# HW8 - KELLY "SCOTT" SIMS

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. For Parts 2 and 3, use the glmnet function in R. Notes on R: • For the elastic net model, what we called lambda in the videos, glmnet calls "alpha"; you can get a range of results by varying alpha from 1 (lasso) to 0 (ridge regression) [and, of course, other values of alpha in between]. • In a function call like glmnet(x,y,family="mgaussian",alpha=1) the predictors x need to be in R's matrix format, rather than data frame format. You can convert a data frame to a matrix using as.matrix – for example, x <- as.matrix(data[,1:n-1]) • Rather than specifying a value of T, glmnet returns models for a variety of values of T.

```r
library(olsrr)
```

```
##
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##
##     rivers
```

```r
data <- read.table('uscrime.txt', header = TRUE, stringsAsFactors = FALSE)
head(data)
```

```
##      M So   Ed  Po1  Po2    LF   M.F Pop   NW    U1  U2 Wealth Ineq
## 1 15.1  1  9.1  5.8  5.6 0.510  95.0  33 30.1 0.108 4.1   3940 26.1
## 2 14.3  0 11.3 10.3  9.5 0.583 101.2  13 10.2 0.096 3.6   5570 19.4
## 3 14.2  1  8.9  4.5  4.4 0.533  96.9  18 21.9 0.094 3.3   3180 25.0
## 4 13.6  0 12.1 14.9 14.1 0.577  99.4 157  8.0 0.102 3.9   6730 16.7
## 5 14.1  0 12.1 10.9 10.1 0.591  98.5  18  3.0 0.091 2.0   5780 17.4
## 6 12.1  0 11.0 11.8 11.5 0.547  96.4  25  4.4 0.084 2.9   6890 12.6
##       Prob    Time Crime
## 1 0.084602 26.2011   791
## 2 0.029599 25.2999  1635
## 3 0.083401 24.3006   578
## 4 0.015801 29.9012  1969
## 5 0.041399 21.2998  1234
## 6 0.034201 20.9995   682
```

