ISyE 6501-HOMEWORK 8

Group Members:
Jinyu, Huang | jhuang472@gatech.edu | GTID: 903522245
Chengqi, Huang | chengqihuang@gatech.edu | GTID: 903534690
Yuefan, Hu | yuefanhu@gatech.edu | GTID:903543027
Jingyu, Li | alanli@gatech.edu | GTID: 903520148

Qusetion 11.1

Stepwise regression

As what our group has explored in Homework 5, we choose AICc as the metric for us to choose model considering the small sample size. We used a modified version of *stepAIC* from library "MASS" called *stepAICc*(https://stat.ethz.ch/pipermail/r-help/2009-April/389888.html). We use the full model (including all predictors) as initial model.

```
> crime = read.table("uscrime.txt", header=TRUE)  # import data
> crime$So = as.factor(crime$So)
```

```
> library(MASS)
> full.model = lm(Crime~., data=crime)
> stepAICc(full.model, direction = "both", steps=2000)
```

```
Start:  AIC=671.13
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=666.16
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=661.87
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=657.87
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=654.18
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=650.88
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=647.99
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=645.43
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


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

Coefficients:
(Intercept)            M           Ed          Po1          M.F
   -6426.10        93.32       180.12       102.65        22.34
         U1           U2         Ineq         Prob
   -6086.63       187.35        61.33     -3796.03
```

The stepwise method showes that model Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob has the lowest AICc. In the model, the coefficients of M, ED, PO1, U2, Ineq and Prob are significant while U1's coefficient is marginally significant. The adjusted R-squared of the model is 0.744, which is relatively high.

```
> model_stepwise = lm(Crime~M+Ed+Po1+M.F+U1+U2+Ineq+Prob, data=crime)
> summary(model_stepwise)
```

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

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
```

Lasso

According to the document of *glmnet* function, the algorithm is to minimize the objective function: $errors + \lambda \sum \|\beta_i\|$, which is not exactly the same approach as what's discussed in the lecture. Instead of setting the lasso constraint $\sum \|\beta_i\| \leq \tau$, running *glmnet* function will generate different values of λ and then find β_i that can minimize the objective function $errors + \lambda \sum \|\beta_i\|$. As we can see from the figures below, when the value of λ becomes larger, the number of non-zero predictor coefficients decreases and the fraction of deviance explained decreases too. This makes sense because when λ gets larger, the influence of penalty term $\sum \|\beta_i\|$ in the objective function $errors + \lambda \sum \|\beta_i\|$ becomes greater. In order to minimize the objective, $\sum \|\beta_i\|$ will get closer to zero so that more coefficients of predictors will equal to zero and less predictors are included in the model. The mechanism of increasing λ is the same as decreasing τ in the constraint.

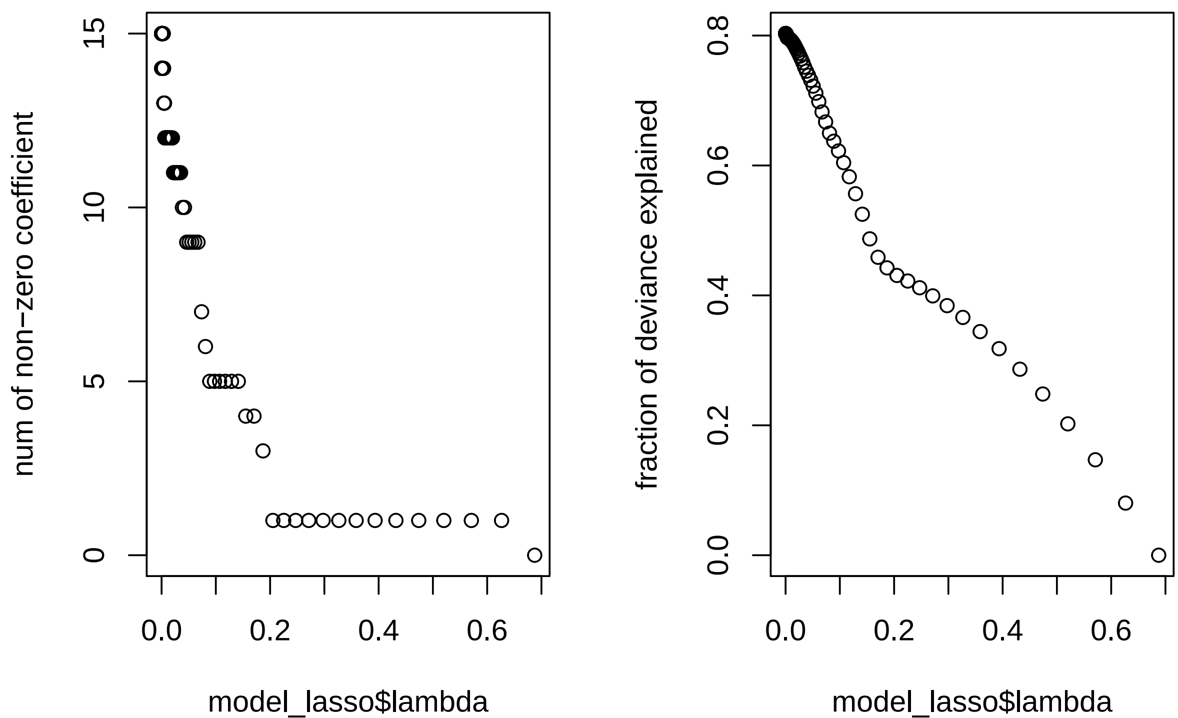
```
> library(glmnet)
```

