ISYE 6051 : Homework 8
3/17/2021

Table of Contents

11.1a Stepwise Regression

Stepwise regression is designed to select the predictor variables that are significantly relevant to use with the response variable. In this case, as in the other problems, the Crime variable will be the response variable for this model.

3 versions of the stepwise regression function (forward, backward and both) were used to compare the predictors selected using the smallest value of AIC to select a model to prevent overfitting.

```
CODE: Step Forward

#library uploads
install.packages("MASS")

library(MASS)

#Read in Data
crimedf <- read.table("uscrime.txt", header = TRUE)
head(crimedf)

#Stepwise Regression
#Create a Linear Regression model with predictors
#Response (intercept-only) model
step_crime <- lm(Crime ~1, data = crimedf)

#All Predictors model
all_stepcrime <- lm(Crime~., data=crimedf)

step_forward <- step(step_crime, direction = "forward", scope=formula(all_stepcrime,
trace = 0))
step_forward$anova
step_forward$coefficients
```

OUTPUT:

Start: AIC=561.02
Crime ~ 1

	Df	Sum of Sq	RSS	AIC
+ Po1	1	3253302	3627626	532.94
+ Po2	1	3058626	3822302	535.39
+ Wealth	1	1340152	5540775	552.84
+ Prob	1	1257075	5623853	553.54
+ Pop	1	783660	6097267	557.34
+ Ed	1	717146	6163781	557.85
+ M.F	1	314867	6566061	560.82
<none>			6880928	561.02
+ LF	1	245446	6635482	561.32
+ Ineq	1	220530	6660397	561.49
+ U2	1	216354	6664573	561.52
+ Time	1	154545	6726383	561.96
+ So	1	56527	6824400	562.64
+ M	1	55084	6825844	562.65
+ U1	1	17533	6863395	562.90
+ NW	1	7312	6873615	562.97

Step: AIC=532.94
Crime ~ Po1

	Df	Sum of Sq	RSS	AIC
+ Ineq	1	739819	2887807	524.22
+ M	1	616741	3010885	526.18
+ M.F	1	250522	3377104	531.57
+ NW	1	232434	3395192	531.82
+ So	1	219098	3408528	532.01
+ Wealth	1	180872	3446754	532.53
<none>			3627626	532.94
+ Po2	1	146167	3481459	533.00
+ Prob	1	92278	3535348	533.72
+ LF	1	77479	3550147	533.92
+ Time	1	43185	3584441	534.37
+ U2	1	17848	3609778	534.70
+ Pop	1	5666	3621959	534.86
+ U1	1	2878	3624748	534.90
+ Ed	1	767	3626859	534.93