## STEPWISE REGRESSION

For stepweise regression, thankfully R already has a module that will handle the very iterative process of adding and subtracting features while training the model and analyzing it. Let's put that funciton to use and build regression models using **Stepwise Regression**. The *ols_step_both_p* function selects features based on p-value. Because of this, we can set the upper and lower bounds of the p-value for the function to take into consideration when it is selecting its features. We will set the "pent" equal to 0.1, meaning variables with p value less than 0.1 will enter into the model. Will will set "prem" equal to 0.3 meaning variables with p value more than 0.3 will not enter into the model

```r
stepwise.model <- lm(Crime ~., data = data)
step <- ols_step_both_p(stepwise.model, pent = 0.1, prem = 0.3, details = TRUE)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. M
## 2. So
## 3. Ed
## 4. Po1
## 5. Po2
## 6. LF
## 7. M.F
## 8. Pop
## 9. NW
## 10. U1
## 11. U2
## 12. Wealth
## 13. Ineq
## 14. Prob
## 15. Time
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - Po1 added
##
##                           Model Summary
## -----------------------------------------------------------------
## R                        0.688       RMSE               283.926
## R-Squared                0.473       Coef. Var           31.370
## Adj. R-Squared           0.461       MSE              80613.907
## Pred R-Squared           0.393       MAE                218.298
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## -----------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## -----------------------------------------------------------------------
## Regression    3253301.823        1    3253301.823    40.357    0.0000
## Residual      3627625.836       45      80613.907
## Total         6880927.660       46
## -----------------------------------------------------------------------
##
##                                     Parameter Estimates
## -----------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t       Sig        lower      upper
## -----------------------------------------------------------------------------------------
## (Intercept)    144.464       126.693                 1.140     0.260    -110.708    399.636
```

```
##        Po1     89.485        14.086        0.688    6.353    0.000      61.114    117.856
## ------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - Ineq added
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.762       RMSE                256.187
## R-Squared               0.580       Coef. Var            28.305
## Adj. R-Squared          0.561       MSE               65631.982
## Pred R-Squared          0.486       MAE                 189.666
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ----------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression   3993120.467        2    1996560.233    30.421    0.0000
## Residual     2887807.193       44      65631.982
## Total        6880927.660       46
## ----------------------------------------------------------------------
##
##                                     Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t        Sig         lower       upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    -944.662       343.947                -2.747    0.009    -1637.841    -251.482
##         Po1     124.148        16.375        0.954    7.582    0.000       91.147     157.149
##        Ineq      40.953        12.198        0.422    3.357    0.002       16.370      65.536
## ----------------------------------------------------------------------------------------------
##
##
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.762       RMSE                256.187
## R-Squared               0.580       Coef. Var            28.305
## Adj. R-Squared          0.561       MSE               65631.982
## Pred R-Squared          0.486       MAE                 189.666
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ----------------------------------------------------------------------
```

```
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    3993120.467        2    1996560.233    30.421    0.0000
## Residual      2887807.193       44      65631.982
## Total         6880927.660       46
## --------------------------------------------------------------------
##
##                                       Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t        Sig         lower       upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    -944.662       343.947                -2.747    0.009    -1637.841    -251.482
##         Po1     124.148        16.375        0.954    7.582    0.000       91.147     157.149
##        Ineq      40.953        12.198        0.422    3.357    0.002       16.370      65.536
## ----------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 3
##
## - Ed added
##
##                           Model Summary
## --------------------------------------------------------------
## R                        0.816       RMSE               231.314
## R-Squared                0.666       Coef. Var           25.557
## Adj. R-Squared           0.642       MSE              53505.987
## Pred R-Squared           0.575       MAE                165.776
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    4580170.224        3    1526723.408    28.534    0.0000
## Residual      2300757.435       43      53505.987
## Total         6880927.660       46
## --------------------------------------------------------------------
##
##                                        Parameter Estimates
## -----------------------------------------------------------------------------------------------
##       model        Beta     Std. Error    Std. Beta      t        Sig         lower        upper
## -----------------------------------------------------------------------------------------------
## (Intercept)    -3275.409       769.137                -4.259    0.000    -4826.521    -1724.297
##         Po1      124.314        14.785        0.955    8.408    0.000       94.497      154.131
##        Ineq       75.058        15.077        0.774    4.978    0.000       44.652      105.463
##          Ed      157.869        47.661        0.457    3.312    0.002       61.752      253.987
## -----------------------------------------------------------------------------------------------
##
```

```
##
##
##                            Model Summary
## -----------------------------------------------------------------
## R                        0.816       RMSE                231.314
## R-Squared                0.666       Coef. Var            25.557
## Adj. R-Squared           0.642       MSE               53505.987
## Pred R-Squared           0.575       MAE                 165.776
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                   ANOVA
## -----------------------------------------------------------------------
##                    Sum of
##                   Squares        DF    Mean Square      F         Sig.
## -----------------------------------------------------------------------
## Regression    4580170.224         3    1526723.408    28.534    0.0000
## Residual      2300757.435        43      53505.987
## Total         6880927.660        46
## -----------------------------------------------------------------------
##
##                                         Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model          Beta    Std. Error    Std. Beta      t        Sig         lower         upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    -3275.409       769.137                 -4.259    0.000    -4826.521    -1724.297
##         Po1      124.314        14.785        0.955     8.408    0.000       94.497      154.131
##        Ineq       75.058        15.077        0.774     4.978    0.000       44.652      105.463
##          Ed      157.869        47.661        0.457     3.312    0.002       61.752      253.987
## ----------------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - M added
##
##                            Model Summary
## -----------------------------------------------------------------
## R                        0.837       RMSE                221.540
## R-Squared                0.700       Coef. Var            24.477
## Adj. R-Squared           0.672       MSE               49079.828
## Pred R-Squared           0.609       MAE                 155.220
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                   ANOVA
## -----------------------------------------------------------------------
##                    Sum of
##                   Squares        DF    Mean Square      F         Sig.
```

```
## ----------------------------------------------------------------------
## Regression    4819574.863         4    1204893.716     24.55    0.0000
## Residual      2061352.797        42      49079.828
## Total         6880927.660        46
## ----------------------------------------------------------------------
##
##                                       Parameter Estimates
## ------------------------------------------------------------------------------------------------
##       model         Beta    Std. Error    Std. Beta      t        Sig         lower        upper
## ------------------------------------------------------------------------------------------------
## (Intercept)    -4249.222       858.514                 -4.950    0.000     -5981.774    -2516.671
##         Po1      129.804        14.377        0.997     9.029    0.000       100.790      158.818
##        Ineq       64.091        15.270        0.661     4.197    0.000        33.276       94.906
##          Ed      166.050        45.797        0.480     3.626    0.001        73.628      258.472
##           M       76.022        34.421        0.247     2.209    0.033         6.557      145.487
## ------------------------------------------------------------------------------------------------
##
##
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.837       RMSE               221.540
## R-Squared                0.700       Coef. Var           24.477
## Adj. R-Squared           0.672       MSE              49079.828
## Pred R-Squared           0.609       MAE                155.220
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ----------------------------------------------------------------------
##                    Sum of
##                   Squares        DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression    4819574.863         4    1204893.716     24.55    0.0000
## Residual      2061352.797        42      49079.828
## Total         6880927.660        46
## ----------------------------------------------------------------------
##
##                                       Parameter Estimates
## ------------------------------------------------------------------------------------------------
##       model         Beta    Std. Error    Std. Beta      t        Sig         lower        upper
## ------------------------------------------------------------------------------------------------
## (Intercept)    -4249.222       858.514                 -4.950    0.000     -5981.774    -2516.671
##         Po1      129.804        14.377        0.997     9.029    0.000       100.790      158.818
##        Ineq       64.091        15.270        0.661     4.197    0.000        33.276       94.906
##          Ed      166.050        45.797        0.480     3.626    0.001        73.628      258.472
##           M       76.022        34.421        0.247     2.209    0.033         6.557      145.487
## ------------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 5
```

```
##
## - Prob added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                       0.859       RMSE                 209.721
## R-Squared               0.738       Coef. Var             23.171
## Adj. R-Squared          0.706       MSE                43982.690
## Pred R-Squared          0.641       MAE                  147.842
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## -----------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## -----------------------------------------------------------------------
## Regression    5077637.365         5    1015527.473    23.089    0.0000
## Residual      1803290.295        41      43982.690
## Total         6880927.660        46
## -----------------------------------------------------------------------
##
##                                     Parameter Estimates
## -----------------------------------------------------------------------------------------------
##       model         Beta    Std. Error    Std. Beta      t        Sig         lower        upper
## -----------------------------------------------------------------------------------------------
## (Intercept)    -4064.574       816.279                 -4.979    0.000    -5713.084    -2416.064
##         Po1      121.229        14.063        0.932     8.621    0.000       92.829      149.629
##        Ineq       68.310        14.559        0.705     4.692    0.000       38.907       97.714
##          Ed      160.153        43.422        0.463     3.688    0.001       72.460      247.845
##           M       79.689        32.620        0.259     2.443    0.019       13.811      145.566
##        Prob    -3867.271      1596.552       -0.227    -2.422    0.020    -7091.573     -642.969
## -----------------------------------------------------------------------------------------------
##
##
##
##                          Model Summary
## -----------------------------------------------------------------
## R                       0.859       RMSE                 209.721
## R-Squared               0.738       Coef. Var             23.171
## Adj. R-Squared          0.706       MSE                43982.690
## Pred R-Squared          0.641       MAE                  147.842
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## -----------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## -----------------------------------------------------------------------
```

```
## Regression    5077637.365         5    1015527.473    23.089    0.0000
## Residual      1803290.295        41      43982.690
## Total         6880927.660        46
## ---------------------------------------------------------------------
##
##                                        Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model          Beta    Std. Error    Std. Beta      t        Sig         lower        upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    -4064.574       816.279                  -4.979    0.000    -5713.084    -2416.064
##         Po1      121.229        14.063        0.932      8.621    0.000       92.829      149.629
##        Ineq       68.310        14.559        0.705      4.692    0.000       38.907       97.714
##          Ed      160.153        43.422        0.463      3.688    0.001       72.460      247.845
##           M       79.689        32.620        0.259      2.443    0.019       13.811      145.566
##        Prob    -3867.271      1596.552       -0.227     -2.422    0.020    -7091.573     -642.969
## ----------------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 6
##
## - U2 added
##
##                           Model Summary
## ---------------------------------------------------------------
## R                         0.875       RMSE               200.690
## R-Squared                 0.766       Coef. Var           22.174
## Adj. R-Squared            0.731       MSE              40276.421
## Pred R-Squared            0.666       MAE                138.674
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## ---------------------------------------------------------------------
##                    Sum of
##                   Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression    5269870.803         6     878311.801    21.807    0.0000
## Residual      1611056.856        40      40276.421
## Total         6880927.660        46
## ---------------------------------------------------------------------
##
##                                        Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model          Beta    Std. Error    Std. Beta      t        Sig         lower        upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    -5040.505       899.843                  -5.602    0.000    -6859.156    -3221.854
##         Po1      115.024        13.754        0.884      8.363    0.000       87.227      142.821
##        Ineq       67.653        13.936        0.698      4.855    0.000       39.488       95.818
##          Ed      196.471        44.754        0.568      4.390    0.000      106.019      286.923
##           M      105.020        33.299        0.341      3.154    0.003       37.719      172.320
##        Prob    -3801.836      1528.097       -0.224     -2.488    0.017    -6890.236     -713.436
```

```
##         U2      89.366       40.906        0.195     2.185    0.035        6.693      172.039
## ----------------------------------------------------------------------------------------------------
## 
## 
## 
##                           Model Summary
## -----------------------------------------------------------------
## R                       0.875       RMSE                 200.690
## R-Squared               0.766       Coef. Var             22.174
## Adj. R-Squared          0.731       MSE                40276.421
## Pred R-Squared          0.666       MAE                  138.674
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
## 
##                                  ANOVA
## -----------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## -----------------------------------------------------------------------
## Regression    5269870.803         6     878311.801    21.807    0.0000
## Residual      1611056.856        40      40276.421
## Total         6880927.660        46
## -----------------------------------------------------------------------
## 
##                                        Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model         Beta    Std. Error    Std. Beta      t        Sig         lower        upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    -5040.505       899.843                 -5.602    0.000    -6859.156    -3221.854
##        Po1       115.024        13.754        0.884     8.363    0.000       87.227      142.821
##       Ineq        67.653        13.936        0.698     4.855    0.000       39.488       95.818
##         Ed       196.471        44.754        0.568     4.390    0.000      106.019      286.923
##          M       105.020        33.299        0.341     3.154    0.003       37.719      172.320
##       Prob     -3801.836      1528.097       -0.224    -2.488    0.017    -6890.236     -713.436
##         U2        89.366        40.906        0.195     2.185    0.035        6.693      172.039
## ----------------------------------------------------------------------------------------------------
## 
## 
## 
## No more variables to be added/removed.
## 
## 
## Final Model Output
## ------------------
## 
##                           Model Summary
## -----------------------------------------------------------------
## R                       0.875       RMSE                 200.690
## R-Squared               0.766       Coef. Var             22.174
## Adj. R-Squared          0.731       MSE                40276.421
## Pred R-Squared          0.666       MAE                  138.674
## -----------------------------------------------------------------
```

