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.

For Parts 2 and 3, use the `glmnet` function in R.

Answer 11.1

Clear global environment, load and preview dataset

```
> rm(list=ls())
> df1 <- read.table("/Users/.../uscrime.txt", header = T, stringsAsFactors = F)
> head(df1, 1)
     M So  Ed Po1 Po2   LF M.F Pop   NW    U1  U2 Wealth Ineq
1 15.1  1 9.1 5.8 5.6 0.51  95  33 30.1 0.108 4.1   3940 26.1
      Prob    Time Crime
1 0.084602 26.2011   791
```

Perform stepwise regression on the full dataset using both AIC and BIC. The summary of AIC result suggests 8 predictors are used after stepwise selection, whereas the summary of BIC suggests only 6 predictors are used with the absence of variables 'M.F' and 'U1'. Given both models yield similar R^2 values, model with BIC criterion seems to be more preferable with fewer factors.

```
> model1 <- lm(Crime~., data = df1)
> model2 <- step(model1, trace = F)
> model3 <- step(model1, k = log(47), trace = F)
> coefficients(model2)
(Intercept)           M          Ed         Po1         M.F
-6426.10102    93.32155   180.12011   102.65316    22.33975
         U1          U2        Ineq        Prob
-6086.63315   187.34512    61.33494 -3796.03183
> coefficients(model3)
(Intercept)           M          Ed         Po1          U2
-5040.50498   105.01957   196.47120   115.02419    89.36604
       Ineq        Prob
   67.65322 -3801.83628
> summary(model2)

Call:
lm(formula = Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob,
    data = df1)

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

> summary(model3)

Call:
lm(formula = Crime ~ M + Ed + Po1 + U2 + Ineq + Prob, data = df1)

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
```

Scale dataset to use for Lasso and Elastic net.

```
> scaled <-
cbind(as.data.frame(scale(df1[,1])),as.data.frame(df1[,2]),as.data.frame(scale(df1[,c(3,4,5,6,7,8,9,10,11,12,13,14,15)])),as.data.f
rame(df1[,16]))
> colnames(scaled) = colnames(df1)
> head(scaled,1)
         M So       Ed        Po1        Po2        LF       M.F
1 0.988693  1 -1.30851 -0.9085105 -0.8666988 -1.266746 -1.120605
          Pop       NW       U1       U2    Wealth     Ineq     Prob
1 -0.09500679 1.943739 0.695106 0.831368 -1.361609 1.679364 1.649763
         Time Crime
1 -0.05599367   791
> library(glmnet)
```

Perform Lasso regression on scaled dataset with nfolds set to 5. Coefficients of the model4 result suggests variables 'Po2' 'LF' and 'Time' are insignificant, and thus the other 12 factors are used to fit model5.

```
> model4 <- cv.glmnet(x = as.matrix(scaled[,-16]), y = as.matrix(scaled$Crime), alpha = 1, nfolds = 5, type.measure = "mse",
family = "gaussian")
> summary(model4)
          Length Class  Mode
lambda     88     -none- numeric
cvm        88     -none- numeric
cvsd       88     -none- numeric
cvup       88     -none- numeric
cvlo       88     -none- numeric
nzero      88     -none- numeric
call        7     -none- call
name        1     -none- character
glmnet.fit 12     elnet  list
lambda.min  1     -none- numeric
lambda.1se  1     -none- numeric
> coefficients(model4, s = model4$lambda.min)
16 x 1 sparse Matrix of class "dgCMatrix"
                     1
(Intercept) 891.752783
M            96.902732
So           39.163699
Ed          156.006320
Po1         299.188384
Po2           .
LF            .
M.F          55.288423
Pop          -5.980508
NW            9.834701
U1          -53.668780
U2           93.430232
Wealth       27.692490
Ineq        218.418342
Prob        -85.783159
Time          .
```