Step: AIC=524.22
Crime ~ Po1 + Ineq

```
              Df Sum of Sq      RSS     AIC
+ Ed           1    587050  2300757  515.53
+ M.F          1    454545  2433262  518.17
+ Prob         1    280690  2607117  521.41
+ LF           1    260571  2627236  521.77
+ Wealth       1    213937  2673871  522.60
+ M            1    181236  2706571  523.17
+ Pop          1    130377  2757430  524.04
<none>                      2887807  524.22
+ NW           1     36439  2851369  525.62
+ So           1     33738  2854069  525.66
+ Po2          1     30673  2857134  525.71
+ U1           1      2309  2885498  526.18
+ Time         1       497  2887310  526.21
+ U2           1       253  2887554  526.21

Step:  AIC=515.53
Crime ~ Po1 + Ineq + Ed

              Df Sum of Sq      RSS     AIC
+ M            1    239405  2061353  512.37
+ Prob         1    234981  2065776  512.47
+ M.F          1    117026  2183731  515.08
<none>                      2300757  515.53
+ Wealth       1     79540  2221218  515.88
+ U2           1     62112  2238646  516.25
+ Time         1     61770  2238987  516.26
+ Po2          1     42584  2258174  516.66
+ Pop          1     39319  2261438  516.72
+ U1           1      7365  2293392  517.38
+ LF           1      7254  2293503  517.39
+ NW           1      4210  2296547  517.45
+ So           1      4135  2296622  517.45

Step:  AIC=512.37
Crime ~ Po1 + Ineq + Ed + M

              Df Sum of Sq      RSS     AIC
+ Prob         1    258063  1803290  508.08
+ U2           1    200988  1860365  509.55
+ Wealth       1    163378  1897975  510.49
<none>                      2061353  512.37
+ M.F          1     74398  1986955  512.64
+ U1           1     50835  2010518  513.20
+ Po2          1     45392  2015961  513.32
+ Time         1     42746  2018607  513.39
```

```
+ NW          1     16488          2044865      513.99
+ Pop         1      8101          2053251      514.19
+ So          1      3189          2058164      514.30
+ LF          1      2988          2058365      514.30

Step:  AIC=508.08
Crime ~ Po1 + Ineq + Ed + M + Prob

              Df Sum of Sq         RSS          AIC
+ U2          1    192233          1611057      504.79
+ Wealth      1     86490          1716801      507.77
+ M.F         1     84509          1718781      507.83
<none>                             1803290      508.08
+ U1          1     52313          1750977      508.70
+ Pop         1     47719          1755571      508.82
+ Po2         1     37967          1765323      509.08
+ So          1     21971          1781320      509.51
+ Time        1     10194          1793096      509.82
+ LF          1       990          1802301      510.06
+ NW          1       797          1802493      510.06

Step:  AIC=504.79
Crime ~ Po1 + Ineq + Ed + M + Prob + U2

              Df Sum of Sq         RSS          AIC
<none>                             1611057      504.79
+ Wealth      1     59910          1551147      505.00
+ U1          1     54830          1556227      505.16
+ Pop         1     51320          1559737      505.26
+ M.F         1     30945          1580112      505.87
+ Po2         1     25017          1586040      506.05
+ So          1     17958          1593098      506.26
+ LF          1     13179          1597878      506.40
+ Time        1      7159          1603898      506.58
+ NW          1       359          1610698      506.78

> step_forward$anova

Step   Df            Deviance          Resid.        Df Resid. Dev       AIC
1      NA            NA                46            6880928            561.0235
2      + Po1         -1 3253301.8      45            3627626            532.9352
3      + Ineq        -1  739818.6      44            2887807            524.2154
4      + Ed          -1  587049.8      43            2300757            515.5343
5      + M           -1  239404.6      42            2061353            512.3701
6      + Prob        -1  258062.5      41            1803290            508.0839
7      + U2          -1  192233.4      40            1611057            504.7859
```

```
> step_forward$coefficients
(Intercept)        Po1        Ineq          Ed           M        Prob
-5040.50498   115.02419    67.65322   196.47120   105.01957 -3801.83628
         U2
   89.36604
```

The smallest AIC is reflected when the stepwise forward function traverses through all of the predictors (first fitting the intercept(response) only and then adding factors) and presents Po1, Ineq, Ed, M, Prob and U2 as the significant predictors to fit a model with the response variable of Crime.

Crime~-5040.5 115.02 67.65 196.5 105.02 -3801.84 (values rounded up)

```
CODE: Stepwise backward

step_backward <- step(step_crime, direction = "backward",
scope=formula(all_stepcrime, trace = 0))
step_backward$anova
step_backward$coefficients
```

```
OUTPUT:

>step_backward$anova

Step   Df          Deviance          Resid.      Df Resid. Dev      AIC
1       NA         NA                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

> step_backward$coefficients
(Intercept)           M          Ed         Po1         M.F          U1
-6426.10102    93.32155   180.12011   102.65316    22.33975 -6086.63315
         U2        Ineq        Prob
  187.34512    61.33494 -3796.03183
```

Unlike the Stepwise Forward functionality, Stepwise Backward has a smaller AIC (which is generally considered to be better to prevent overfitting) and more prediction variables are selected M, Ed, Po1, M.F. U1, U2, Ineq and Prob. The backward functionality first takes all of the predictors and traverses through them using a predictor -1 process until the lowest AIC is achieved.