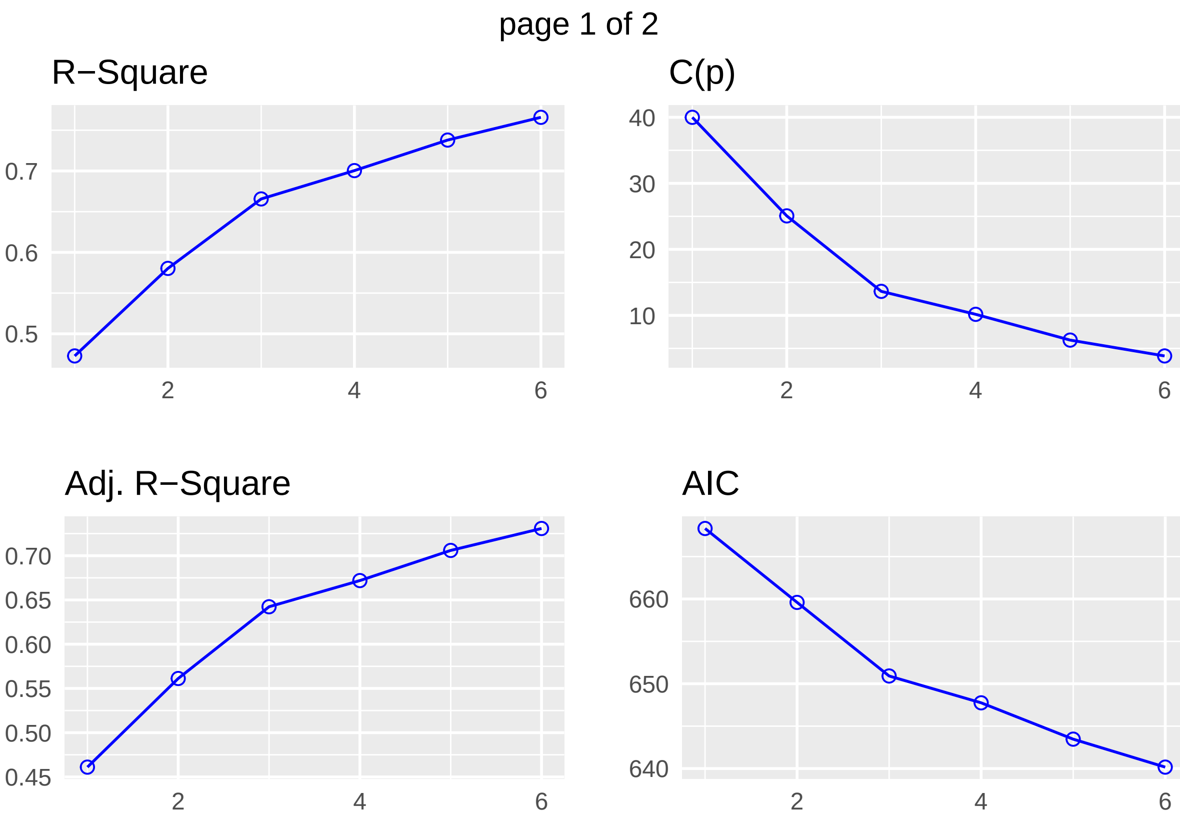
```
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## ----------------------------------------------------------------------
##                     Sum of
##                    Squares        DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression    5269870.803         6     878311.801    21.807    0.0000
## Residual      1611056.856        40      40276.421
## Total         6880927.660        46
## ----------------------------------------------------------------------
##
##                                        Parameter Estimates
## -----------------------------------------------------------------------------------------------------
##       model          Beta    Std. Error    Std. Beta       t        Sig         lower        upper
## -----------------------------------------------------------------------------------------------------
## (Intercept)    -5040.505       899.843                  -5.602    0.000    -6859.156    -3221.854
##         Po1      115.024        13.754        0.884      8.363    0.000       87.227      142.821
##        Ineq       67.653        13.936        0.698      4.855    0.000       39.488       95.818
##          Ed      196.471        44.754        0.568      4.390    0.000      106.019      286.923
##           M      105.020        33.299        0.341      3.154    0.003       37.719      172.320
##        Prob    -3801.836      1528.097       -0.224     -2.488    0.017    -6890.236     -713.436
##          U2       89.366        40.906        0.195      2.185    0.035        6.693      172.039
## -----------------------------------------------------------------------------------------------------
```

