
## U1          -6086.63    3339.27  -1.823  0.07622 .
## U2            187.35      72.48   2.585  0.01371 *
## Ineq           61.33      13.96   4.394 8.63e-05 ***
## Prob        -3796.03    1490.65  -2.547  0.01505 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 195.5 on 38 degrees of freedom
## Multiple R-squared:  0.7888, Adjusted R-squared:  0.7444
## F-statistic: 17.74 on 8 and 38 DF,  p-value: 1.159e-10
```

```
#Note: The summary shows M.F and U1 are not significant based on the p-value
#(0.05). We can further test this by removing both M.F and U1

#New model
```

```r
new_1<-lm(Crime ~ M + Ed + Po1 + U2 + Ineq + Prob, data = crime_data)

#New model summary
summary(new_1)
```

```
##
## Call:
## lm(formula = Crime ~ M + Ed + Po1 + U2 + Ineq + Prob, data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -470.68  -78.41  -19.68  133.12  556.23
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5040.50     899.84  -5.602 1.72e-06 ***
## M             105.02      33.30   3.154  0.00305 **
## Ed            196.47      44.75   4.390 8.07e-05 ***
## Po1           115.02      13.75   8.363 2.56e-10 ***
## U2             89.37      40.91   2.185  0.03483 *
## Ineq           67.65      13.94   4.855 1.88e-05 ***
## Prob        -3801.84    1528.10  -2.488  0.01711 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 200.7 on 40 degrees of freedom
## Multiple R-squared:  0.7659, Adjusted R-squared:  0.7307
## F-statistic: 21.81 on 6 and 40 DF,  p-value: 3.418e-11
```

```r
#Now to further evaluate this model let use cross validation
#Linear regression model using cross validation
set.seed(123)
train.control <- trainControl(method = "cv", number = 5)
model1<-train(Crime ~ M + Ed + Po1 + U2 + Ineq + Prob,
             data = crime_data, method='lm',  #Linear Model2 with CV
             trControl=train.control)

# CV Model summary
summary(model1)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -470.68  -78.41  -19.68  133.12  556.23
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5040.50     899.84  -5.602 1.72e-06 ***
## M             105.02      33.30   3.154  0.00305 **
## Ed            196.47      44.75   4.390 8.07e-05 ***
```

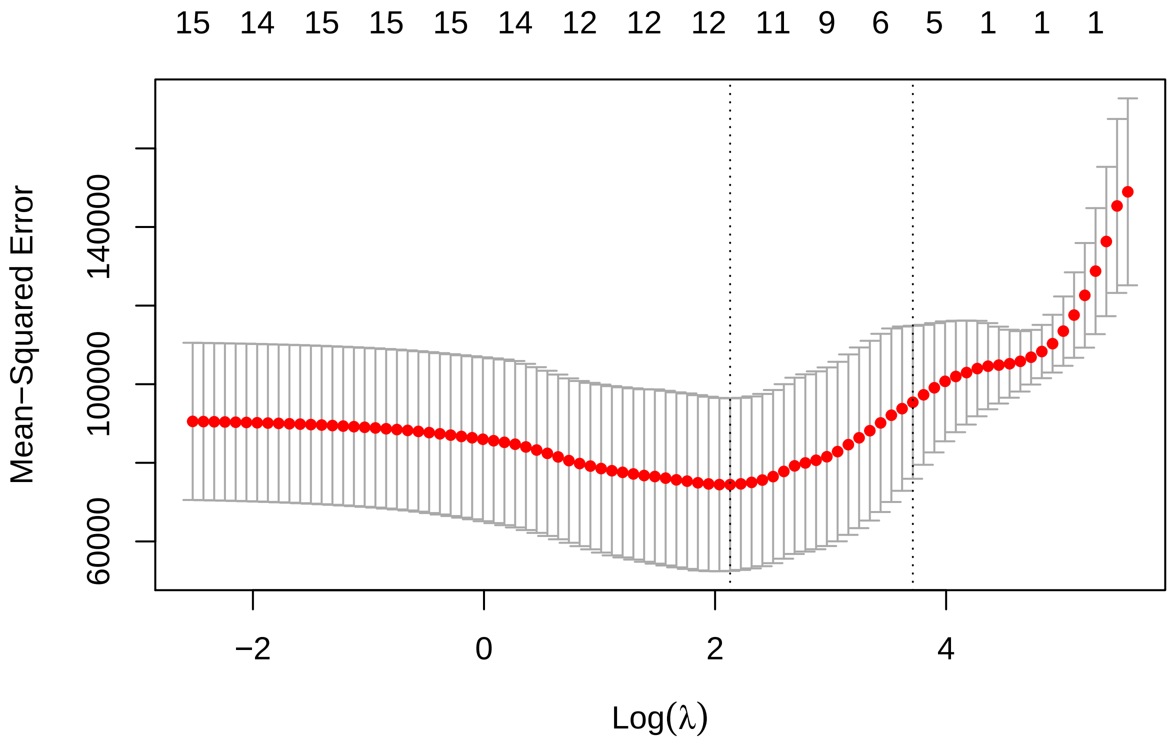
```
## Po1           115.02      13.75   8.363 2.56e-10 ***
## U2             89.37      40.91   2.185  0.03483 *
## Ineq           67.65      13.94   4.855 1.88e-05 ***
## Prob        -3801.84    1528.10  -2.488  0.01711 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 200.7 on 40 degrees of freedom
## Multiple R-squared:  0.7659, Adjusted R-squared:  0.7307
## F-statistic: 21.81 on 6 and 40 DF,  p-value: 3.418e-11
```

Conclusion: Stepwise Regression

The whole purpose of stepwise regression is to eliminate insignificant factors and make the model simpler. This should not be at the cost of losing important variables.We can see that the new model with only 6 factors produced an adjusted R2 of 73%. The previous model with all the 8 factors produced an adjusted R2 of 74%. So, the metrics for both models is close so we might as well keep the 8 factors i n our model.

2. LASSO

