ISYE 6501 Intro Analytics Modeling – HW8

Question 11.1 Using the crime data set `uscrime.txt` build a regression model

1. Stepwise regression

In the first step, I scaled all 15 independent variables, and split the whole dataset into training (0.7) and setting (0.3) set. Then I created a model using all 15 scaled independent variables and tried backward variable selection using the step function. At each step, the model will drop one variable which brings the biggest decrement for AIC, then calculated the F-Value also the P-value of the Old and New AICs. In this case, the variable selection was stopped when the AIC stop decreasing. In the end, I get a model with only 9 variables selected and AIC 417 smaller than the full model 425.5. Below is the detail of each step.

```
> model_step <- step(full_model,direction = "backward",test
= "F")
Start:  AIC=425.5
Crime ~ M + So + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 +
U2 + Wealth + Ineq + Prob + Time
         Df Deviance    AIC F value   Pr(>F)
- Wealth  1   556374 423.63  0.0619 0.806848
- Pop     1   557037 423.66  0.0799 0.781332
- M.F     1   557732 423.70  0.0987 0.757732
- So      1   565127 424.11  0.2989 0.592635
- LF      1   581312 424.98  0.7370 0.404123
<none>        554087 425.50
- NW      1   598536 425.89  1.2033 0.289960
- U1      1   600828 426.01  1.2654 0.278315
- U2      1   643275 428.12  2.4145 0.141060
- M       1   659600 428.90  2.8564 0.111680
- Ed      1   691913 430.38  3.7312 0.072523 .
- Ineq    1   699446 430.72  3.9351 0.065901 .
- Time    1   734036 432.22  4.8715 0.043300 *
- Prob    1   797152 434.77  6.5802 0.021539 *
- Po2     1   809589 435.25  6.9168 0.018933 *
- Po1     1   962983 440.63 11.0695 0.004596 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1
Step:  AIC=423.63
Crime ~ M + So + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 +
U2 + Ineq + Prob + Time
       Df Deviance    AIC F value  Pr(>F)
- Pop   1   557434 421.68  0.0305 0.86362
- M.F   1   564256 422.06  0.2267 0.64045
- So    1   576818 422.74  0.5879 0.45439
- LF    1   584014 423.13  0.7948 0.38585
<none>      556374 423.63
- NW    1   601174 424.03  1.2883 0.27307
- U1    1   635251 425.74  2.2683 0.15153
- M     1   665726 427.19  3.1447 0.09521 .
- U2    1   681758 427.93  3.6057 0.07577 .
- Ed    1   716903 429.48  4.6164 0.04732 *
- Ineq  1   748116 430.81  5.5140 0.03205 *
- Time  1   757026 431.17  5.7703 0.02880 *
- Po2   1   831952 434.10  7.9250 0.01245 *
- Prob  1   842912 434.50  8.2401 0.01110 *
- Po1   1  1006008 439.99 12.9304 0.00242 **
---
Step:  AIC=421.68
Crime ~ M + So + Ed + Po1 + Po2 + LF + M.F + NW + U1 + U2 +
Ineq + Prob + Time
       Df Deviance    AIC F value   Pr(>F)
- M.F   1   573880 420.59  0.5016 0.488411
- So    1   585629 421.21  0.8599 0.366753
- LF    1   590523 421.47  1.0091 0.329202
<none>      557434 421.68
- NW    1   608074 422.38  1.5444 0.230840