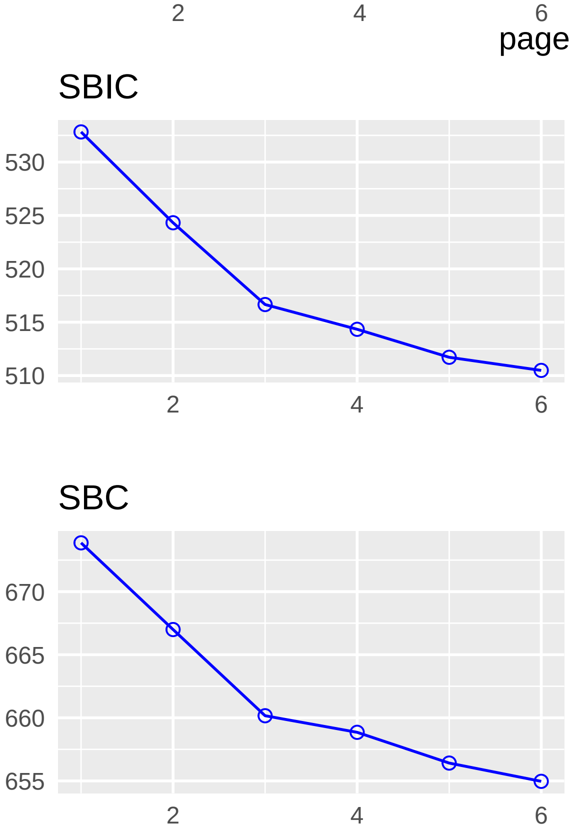
```
step
```

```
##
##                               Stepwise Selection Summary
## ------------------------------------------------------------------------------------------
##                      Added/                  Adj.
## Step    Variable    Removed     R-Square    R-Square     C(p)        AIC         RMSE
## ------------------------------------------------------------------------------------------
##    1       Po1       addition       0.473       0.461    39.9970    668.3155    283.9259
##    2      Ineq       addition       0.580       0.561    25.0710    659.5957    256.1874
##    3       Ed        addition       0.666       0.642    13.6390    650.9145    231.3136
##    4        M        addition       0.700       0.672    10.1620    647.7503    221.5397
##    5      Prob       addition       0.738       0.706     6.2580    643.4641    209.7205
##    6       U2        addition       0.766       0.731     3.8600    640.1661    200.6899
## ------------------------------------------------------------------------------------------
```