```r
#Let first separate response and predictor for the algorithm
X<-crime_data[,1:15]
y<-crime_data[16]

#We need to also scale the data except the response data
X_scale<-scale(X)

# Setting alpha = 1 implements Lasso regressions
set.seed(42)
lasso_cv <- cv.glmnet(x=as.matrix(X_scale), y=as.matrix(y), alpha = 1,
                      nfolds = 5,type.measure="mse",family="gaussian")

# Plot cross-validation results
plot(lasso_cv)
```




```r
#Lambda that gives minimum mean cross-validated error
lasso_cv$lambda.min
```

```
## [1] 8.417125
```

```r
#Selected Lambda and the corresponding coefficients
coef(lasso_cv, s=lasso_cv$lambda.min)
```

```
## 16 x 1 sparse Matrix of class "dgCMatrix"
##                      1
## (Intercept) 905.085106
## M            91.605916
## So           20.875623
## Ed          143.159609
## Po1         302.645106
## Po2           .
## LF            .
## M.F          55.995942
## Pop           .
## NW            7.047715
## U1          -41.517302
## U2           78.040254
## Wealth       10.762412
## Ineq        198.440786
## Prob        -83.952696
## Time          .
```

```r
#Now using the above variables from the Lasso regression results and run a new
#regression model

new_from_lasso<-lm(Crime ~ M+So+Ed+Po1+M.F+NW+U1+U2+Ineq+Prob+Wealth,
                   data = crime_data)
summary(new_from_lasso)
```

```
##
## Call:
## lm(formula = Crime ~ M + So + Ed + Po1 + M.F + NW + U1 + U2 +
##     Ineq + Prob + Wealth, data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -408.38  -96.14   -1.39  114.80  454.53
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.757e+03  1.313e+03  -5.147 1.03e-05 ***
## M            9.148e+01  3.893e+01   2.350  0.02454 *
## So           3.335e+01  1.237e+02   0.270  0.78905
## Ed           1.746e+02  5.589e+01   3.124  0.00357 **
## Po1          9.277e+01  2.019e+01   4.596 5.41e-05 ***
## M.F          2.189e+01  1.453e+01   1.506  0.14101
## NW           1.549e+00  5.559e+00   0.279  0.78209
## U1          -5.248e+03  3.600e+03  -1.458  0.15380
## U2           1.667e+02  7.853e+01   2.123  0.04089 *
## Ineq         6.693e+01  2.022e+01   3.310  0.00217 **
## Prob        -3.854e+03  1.770e+03  -2.177  0.03627 *
## Wealth       7.626e-02  9.737e-02   0.783  0.43878
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 201.3 on 35 degrees of freedom
## Multiple R-squared:  0.794,  Adjusted R-squared:  0.7292
## F-statistic: 12.26 on 11 and 35 DF,  p-value: 5.334e-09
```