- U1    1   668150 425.30  3.3765 0.083676 .
- M     1   672179 425.49  3.4994 0.078709 .
- U2    1   685261 426.08  3.8983 0.064808 .
- Ed    1   717459 427.51  4.8803 0.041178 *
- Ineq  1   757538 429.19  6.1025 0.024382 *
- Time  1   794839 430.68  7.2401 0.015471 *
- Prob  1   844128 432.55  8.7433 0.008829 **
- Po2   1   849369 432.74  8.9031 0.008337 **
- Po1   1  1015687 438.28 13.9753 0.001636 **
---
Step:  AIC=420.59
Crime ~ M + So + Ed + Po1 + Po2 + LF + NW + U1 + U2 + Ineq +
Prob + Time
       Df Deviance    AIC F value   Pr(>F)
- LF    1   592923 419.60  0.5973 0.449646
- So    1   594711 419.69  0.6534 0.429468
<none>      573880 420.59
- NW    1   613006 420.63  1.2272 0.282535
- U1    1   671744 423.47  3.0695 0.096791 .
- U2    1   692493 424.41  3.7203 0.069675 .
- M     1   747873 426.80  5.4574 0.031253 *
- Ed    1   761621 427.36  5.8886 0.025965 *
- Ineq  1   807447 429.17  7.3259 0.014446 *
- Time  1   827936 429.95  7.9686 0.011270 *
- Prob  1   860359 431.14  8.9855 0.007726 **
- Po2   1   868609 431.43  9.2443 0.007038 **
- Po1   1  1044159 437.14 14.7505 0.001198 **
---
Step:  AIC=419.6
Crime ~ M + So + Ed + Po1 + Po2 + NW + U1 + U2 + Ineq + Prob
+ Time
       Df Deviance    AIC F value   Pr(>F)
- So    1   604080 418.18  0.3575 0.556944
- NW    1   617267 418.85  0.7801 0.388149
<none>      592923 419.60
- U1    1   674029 421.57  2.5990 0.123416
- U2    1   703499 422.90  3.5434 0.075191 .
- Ed    1   763342 425.43  5.4610 0.030548 *
- M     1   763836 425.45  5.4769 0.030335 *
- Ineq  1   812153 427.35  7.0252 0.015784 *
- Time  1   830603 428.05  7.6164 0.012468 *
- Prob  1   868749 429.44  8.8387 0.007815 **
- Po2   1   874694 429.65  9.0293 0.007283 **
- Po1   1  1055292 435.47 14.8164 0.001081 **
---
Step:  AIC=418.18
Crime ~ M + Ed + Po1 + Po2 + NW + U1 + U2 + Ineq + Prob +
Time
       Df Deviance    AIC F value    Pr(>F)
- NW    1   620359 417.00  0.5389 0.4713895
<none>      604080 418.18
- U1    1   674043 419.57  2.3163 0.1436767
- U2    1   704816 420.96  3.3352 0.0827797 .
- M     1   764848 423.49  5.3227 0.0318744 *
- Ed    1   786577 424.36  6.0421 0.0232162 *
- Ineq  1   812551 425.37  6.9021 0.0161475 *
- Time  1   832006 426.10  7.5462 0.0124271 *
- Po2   1   882811 427.94  9.2283 0.0064979 **
- Prob  1   891734 428.25  9.5237 0.0058262 **
- Po1   1  1061693 433.66 15.1507 0.0009047 ***
---
Step:  AIC=417
Crime ~ M + Ed + Po1 + Po2 + U1 + U2 + Ineq + Prob + Time
       Df Deviance    AIC F value    Pr(>F)
<none>      620359 417.00
- U1    1   683262 417.99  2.1294 0.1592990
- U2    1   729860 420.04  3.7068 0.0678426 .
- Ed    1   816859 423.53  6.6518 0.0174974 *
- Time  1   858817 425.08  8.0721 0.0097820 **
- M     1   881135 425.88  8.8276 0.0072864 **
- Po2   1   888000 426.12  9.0600 0.0066678 **
- Prob  1   912081 426.95  9.8752 0.0049166 **
- Ineq  1  1029886 430.71 13.8631 0.0012565 **
- Po1   1  1067954 431.84 15.1517 0.0008398 ***
```