***Analysis*** *We can see above that each step is a concatenation of the previous model, starting at* Po1*, and the next feature. We can see at the bottom of the iterations, we ended up with a model that had 15 candidates for features but now is constructed by only 6 of them. The stepwise selection summary shows that with each feature addition, RMSE, AIC, and C(p) were dropping, while R-squared and Adjusted R-squared are increasing. This is the effect we like to see. We can plot the model to illustrate how much these metrics changed with each addition of a new feature.

```
plot(step)
```




```r
#Unseen data given a previous HW assignment
new.data <- data.frame("M"= 14,
"So" = 0,
```

```
"Ed" = 10,
"Po1" = 12,
"Po2" = 15.5,
"LF" = .640,
"M.F" = 94,
"Pop" = 150,
"NW" = 1.1,
"U1" = .120,
"U2" = 3.6,
"Wealth" = 3200,
"Ineq" = 20.1,
"Prob" = .04,
"Time" = 39)
```

Let's use the best model from Stepwise Regression to predict on the unseen data that was presented in Week 5 for the Crime data. Predictions on previous models were in the range of 1100 to 1350. So we should expect to see a prediction from this model within that range.

```
best_step.model <- step$model
unseen_step <- new.data[,c('Po1', 'Ineq', 'Ed','M','Prob','U2')]
predict(best_step.model, unseen_step)
```

```
##        1
## 1304.245
```

The prediction is well within the expected range of Crime predictions.

# LASSO REGRESSION

Next, we will implement LASSO Regression, which can consequently drive feature coeffecicents to zero with a high enough hyperparameter of alpha (lambda). Let's see what features LASSO selects vs Stepwise

```
require(glmnet)
```

```
## Loading required package: glmnet

## Loading required package: Matrix

## Loading required package: foreach

## Loaded glmnet 2.0-16
```

```
require(caret)
```

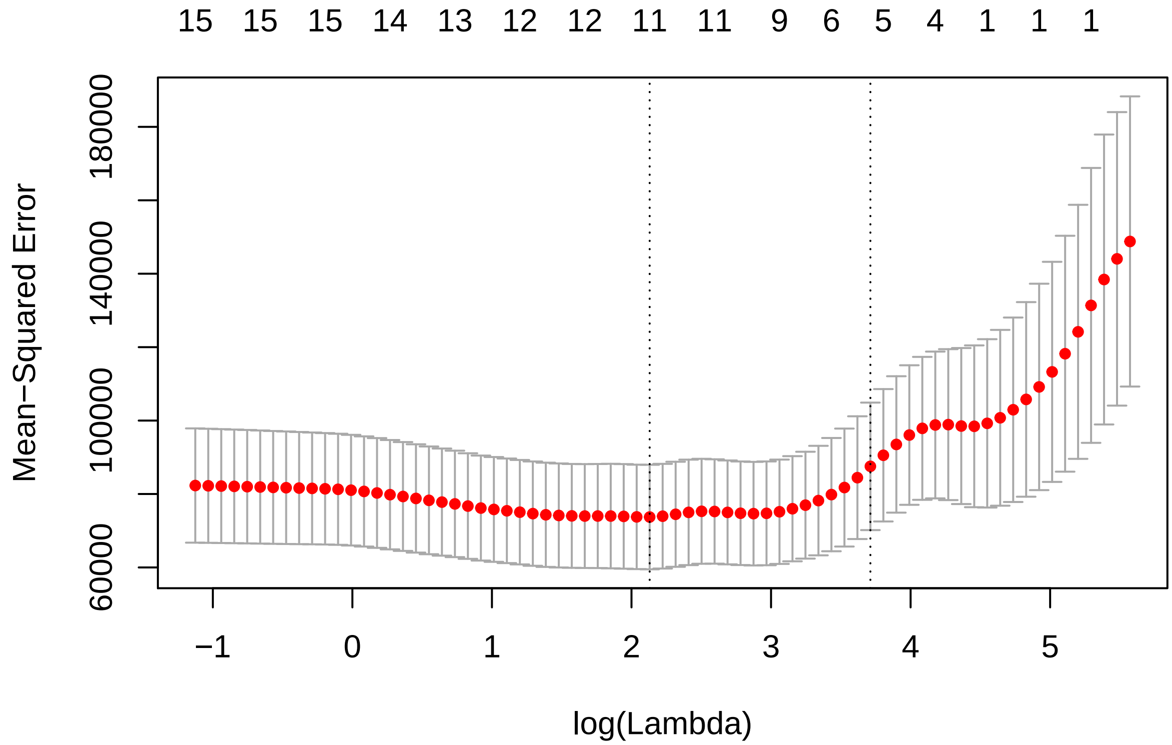
```
## Loading required package: caret

## Loading required package: lattice

## Loading required package: ggplot2
```

```
set.seed(42)

#Scale the train data
x <- as.data.frame(data[,1:15])
pp = preProcess(x)
scaled_x <- as.matrix(predict(pp, x))
#Scale the test data
scaled_test <- as.matrix(predict(pp, new.data))
```