```r
#Further based on the p-value we can remove So,M.F,NW,U1,Wealth and run a new model
mod<-lm(Crime ~ M+Ed+Po1+U2+Ineq+Prob,
        data = crime_data)
summary(mod)
```

```
##
## Call:
## lm(formula = Crime ~ M + Ed + Po1 + U2 + Ineq + Prob, data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -470.68  -78.41  -19.68  133.12  556.23
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5040.50     899.84  -5.602 1.72e-06 ***
```

```
## M              105.02      33.30   3.154  0.00305 **
## Ed             196.47      44.75   4.390 8.07e-05 ***
## Po1            115.02      13.75   8.363 2.56e-10 ***
## U2              89.37      40.91   2.185  0.03483 *
## Ineq            67.65      13.94   4.855 1.88e-05 ***
## Prob         -3801.84    1528.10  -2.488  0.01711 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 200.7 on 40 degrees of freedom
## Multiple R-squared:  0.7659, Adjusted R-squared:  0.7307
## F-statistic: 21.81 on 6 and 40 DF,  p-value: 3.418e-11
```

```r
#The following model have all variables with significant values so we can keep
#it.Also, it is the same model we got through stepwise regression
```

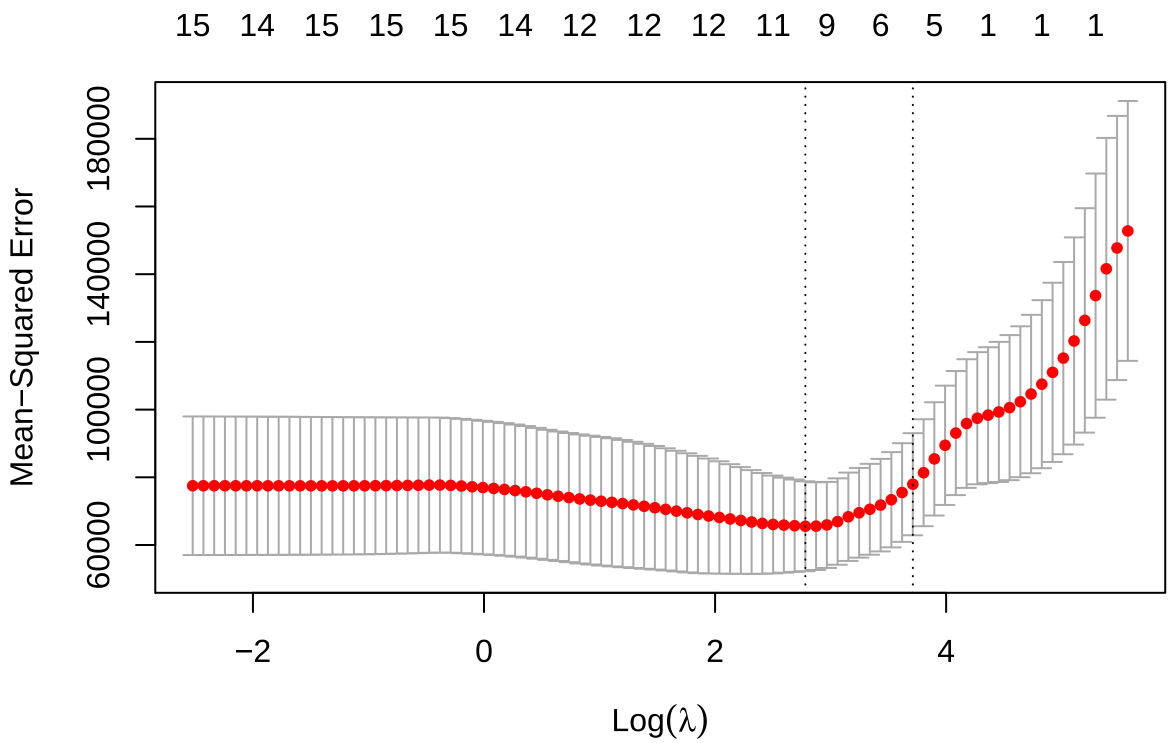
Conclusion: LASSO

We used the Lasso algorithm where the alpha value is 1 for variable selection. The model gave 11 significant variables and these were subsequently used to build a regression model. Further the p-value was used to eliminate the insignificant variable and a final regression model was build which was the same as stepwise regression.

2. Elastic Net

```r
#For elastic net we use different values of alpha and see which model gives
#us the best result.