Then I also tried the stepwise regression combined backward and forward selection. As you can see, at each step, not only the AIC after removing one variable is considered, the AIC after adding one variable is also considered. However, in this case, adding any variable in any step won't bring any AIC decrement. So, this method works the same as the backward selection in this case and I got the same variable set at the end.

```
> model_step <- step(full_model,direction = "both",test =
"F")
Start:  AIC=425.5
Crime ~ M + So + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 +
U2 + Wealth + Ineq + Prob + Time
         Df Deviance    AIC F value   Pr(>F)
- Wealth  1   556374 423.63  0.0619 0.806848
- Pop     1   557037 423.66  0.0799 0.781332
- M.F     1   557732 423.70  0.0987 0.757732
- So      1   565127 424.11  0.2989 0.592635
- LF      1   581312 424.98  0.7370 0.404123
<none>        554087 425.50
- NW      1   598536 425.89  1.2033 0.289960
- U1      1   600828 426.01  1.2654 0.278315
- U2      1   643275 428.12  2.4145 0.141060
- M       1   659600 428.90  2.8564 0.111680
- Ed      1   691913 430.38  3.7312 0.072523 .
- Ineq    1   699446 430.72  3.9351 0.065901 .
- Time    1   734036 432.22  4.8715 0.043300 *
- Prob    1   797152 434.77  6.5802 0.021539 *
- Po2     1   809589 435.25  6.9168 0.018933 *
- Po1     1   962983 440.63 11.0695 0.004596 **
---
Step:  AIC=423.63
Crime ~ M + So + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 +
U2 + Ineq + Prob + Time
         Df Deviance    AIC F value  Pr(>F)
- Pop     1   557434 421.68  0.0305 0.86362
- M.F     1   564256 422.06  0.2267 0.64045
- So      1   576818 422.74  0.5879 0.45439
- LF      1   584014 423.13  0.7948 0.38585
<none>        556374 423.63
- NW      1   601174 424.03  1.2883 0.27307
+ Wealth  1   554087 425.50  0.0619 0.80685
- U1      1   635251 425.74  2.2683 0.15153
- M       1   665726 427.19  3.1447 0.09521 .
- U2      1   681758 427.93  3.6057 0.07577 .
- Ed      1   716903 429.48  4.6164 0.04732 *
- Ineq    1   748116 430.81  5.5140 0.03205 *
- Time    1   757026 431.17  5.7703 0.02880 *
- Po2     1   831952 434.10  7.9250 0.01245 *
- Prob    1   842912 434.50  8.2401 0.01110 *
- Po1     1  1006008 439.99 12.9304 0.00242 **
---
Step:  AIC=421.68
Crime ~ M + So + Ed + Po1 + Po2 + LF + M.F + NW + U1 + U2 +
Ineq + Prob + Time
         Df Deviance    AIC F value   Pr(>F)
- M.F     1   573880 420.59  0.5016 0.488411
- So      1   585629 421.21  0.8599 0.366753
- LF      1   590523 421.47  1.0091 0.329202
<none>        557434 421.68
- NW      1   608074 422.38  1.5444 0.230840
+ Pop     1   556374 423.63  0.0305 0.863622
+ Wealth  1   557037 423.66  0.0114 0.916324
- U1      1   668150 425.30  3.3765 0.083676 .
- M       1   672179 425.49  3.4994 0.078709 .
- U2      1   685261 426.08  3.8983 0.064808 .
- Ed      1   717459 427.51  4.8803 0.041178 *
- Ineq    1   757538 429.19  6.1025 0.024382 *
- Time    1   794839 430.68  7.2401 0.015471 *
- Prob    1   844128 432.55  8.7433 0.008829 **
- Po2     1   849369 432.74  8.9031 0.008337 **
- Po1     1  1015687 438.28 13.9753 0.001636 **
---
Step:  AIC=420.59
Crime ~ M + So + Ed + Po1 + Po2 + LF + NW + U1 + U2 + Ineq +
Prob + Time
        Df Deviance    AIC F value   Pr(>F)
- LF      1   592923 419.60  0.5973 0.449646