Result from Lasso regression consists of 12 factors as opposed to stepwise AIC method with 8 factors Summary of the model5 suggests there are presence of variables whose p-value is well above .05 threshold, so model6 was built using significant variables in model5.

```
> model5 <- lm(Crime~M+So+Ed+Po1+M.F+Pop+NW+U1+U2+Wealth+Ineq+Prob, data = scaled)
> summary(model5)

Call:
lm(formula = Crime ~ M + So + Ed + Po1 + M.F + Pop + NW + U1 +
    U2 + Wealth + Ineq + Prob, data = scaled)

Residuals:
    Min      1Q  Median      3Q     Max
-434.18 -107.01   18.55  115.88  470.32

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   897.29      51.91  17.286  < 2e-16 ***
M             112.71      49.35   2.284  0.02876 *
```

```
So          22.89    125.35   0.183  0.85621
Ed         195.70     62.94   3.109  0.00378 **
Po1        293.18     64.99   4.511 7.32e-05 ***
M.F         48.92     48.12   1.017  0.31656
Pop        -33.25     45.63  -0.729  0.47113
NW          19.16     57.71   0.332  0.74195
U1         -89.76     65.68  -1.367  0.18069
U2         140.78     66.77   2.108  0.04245 *
Wealth      83.30     95.53   0.872  0.38932
Ineq       285.77     85.19   3.355  0.00196 **
Prob       -92.75     41.12  -2.255  0.03065 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 202.6 on 34 degrees of freedom
Multiple R-squared:  0.7971,  Adjusted R-squared:  0.7255
F-statistic: 11.13 on 12 and 34 DF,  p-value: 1.52e-08
```

Model6 is fitted using 6 factors. Summary of the model suggests all factors are significant with R^2 value similar to previous models. Notice that the result from model6 is identical to the result in stepwise BIC method.

```
> model6 <- lm(Crime~M+Ed+Po1+U2+Ineq+Prob, data = scaled)
> summary(model6)

Call:
lm(formula = Crime ~ M + Ed + Po1 + U2 + Ineq + Prob, data = scaled)

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)   905.09      29.27  30.918  < 2e-16 ***
M             131.98      41.85   3.154  0.00305 ** 
Ed            219.79      50.07   4.390 8.07e-05 ***
Po1           341.84      40.87   8.363 2.56e-10 ***
U2             75.47      34.55   2.185  0.03483 *  
Ineq          269.91      55.60   4.855 1.88e-05 ***
Prob          -86.44      34.74  -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
```

Perform Elastic net on scaled dataset.

```
> r2=c()
> for (i in 0:100) {
+     model7 = cv.glmnet(x = as.matrix(scaled[,-16]), y = as.matrix(scaled$Crime), alpha = i/100, nfolds = 5, type.measure = "mse", family = "gaussian")
+ m = which(model7$glmnet.fit$lambda == model7$lambda.min)
+ r2 = cbind(r2, model7$glmnet.fit$dev.ratio[m])
+ }
> r2
        [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
```

```
[1,] 0.7743273 0.7084704 0.7559849 0.7388612 0.7378399 0.7633095
          [,7]      [,8]      [,9]     [,10]     [,11]     [,12]
[1,] 0.7512885 0.7872388 0.721134 0.7616028 0.7781487 0.7701326
         [,13]     [,14]     [,15]     [,16]     [,17]     [,18]
[1,] 0.7631761 0.7655552 0.7642482 0.754669 0.7677372 0.7720571
        [,19]     [,20]     [,21]     [,22]     [,23]     [,24]
[1,] 0.770546 0.771746 0.8029381 0.7607435 0.7657714 0.7632132
         [,25]     [,26]     [,27]     [,28]     [,29]     [,30]
[1,] 0.7453186 0.7521779 0.7105853 0.7634324 0.7887367 0.7391919
         [,31]     [,32]     [,33]     [,34]     [,35]     [,36]
[1,] 0.7906996 0.7789005 0.7592952 0.6787222 0.7928193 0.7810653
         [,37]     [,38]     [,39]     [,40]     [,41]     [,42]
[1,] 0.7623499 0.7935228 0.7824468 0.7939137 0.7871829 0.7915803
         [,43]     [,44]     [,45]     [,46]     [,47]     [,48]
[1,] 0.7790571 0.7936687 0.7307044 0.7524547 0.788943 0.7536114
         [,49]     [,50]     [,51]     [,52]     [,53]     [,54]
[1,] 0.7333959 0.760309 0.6974056 0.7612808 0.7665051 0.7508373
         [,55]     [,56]     [,57]     [,58]     [,59]     [,60]
[1,] 0.7749513 0.7811466 0.7949608 0.7303755 0.7763342 0.7823801
         [,61]     [,62]     [,63]     [,64]     [,65]     [,66]
[1,] 0.7870555 0.7930278 0.748668 0.7874556 0.7932021 0.6810646
         [,67]     [,68]     [,69]     [,70]     [,71]     [,72]
[1,] 0.7184253 0.7906718 0.7358419 0.7908232 0.7882172 0.7793365
         [,73]     [,74]     [,75]     [,76]     [,77]     [,78]
[1,] 0.7920349 0.7884898 0.8030258 0.7954808 0.7771676 0.7701247
         [,79]     [,80]     [,81]     [,82]     [,83]     [,84]
[1,] 0.7774003 0.7775113 0.7890336 0.7939008 0.7613226 0.7416594
         [,85]     [,86]     [,87]     [,88]     [,89]     [,90]
[1,] 0.7559768 0.778112 0.749483 0.7071762 0.7669546 0.7784631
         [,91]     [,92]     [,93]     [,94]     [,95]     [,96]
[1,] 0.7896392 0.728912 0.7674452 0.646746 0.7936412 0.6967131
         [,97]     [,98]     [,99]    [,100]    [,101]
[1,] 0.7577722 0.7104765 0.7955081 0.7900626 0.7682853
> alpha = (which.max(r2)-1)/100
> alpha
[1] 0.74
```