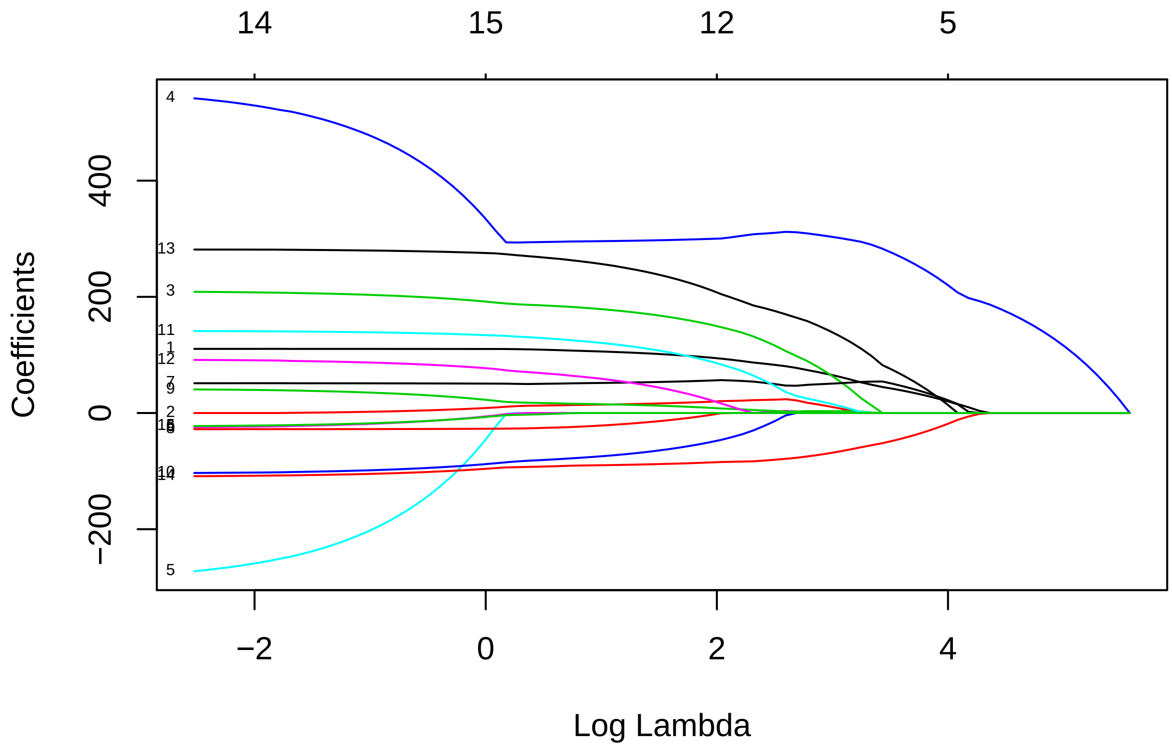
```r
#target variable
y <- data[,16]

lasso <- cv.glmnet(scaled_x,y, family = 'gaussian', alpha = 1, parallel = TRUE)
```

```
## Warning: executing %dopar% sequentially: no parallel backend registered
```

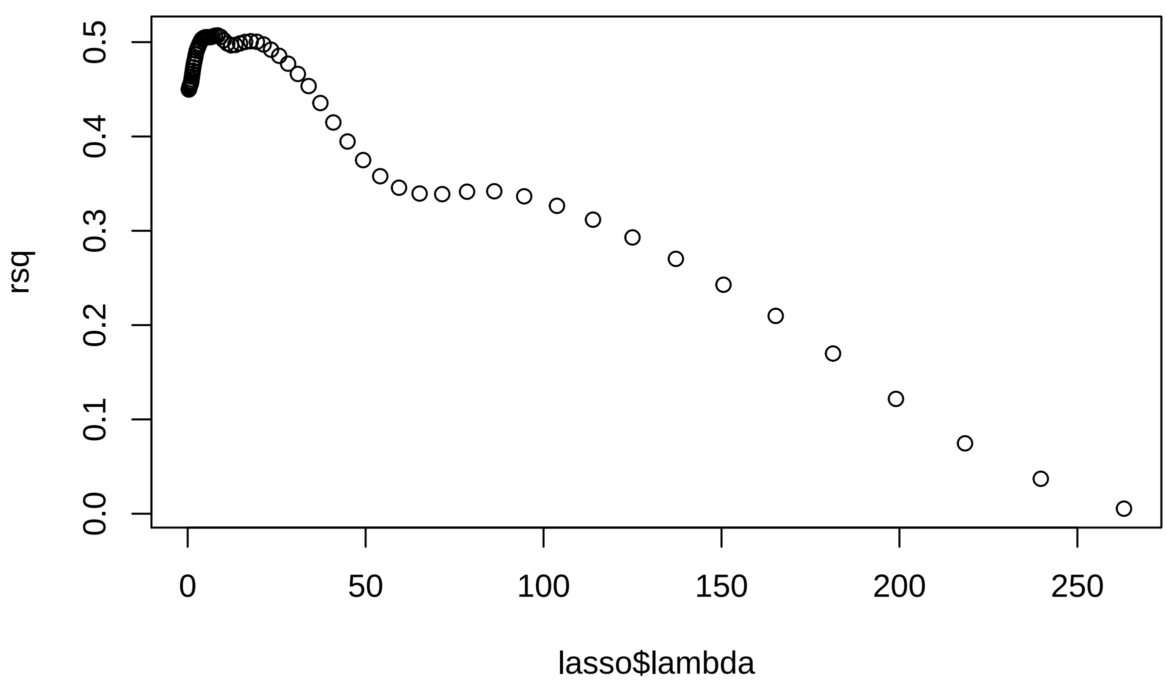
```r
plot(lasso)
```




```r
plot(lasso$glmnet.fit, xvar='lambda', label = TRUE)
```




```
rsq = 1 - lasso$cvm/var(y)
plot(lasso$lambda,rsq)
```




```
coeffecients <- coef(lasso, s=lasso$lambda.min)
coeffecients
```

```
## 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           .
```

```
paste('Minimum MSE: ',min(lasso$cvm))
```

```
## [1] "Minimum MSE:  73725.9799218528"
```

```
paste('Maximum R^2: ', max(rsq))
```

```
## [1] "Maximum R^2:  0.507131124727598"
```