- So      1   594711 419.69  0.6534 0.429468
<none>        573880 420.59
- NW      1   613006 420.63  1.2272 0.282535
+ M.F     1   557434 421.68  0.5016 0.488411
+ Pop     1   564256 422.06  0.2900 0.597230
+ Wealth  1   573078 422.54  0.0238 0.879246
- U1      1   671744 423.47  3.0695 0.096791 .
- U2      1   692493 424.41  3.7203 0.069675 .
- M       1   747873 426.80  5.4574 0.031253 *
- Ed      1   761621 427.36  5.8886 0.025965 *
- Ineq    1   807447 429.17  7.3259 0.014446 *
- Time    1   827936 429.95  7.9686 0.011270 *
- Prob    1   860359 431.14  8.9855 0.007726 **
- Po2     1   868609 431.43  9.2443 0.007038 **
- Po1     1  1044159 437.14 14.7505 0.001198 **
---
Step:  AIC=419.6
Crime ~ M + So + Ed + Po1 + Po2 + NW + U1 + U2 + Ineq + Prob
    + Time
         Df Deviance    AIC F value   Pr(>F)
- So      1   604080 418.18  0.3575 0.556944
- NW      1   617267 418.85  0.7801 0.388149
<none>        592923 419.60
+ LF      1   573880 420.59  0.5973 0.449646
+ Pop     1   584017 421.13  0.2745 0.606732
+ M.F     1   590523 421.47  0.0732 0.789859
- U1      1   674029 421.57  2.5990 0.123416
+ Wealth  1   592922 421.60  0.0000 0.994678
- U2      1   703499 422.90  3.5434 0.075191 .
- Ed      1   763342 425.43  5.4610 0.030548 *
- M       1   763836 425.45  5.4769 0.030335 *
- Ineq    1   812153 427.35  7.0252 0.015784 *
- Time    1   830603 428.05  7.6164 0.012468 *
- Prob    1   868749 429.44  8.8387 0.007815 **
- Po2     1   874694 429.65  9.0293 0.007283 **
- Po1     1  1055292 435.47 14.8164 0.001081 **
---
Step:  AIC=418.18
Crime ~ M + Ed + Po1 + Po2 + NW + U1 + U2 + Ineq + Prob +
    Time
         Df Deviance    AIC F value    Pr(>F)
- NW      1   620359 417.00  0.5389 0.4713895
<none>        604080 418.18
+ Pop     1   590294 419.46  0.4437 0.5133345
- U1      1   674043 419.57  2.3163 0.1436767
+ So      1   592923 419.60  0.3575 0.5569439
+ LF      1   594711 419.69  0.2993 0.5906693
+ M.F     1   602090 420.07  0.0628 0.8047846
+ Wealth  1   603212 420.13  0.0273 0.8704117
- U2      1   704816 420.96  3.3352 0.0827797 .
- M       1   764848 423.49  5.3227 0.0318744 *
- Ed      1   786577 424.36  6.0421 0.0232162 *
- Ineq    1   812551 425.37  6.9021 0.0161475 *
- Time    1   832006 426.10  7.5462 0.0124271 *
- Po2     1   882811 427.94  9.2283 0.0064979 **
- Prob    1   891734 428.25  9.5237 0.0058262 **
- Po1     1  1061693 433.66 15.1507 0.0009047 ***
---
Step:  AIC=417
Crime ~ M + Ed + Po1 + Po2 + U1 + U2 + Ineq + Prob + Time
         Df Deviance    AIC F value    Pr(>F)
<none>        620359 417.00
- U1      1   683262 417.99  2.1294 0.1592990
+ NW      1   604080 418.18  0.5389 0.4713895
+ Pop     1   607281 418.34  0.4307 0.5191369
+ So      1   617267 418.85  0.1002 0.7549176
+ LF      1   617454 418.85  0.0941 0.7622308
+ M.F     1   618980 418.93  0.0445 0.8349719
+ Wealth  1   620304 419.00  0.0018 0.9668903
- U2      1   729860 420.04  3.7068 0.0678426 .
- Ed      1   816859 423.53  6.6518 0.0174974 *
- Time    1   858817 425.08  8.0721 0.0097820 **
- M       1   881135 425.88  8.8276 0.0072864 **
- Po2     1   888000 426.12  9.0600 0.0066678 **
- Prob    1   912081 426.95  9.8752 0.0049166 **
- Ineq    1  1029886 430.71 13.8631 0.0012565 **
- Po1     1  1067954 431.84 15.1517 0.0008398 ***
---
```