# Build the model using the values of alpha between 1-0

metric<-list()
rng<-as.list(seq(from = 0, to = 100, by = 1))

for (i in 1:length(rng)) {
  set.seed(42)
  elastic<-cv.glmnet(x=as.matrix(X_scale), y=as.matrix(y),
                        alpha=i/100,nfolds = 10)
  metric[i]=elastic$glmnet.fit$dev.ratio[which(elastic$glmnet.fit$lambda==
                                                        elastic$lambda.min)]
}

# Getting the best alpha from the metric list above
mat<-data.frame(Alpha=unlist(rng), R2=unlist(metric))

#Sorting the dataframe for best alpha
data_sort<-mat[order(-mat$R2),]

#Best Alpha
best_alpha<-data_sort[1,1]

#Now building a elastic net using the best alpha and evaluating the results
```

```r
best_mod<-cv.glmnet(x=as.matrix(X_scale), y=as.matrix(y),
                    alpha=best_alpha,nfolds = 10)

# Plot cross-validation results
plot(best_mod)
```




```r
#Lambda that gives minimum mean cross-validated error
best_mod$lambda.min
```

```
## [1] 16.14329
```

```r
#Selected Lambda and the corresponding coefficients
coef(best_mod, s=best_mod$lambda.min)
```

```
## 16 x 1 sparse Matrix of class "dgCMatrix"
##                        1
## (Intercept) 905.0851064
## M            74.0174295
## So           17.7133651
## Ed           89.5805899
## Po1         309.1881837
## Po2            .
## LF            0.4790569
## M.F          48.4916943
## Pop            .
## NW            3.2640901
## U1             .
## U2           24.9595863
```

```
## Wealth        .
## Ineq        158.3327124
## Prob        -74.8230648
## Time          .
```

```r
#Now using the above variables from the Elastic net and run a new
#regression model

new_from_elastic<-lm(Crime ~ M+So+Ed+Po1+LF+M.F+NW+U2+Ineq+Prob,
                   data = crime_data)
summary(new_from_elastic)
```

```
##
## Call:
## lm(formula = Crime ~ M + So + Ed + Po1 + LF + M.F + NW + U2 +
##     Ineq + Prob, data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -410.55 -121.42    5.76  110.54  550.24
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5725.4999  1184.4847  -4.834 2.49e-05 ***
## M              90.4022    39.7548   2.274  0.02903 *
## So            122.7918   129.8958   0.945  0.35080
## Ed            168.7274    57.8799   2.915  0.00608 **
## Po1           112.2077    16.6064   6.757 6.86e-08 ***
## LF            622.6363  1227.0766   0.507  0.61496
## M.F            10.6943    14.6530   0.730  0.47021
## NW             -0.2803     5.7955  -0.048  0.96169
## U2             86.2961    49.4882   1.744  0.08973 .
## Ineq           57.4296    17.2560   3.328  0.00203 **
## Prob        -4360.5001  1769.9432  -2.464  0.01867 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 206.6 on 36 degrees of freedom
## Multiple R-squared:  0.7767, Adjusted R-squared:  0.7147
## F-statistic: 12.53 on 10 and 36 DF,  p-value: 5.374e-09
```

```r
#Further based on the p-value we can remove So,LF,M.F,NW and run a new model
mod_last<-lm(Crime ~ M+Ed+Po1+U2+Ineq+Prob,
        data = crime_data)
summary(mod_last)
```

```
##
## Call:
## lm(formula = Crime ~ M + Ed + Po1 + U2 + Ineq + Prob, data = crime_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -470.68  -78.41  -19.68  133.12  556.23
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5040.50     899.84  -5.602 1.72e-06 ***
## M             105.02      33.30   3.154  0.00305 **
## Ed            196.47      44.75   4.390 8.07e-05 ***
## Po1           115.02      13.75   8.363 2.56e-10 ***
## U2             89.37      40.91   2.185  0.03483 *
## Ineq           67.65      13.94   4.855 1.88e-05 ***
## Prob        -3801.84    1528.10  -2.488  0.01711 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 200.7 on 40 degrees of freedom
## Multiple R-squared:  0.7659, Adjusted R-squared:  0.7307
## F-statistic: 21.81 on 6 and 40 DF,  p-value: 3.418e-11
```

Conclusion: Elastic Net

We used the Elastic net algorithm where the alpha value is between 0-1 for variable selection. The model gave 10 significant variables and these were subsequently used to build a regression model. Further the p-value was used to eliminate the insignificant variable and a final regression model was build which was the same as Lasso and stepwise regression.