Above in the first graph, we are basically witnessing the relationship between the regularization tuning parameter lambda, and Mean Squared Error. The two dotted lines is lambda min and lambda 1se (standard error) which is the best lamda that minimizes mean squared error and the lambda that is 1 standard error from lambda min. At the very top of the graph, the numbers represent how many non zero coeffecients there are. Therefore, the cross validated model apepars to have selected 11 non zero features and a lambda min of approx. 2.1(log) . The second graph shows how the coeffecients for all 15 features are adjusted as lambda is adjusted. We can see that as lambda increases, more and more features trend towards zero. Also, we can see that the most optimal model selected 11 of the initial 15 features where Po2, LF, Pop and Time are all zero. The minimum cross validated MSE (minimum square error) was 73725 for this model selection in comparison to stepwise's 40,276. Stepwise's R^2 was also better at 0.766, and the lasso model is a mere 0.51. Next Let's next predict Crime using this model and the new data. Again we are looking for values between 1000 and 1350

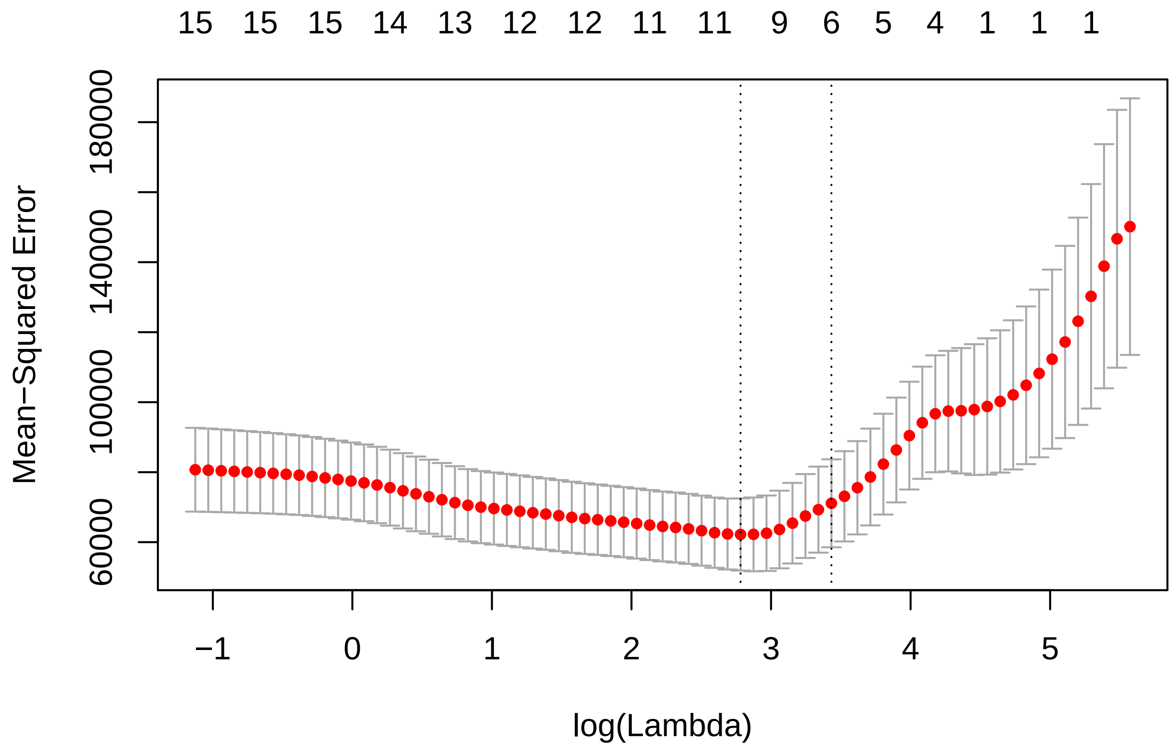
```
coeffecients[1] + coeffecients[2]*scaled_test[1] + coeffecients[3]*scaled_test[2] + coeffecients[4]*sca
```

```
## [1] 1097.259
```