Crime~-6426.1 93.32 180.12 102.65 22.34 -6086.63 187.35 61.33 -3796.03

```
CODE: STEPWISE BOTH

step_both <- step(step_crime, direction = "both", scope=formula(all_stepcrime), trace =
0)
step_both$anova
step_both$coefficients
```

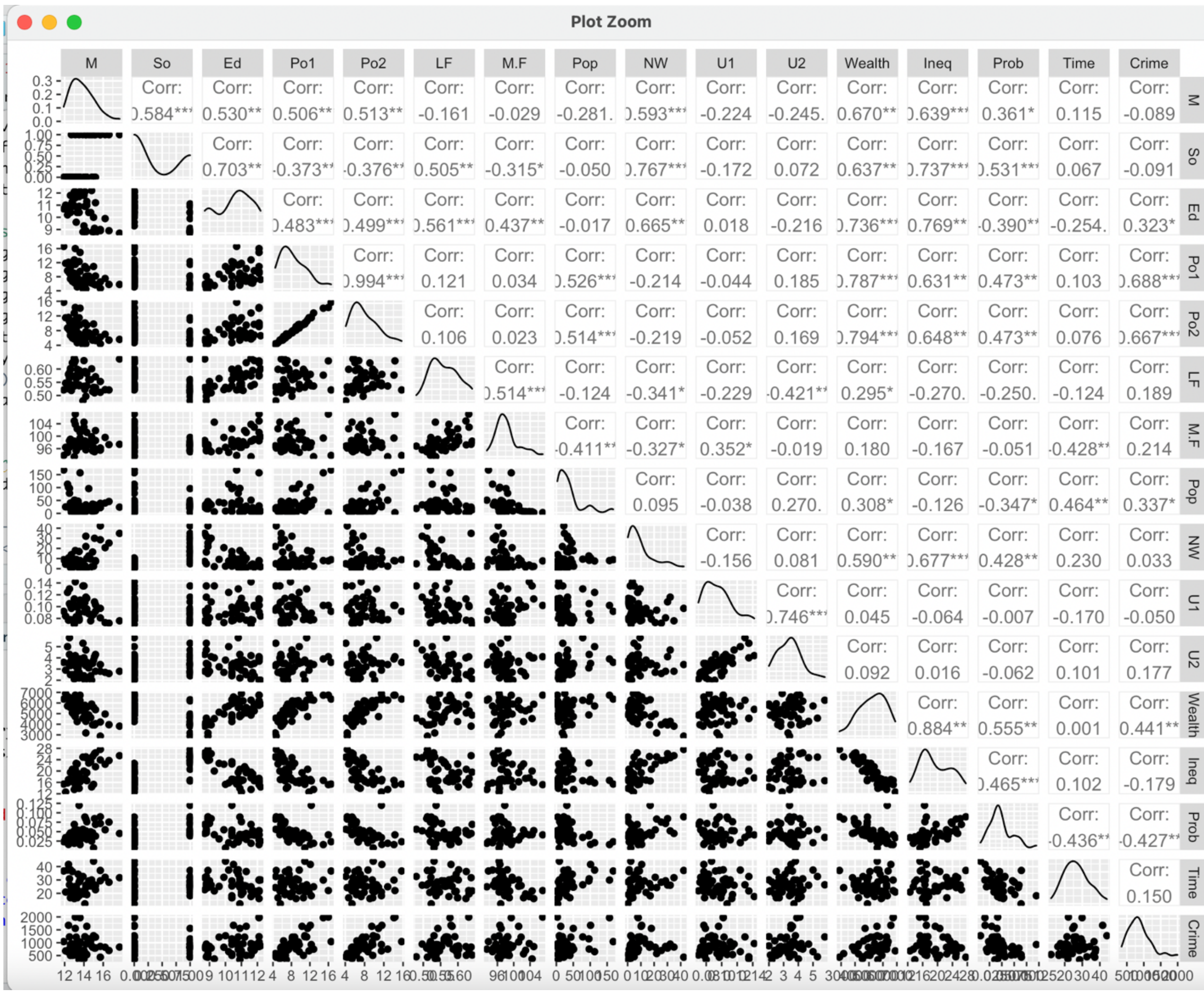
```
OUTPUT:

>step_both$anova

Step    Df      Deviance Resid.   Df Resid. Dev        AIC
1       NA      NA                46       6880928     561.0235
2       + Po1   -1 3253301.8      45       3627626     532.9352
3       + Ineq  -1  739818.6      44       2887807     524.2154
4       + Ed    -1  587049.8      43       2300757     515.5343
5       + M     -1  239404.6      42       2061353     512.3701
6       + Prob  -1  258062.5      41       1803290     508.0839
7       + U2    -1  192233.4      40       1611057     504.7859
> step_both$coefficients
(Intercept)        Po1        Ineq          Ed           M        Prob
-5040.50498   115.02419    67.65322   196.47120   105.01957 -3801.83628
         U2
   89.36604
```

Similar to the Stepwise Forward function, Stepwise Both finds an AIC of 504.7859. Similar to the forward function predictors are added that add significant value. An added step in process is that the both function also removes predictors that no longer add significant value. The final list of predictors is Po1, Ineq, Ed, M, Prob and U2. This is also similar to my findings in homework 8.2 where the p-values for these same 6 predictors showed the best fit.