```
Loading required package: Matrix

Loading required package: foreach

Loaded glmnet 2.0-18
```

```
> library(Matrix)
> library(foreach)
> model_lasso = glmnet(as.matrix(crime[,-16]), as.matrix(crime[,16]),
+                      family="mgaussian", alpha=1,
+                      standardize=TRUE, standardize.response=TRUE)
> par(mfrow=c(1,2))
> plot(model_lasso$lambda, model_lasso$df, ylab="num of non-zero coefficient")
> plot(model_lasso$lambda, model_lasso$dev.ratio, ylab="fraction of deviance explained")
```



Because there is no way for us to do validation or testing based on this small sample. The way helping us to select value of λ is to compare metrics like AICc and adjusted R-squared comprehensively. Considering the tradeoff between model complexity(number of predictors) and model fitting(fraction of deviance explained), we propose that we set $\lambda = 0.025, 0.05, 0.075, 0.1$ as example and the results showed below (if needed, we can traverse all the lambdas we get). Among these models, we tend to choose Crime~M+Ed+Po1+M.F+NW+Ineq+Prob.

```
> model_lasso_new = glmnet(as.matrix(crime[,-16]), as.matrix(crime[,16]),
+                         family="mgaussian", alpha=1, lambda=c(0.025,0.05,0.075,0.1),
+                         standardize=TRUE, standardize.response=TRUE)
> model_lasso_new$beta
```

```
15 x 4 sparse Matrix of class "dgCMatrix"
              s0           s1           s2            s3
M         29.55254  3.824356e+01   53.5211898  7.028494e+01
So         .        .              25.4476785  4.485523e+01
Ed         .        9.477320e+00   62.3025152  1.211070e+02
Po1       88.64104  9.710730e+01  102.4935761  1.030196e+02
Po2        .        .               .          .
LF         .        .               .          .
M.F       15.04436  1.832686e+01   17.0521736  1.858082e+01
Pop        .        .               .          .
NW         .        3.242865e-02    0.3155396  5.658636e-01
U1         .        .               .         -1.874393e+03
U2         .        .              20.8499293  8.156665e+01
Wealth     .        .               .          3.490447e-03
```

```
Ineq      15.33564  2.391696e+01    35.9800390  4.754845e+01
Prob   -1829.33474 -2.407847e+03 -3076.4573724 -3.665531e+03
Time         .           .                 .            .
```