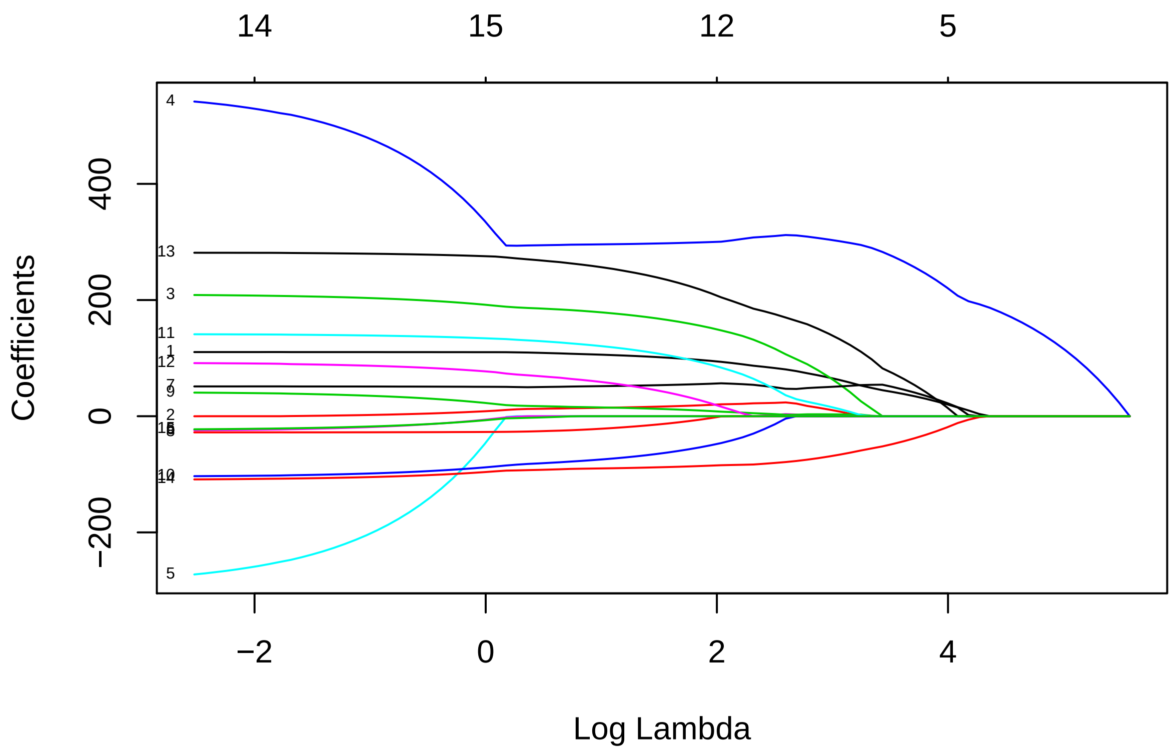
This model predict 1097, in comparison to the 1304 the Stepwise Model predicted. Next we will use an elastic net model which uses a combination of LASSO and RIDGE regression

## Elastic Net

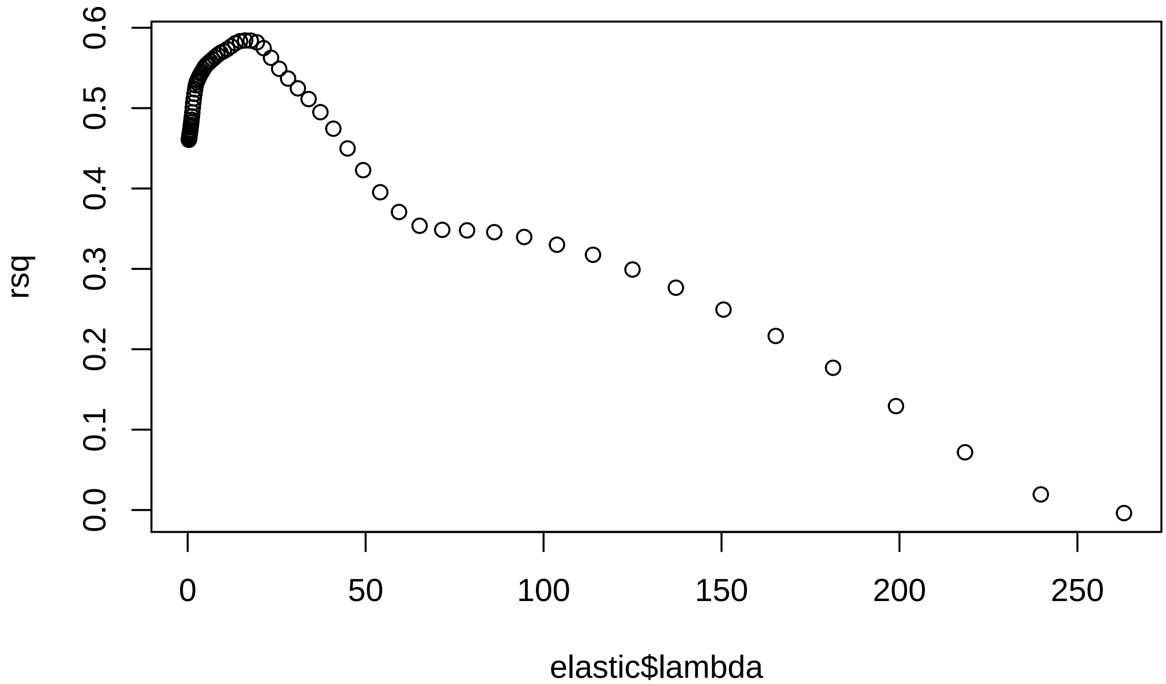
```
set.seed(24)
elastic <- cv.glmnet(x = scaled_x, y = y, family = 'gaussian')
plot(elastic)
```




```
plot(elastic$glmnet.fit, xvar='lambda', label = TRUE)
```




```
coeffecients <- coef(elastic, s = 'lambda.min')
rsq = 1 - elastic$cvm/var(y)
plot(elastic$lambda,rsq)
```




```r
coeffecients
```

```
## 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
paste('Minimum MSE: ',min(elastic$cvm))
```

```
## [1] "Minimum MSE:  62201.8759075783"
```

```r
paste('Maximum R^2: ', max(rsq))
```

```
## [1] "Maximum R^2:  0.584171432500491"
```

The elatic net model improved upon the LASSO model slightly with a MSE score of 62,201 and R^2 score of 0.58. This however is still not as good as the stepwise method. This can mean one of two things, either stepwise was a better method of model selection or it is overfitting the data. Finally, let's predict on the unseen data with the coeffecietns from elastic net

```
coeffecients[1] + coeffecients[2]*scaled_test[1] + coeffecients[3]*scaled_test[2] + coeffecients[4]*scal
```

```
## [1] 1204.24
```