Crime~-5040.50 115.02 67.65 196.47 105.02 -3801.84

11.1b LASSO

Similar to Stepwise Regression, LASSO creates a model that selects fewer predictors in order to avoid overfitting. This model will also focus on Crime as the response variable and finding the best set of prediction variables.
Using plotting there were 6 predictors that had a linear correlation value with the Crime response near 1 – Ineq, Wealth, U2, NW, Po2 and ED.



Running a simple linear regression using these predictors produced a model that showed the fit of the selected predictors. 2 of the 6 predictors chosen as part of the plotting correlation show significant association with Crime – M, Ed, Ineq and Prob.

```
CODE:

reg <- lm(Crime~., data=crimedf)
> summary(reg)
```

```
OUTPUT:
Call:
lm(formula = Crime ~ ., data = crimedf)

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
```

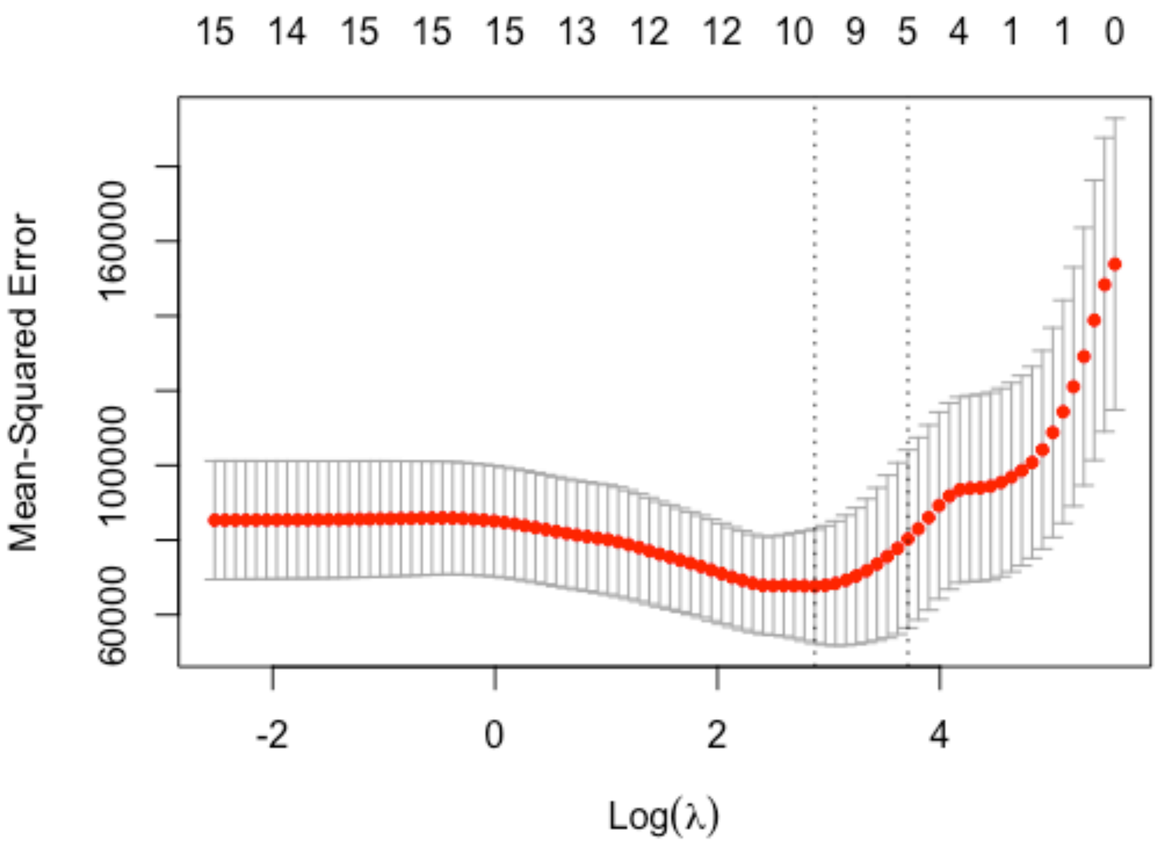
The glmnet function was now used to evaluate predictors using LASSO regression where alpha=1. For brevity, all of the output of Lambda values is not reflected for 1-99. Samples are displayed

```
CODE:

lasso_crime <- glmnet(scale(matrixcrime), dataresponse, family = "mgaussian",
alpha=1)
> lasso_crime
>cv.lasso_crime <- cv.glmnet(matrixcrime, dataresponse, alpha=1)
>plot(cv.lasso_crime, xvar = "lambda", label=T)

>abline(v=cv.lasso_crime$lambda.min, col = "red", lty=2)
>abline(v=cv.lasso_crime$lambda.1se, col="blue", lty=2)
>coef(cv.lasso_crime, s = cv.lasso_crime$lambda.min)