```
> cat("lambda=0.1\n")
```

```
lambda=0.1
```

```
> model = lm(Crime~M+Po1+M.F+Ineq+Prob, data=crime)
> AICc(model)
```

```
[1] 651.01
```

```
> summary(model)$adj.r.squared
```

```
[1] 0.6752244
```

```
> cat("lambda=0.075\n")
```

```
lambda=0.075
```

```
> model = lm(Crime~M+Ed+Po1+M.F+NW+Ineq+Prob, data=crime)
> AICc(model)
```

```
[1] 649.789
```

```
> summary(model)$adj.r.squared
```

```
[1] 0.7071524
```

```
> cat("lambda=0.05\n")
```

```
lambda=0.05
```

```
> model = lm(Crime~M+So+Ed+Po1+M.F+NW+U2+Ineq+Prob, data=crime)
> AICc(model)
```

```
[1] 651.8072
```

```
> summary(model)$adj.r.squared
```

```
[1] 0.7204581
```

```
> cat("lambda=0.025\n")
```

```
lambda=0.025
```

```
> model = lm(Crime~M+So+Ed+Po1+M.F+NW+U1+U2+Wealth+Ineq+Prob, data=crime)
> AICc(model)
```

```
[1] 655.1874
```

```
> summary(model)$adj.r.squared
```

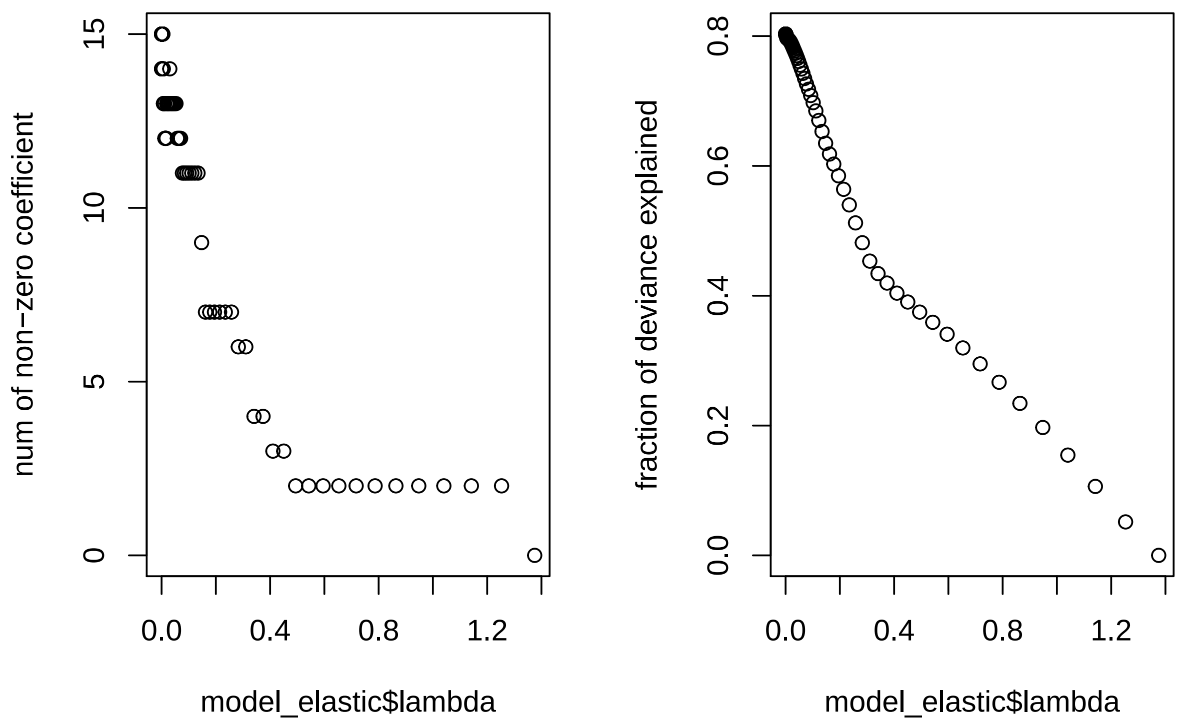
```
[1] 0.7292128
```

Elastic net

We set $\alpha = 0.5$ in *glmnet* function, so the algorithm is to minimize the objective function: $errors + 0.5\lambda(\sum \|\beta_i\| + \sum \beta_i^2)$. Same as Lasso, when the value of λ becomes larger, the number of non-zero predictor coefficients decreases and the fraction of deviance explained decreases too.