Result of the Elastic net model yielded 12 predictors which is identical to Lasso (12) but opposed to stepwise (8). Noting that the summary of the model still shows factors whose p-value is greater than .05 threshold. After removing these factors, the result would be identical to the result in previous model6 and stepwise BIC. Both elastic net and Lasso model has R^2 value of .7971, whereas stepwise AIC model has R^2 value of .7888. Given that the R^2 value for three models are similar, and stepwise model uses fewer factors for prediction, therefore stepwise model seems to perform better in the given dataset.

```
> model8 <- cv.glmnet(x=as.matrix(scaled[,-16]), y=as.matrix(scaled$Crime), alpha = alpha, nfolds = 5, type.measure = "mse", family = "gaussian")
> coefficients(model8, s=model8$lambda.min)
16 x 1 sparse Matrix of class "dgCMatrix"
                    1
(Intercept) 892.17530
M            99.63993
So           37.92256
Ed          164.15544
Po1         292.56919
Po2           .
LF            .
M.F          56.02197
Pop         -11.77770
NW           14.97051
```

```
U1        -64.48262
U2        106.18145
Wealth     42.93804
Ineq      230.00920
Prob      -88.45140
Time       .
> model9 <- lm(Crime ~ M+So+Ed+Po1+M.F+Pop+NW+U1+U2+Wealth+Ineq+Prob, data = scaled)
> summary(model9)

Call:
lm(formula = Crime ~ M + So + Ed + Po1 + M.F + Pop + NW + U1 +
   U2 + Wealth + Ineq + Prob, data = scaled)

Residuals:
   Min     1Q  Median     3Q    Max
-434.18 -107.01   18.55  115.88  470.32

Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept)   897.29      51.91  17.286  < 2e-16 ***
M           112.71      49.35   2.284  0.02876 *
So           22.89     125.35   0.183  0.85621
Ed          195.70      62.94   3.109  0.00378 **
Po1         293.18      64.99   4.511 7.32e-05 ***
M.F          48.92      48.12   1.017  0.31656
Pop         -33.25      45.63  -0.729  0.47113
NW           19.16      57.71   0.332  0.74195
U1          -89.76      65.68  -1.367  0.18069
U2          140.78      66.77   2.108  0.04245 *
Wealth       83.30      95.53   0.872  0.38932
Ineq        285.77      85.19   3.355  0.00196 **
Prob        -92.75      41.12  -2.255  0.03065 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 202.6 on 34 degrees of freedom
Multiple R-squared:  0.7971,  Adjusted R-squared:  0.7255
F-statistic: 11.13 on 12 and 34 DF,  p-value: 1.52e-08
```