Call:  glmnet(x = scale(matrixcrime), y = dataresponse, family = "mgaussian",      alpha
= 1)
```

```
OUTPUT:

Df  %  Dev            Lambda
1      0  0.00        263.100
2      1  8.03        239.700
3      1 14.69        218.400
4      1 20.22        199.000
5      1 24.82        181.300
6      1 28.63        165.200
7      1 31.80        150.600
8      1 34.43        137.200
9      1 36.61        125.000
10     1 38.42        113.900
…
50     12 79.48       2.756
51     12 79.52       2.511
52     12 79.55       2.288
53     13 79.58       2.085
54     13 79.61       1.900
55     13 79.63       1.731
56     13 79.65       1.577
57     14 79.66       1.437
58     14 79.67       1.309
59     15 79.69       1.193
…
93     15 80.30       0.050
94     15 80.30       0.046
```

```
95       15 80.31        0.042
96       15 80.31        0.038
97       15 80.31        0.035
98       15 80.31        0.032
99       15 80.31        0.029
```

```
coef(cv.lasso_crime, s = cv.lasso_crime$lambda.min)
16 x 1 sparse Matrix of class "dgCMatrix"
                        1
(Intercept) -3828.8353017
M              56.1008808
So             30.7597658
Ed             70.8167194
Po1           103.2100909
Po2              .
LF               .
M.F            16.7898439
Pop              .
NW              0.3226147
U1               .
U2             24.9099830
Wealth           .
Ineq           37.7315902
Prob        -3179.3760049
Time             .
```