```
> model_elastic = glmnet(as.matrix(crime[,-16]), as.matrix(crime[,16]),
+                       family="mgaussian", alpha=0.5,
```

```
+                      standardize=TRUE, standardize.response=TRUE)
> par(mfrow=c(1,2))
> plot(model_elastic$lambda, model_elastic$df, ylab="num of non-zero coefficient")
> plot(model_elastic$lambda, model_elastic$dev.ratio, ylab="fraction of deviance explained")
```



Similarly, we set $\lambda = 0.05, 0.1, 0.15, 0.2$ as example and the results showed below (if needed, we can traverse all the lambdas we get). Among these models, we tend to choose Crime~M+So+Ed+Po1+Po2+M.F+NW+Ineq+Prob.

Another thing that is valuable to pay attention is that compare to Lasso, Elastic Net is not that sufficient to exclude some of the highly correlated predictors. As we know, Po1 and Po2 have correlation approximating to 1. In Lasso, only Po1 is included but here, the two variables are included.

```
> model_elastic_new = glmnet(as.matrix(crime[,-16]), as.matrix(crime[,16]),
+                      family="mgaussian", alpha=0.5, lambda=c(0.05,0.1,0.15,0.2),
+                      standardize=TRUE, standardize.response=TRUE)
> model_elastic_new$beta
```

```
15 x 4 sparse Matrix of class "dgCMatrix"
                 s0            s1          s2           s3
M         21.155227    30.0171507    45.74293  6.422112e+01
So         .            0.7152321    37.27261  5.749470e+01
Ed         .            3.8958268    44.22839  1.006003e+02
Po1       49.429737    55.9912810    63.52153  7.446260e+01
Po2       29.728811    32.6061047    32.57495  2.538479e+01
LF         .            .           184.51159  2.078840e+02
```

```
M.F       14.346459    18.5070480    18.13616  1.951890e+01
Pop         .            .             .        .
NW         1.439407     1.6337525     1.47710  1.291836e+00
U1          .            .             .      -1.617839e+03
U2          .            .           21.69812  7.521554e+01
Wealth      .            .             .        5.355787e-03
Ineq       8.876642    16.0317562    26.48426  4.034065e+01
Prob   -1921.571834 -2483.9855651 -3118.47146 -3.699059e+03
Time        .            .             .        .
```

```
> cat("lambda=0.2\n")
```

```
lambda=0.2
```

```
> model = lm(Crime~M+Po1+Po2+M.F+Ineq+Prob, data=crime)
> AICc(model)
```

```
[1] 653.8255
```

```
> summary(model)$adj.r.squared
```

```
[1] 0.667828
```

```
> cat("lambda=0.15\n")
```

```
lambda=0.15
```

```
> model = lm(Crime~M+So+Ed+Po1+Po2+M.F+NW+Ineq+Prob, data=crime)
> AICc(model)
```

```
[1] 654.4543
```

```
> summary(model)$adj.r.squared
```

```
[1] 0.7042624
```

```
> cat("lambda=0.1\n")
```

```
lambda=0.1
```

```
> model = lm(Crime~M+So+Ed+Po1+Po2+LF+M.F+NW+U2+Ineq+Prob, data=crime)
> AICc(model)
```

```
[1] 658.3654
```

```
> summary(model)$adj.r.squared
```

```
[1] 0.7102697
```

```
> cat("lambda=0.05\n")
```

```
lambda=0.05
```

```
> model = lm(Crime~M+So+Ed+Po1+Po2+LF+M.F+NW+U1+U2+Wealth+Ineq+Prob, data=crime)
> AICc(model)
```

```
[1] 662.6121
```

```
> summary(model)$adj.r.squared
```

```
[1] 0.7190197
```