The plot shows Lambda between 3 and 4 therefore 4 significant predictors – M, Ed, Po1, and Ineq – were chosen to run a simple linear regression using scaled data from uscrime.txt.

```
OUTPUT:

Call:
lm(formula = Crime ~ M + Ed + Po1 + Ineq, data = scale_crime)

Residuals:
     Min       1Q   Median       3Q      Max
-1.37276 -0.23757  0.01956  0.35609  1.49145

Coefficients:
             Estimate    Std. Error   t value   Pr(>|t|)
(Intercept)  -1.506e-16  8.355e-02    0.000     1.000000
M            2.470e-01   1.119e-01    2.209     0.032714 *
Ed           4.803e-01   1.325e-01    3.626     0.000773 ***
Po1          9.974e-01   1.105e-01    9.029     2.16e-11 ***
```

```
Ineq            6.611e-01     1.575e-01     4.197           0.000137 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5728 on 42 degrees of freedom
Multiple R-squared:  0.7004,        Adjusted R-squared:  0.6719
F-statistic: 24.55 on 4 and 42 DF,  p-value: 1.595e-10
```

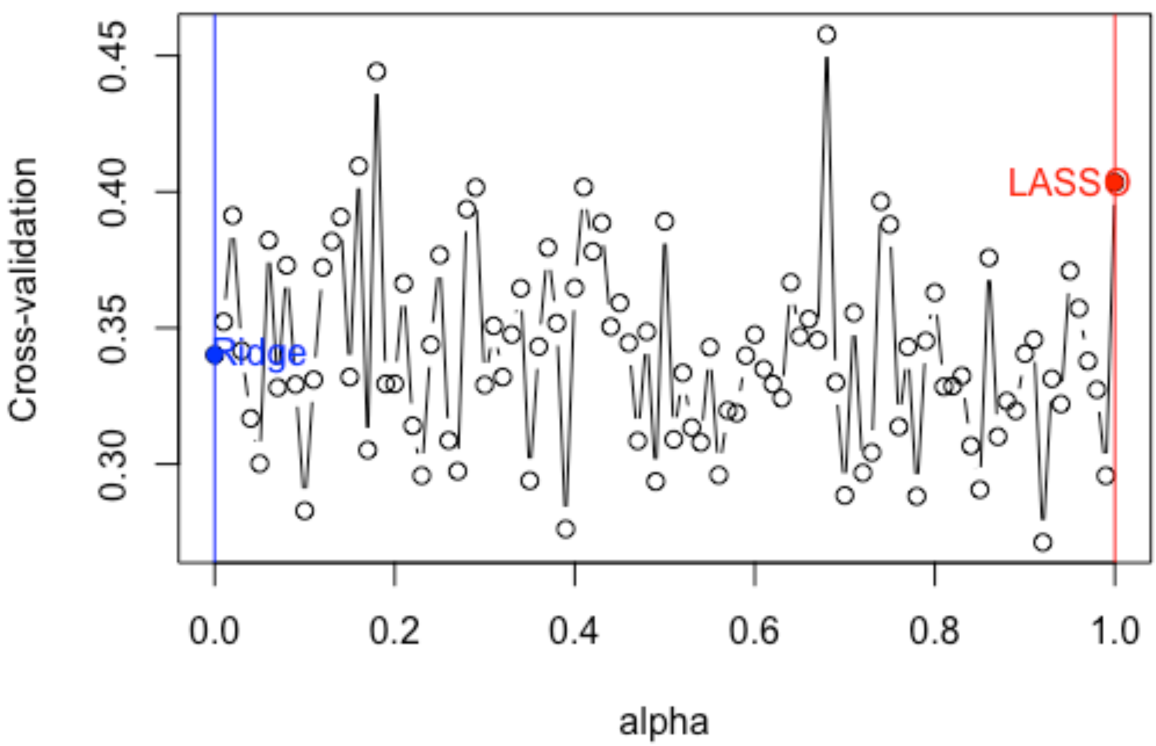
R^2 is 0.6719 so this is an adequate fit. Each of the variables could definitely be a predictor for crime rates – number of males, education, police expenditures and income inequality. However, running a model with 8 predictors -M, So, ED, Po1, M.F, NW, U2 and Ineq (adding the southern state indicator, number of men versus women, age and racial factors) - gave an R^2 of 0.6787 which is slightly better set of predictors for the documented crime rate.

11.1c Elastic net

Elastic net uses both LASSO regression (where predictor variable coefficients can go to 0) and Ridge regression (where predictor variable coefficients can go close to 0).

```
#11.1c Elastic net

set.seed(17)
crimedfScaled <- scale(crimedf)

#Define alpha from 0 to 1 in order to look at Ridge and LASSO
alpha_sequence <- seq(from = 0, to = 1, by = 0.01)
cv_error <- rep(NA, times = length(alpha_sequence))
for( i in 1:length(alpha_sequence)){
  cv_enet <- cv.glmnet( x = as.matrix(crimedfScaled[,c(Crime~., M + So + Ed + Po1 +
M.F + NW + U2 + Ineq)]),
                        y = as.matrix(crimedfScaled[,16]),
                        family = "gaussian", alpha = alpha_sequence[i],
                        nfolds = 5 )
  cv_error[i] <- min(cv_enet$cvm)
}
plot( alpha_sequence, cv_error,
      xlab = "alpha", ylab = "Cross-validation",
      type = "b")
abline(v = 0, col = "blue")
text( x = 0.05, y = cv_error[1], labels = "Ridge", col = "blue" )
points(x = 0, y = cv_error[1], col = "blue", pch = 16 )
abline(v = 1, col = "red")
```

```
text( x = 0.95, y = cv_error[length(cv_error)], labels = "LASSO", col = "red"
)
points(x = 1, y = cv_error[length(cv_error)], col = "red", pch = 16 )
```



An alpha value of ~ 0.7 provides the best fit with 8 predictors.