To test if variable decrement will improve the model, I compared MSE_Train and MSE_Test for the model using all 15 variables and the model only using those 9 variables. Although the model after the variable selection has a bigger training error, the test error is much smaller.

Model	MSE_TRAIN	MSE_TEST
Full model	17873.77	94588.46
Model after variable selection	20011.57	84837.4

2. Lasso

I tried to using LASSO regression with cross-validation on the same training set. I get MSE_Train 99789 and MSE_Test 60876.36, which is better than the backward selected model. From the coefficients, the model only chose 5 variables: M, Po1, LF, M.F, Ineq. We can see, Po2 is not selected in this model, which has a very high correlation with Crime from the single variable regression (hw5). The reason is that there is collinearity between Po1 and Po2, and LASSO can also deal with multicollinearity to reduce model noise.

```
Call:  cv.glmnet(x = as.matrix(train[, c(1:15)]), y = train$Crime, weights = NULL,      offset = NULL, lambda = NULL, type.measure
= c("default",          "mse", "deviance", "class", "auc", "mae", "C"), nfolds = 10,      foldid = NULL, alignment = c("lambda",
"fraction"), grouped = TRUE,      keep = FALSE, parallel = FALSE, gamma = c(0, 0.25, 0.5, 0.75,          1), relax = FALSE,
trace.it = 0, family = "gaussian",      alpha = 1)
Measure: Mean-Squared Error
    Lambda Measure    SE Nonzero
min  39.21   83053 17346       5
1se  75.20   99789 15889       4
```

```
> coef(lasso_glm, s = "lambda.min")
16 x 1 sparse Matrix of class "dgCMatrix"
                    1
(Intercept) 913.64927
M            34.50857
So            .
Ed            .
Po1         307.28274
Po2           .
LF           29.98917
M.F          55.52510
Pop           .
NW            .
U1            .
U2            .
Wealth        .
Ineq         66.55190
Prob          .
Time          .
```

3. Elastic net

Similarly, I also train the same cross-validation model on the elastic net model. I get MSE_Train 106337 and MSE_Test 66711.07, which performs slightly worse than LASSO. The variable selected are M, Po1, Po2, LF, M.F, Ineq, Prob. Two more variables are selected comparing to Lasso. Spuriously, Ridge model cross-validation using all the variables gave the smallest test error. It indicates we probably will lose some important information by reducing the amount of variable.

Model	Variable Selected	MSE_TRAIN	MSE_TEST
Lasso	5	99789	59147.18
Elastic Net	7	106337	69829.92
Ridge	15 (all)	101888	46963.7

```
Call:  cv.glmnet(x = as.matrix(train[, c(1:15)]), y = train$Crime, weights = NULL,      offset = NULL, lambda = NULL, type.measure
= c("default",          "mse", "deviance", "class", "auc", "mae", "C"), nfolds = 10,      foldid = NULL, alignment = c("lambda",
"fraction"), grouped = TRUE,      keep = FALSE, parallel = FALSE, gamma = c(0, 0.25, 0.5, 0.75,          1), relax = FALSE,
trace.it = 0, family = "gaussian",      alpha = 0.5)

Measure: Mean-Squared Error
```

```
    Lambda Measure    SE Nonzero
min   65.1   90291 17246       7
1se  137.0  106337 16114       5
```

```
> coef(elnet_glm, s = "lambda.min")
16 x 1 sparse Matrix of class "dgCMatrix"
                    1
(Intercept) 912.20016
M            34.31614
So            .
Ed            .
Po1         209.90233
Po2          79.13101
LF           32.30476
M.F          58.28796
Pop           .
NW            .
U1            .
U2            .
Wealth        .
Ineq         70.10120
Prob        -22.14290
Time          .
```

Code

```
# ISYE 6501 Intro Analytics Modeling - HW8
# IP uscrime.txt

# Loading and examining data
df<-read.delim("uscrime.txt", header = TRUE, sep = "\t")
fn <- function(x) scale(x,  scale = TRUE)
df_scaled<-as.data.frame(lapply(df[,-16], fn))
df_scaled$Crime<-df$Crime

#splite train and test
set.seed(666)
g <- sample(1:2,size=nrow(df_scaled),replace=TRUE,prob=c(0.7,0.3))
train <- df_scaled[g==1,]
test <- df_scaled[g==2,]

# Fit glm model: gaussian model backward stepwise&both
full_model<-glm(Crime~.,family = gaussian,train)
MSE_train_full<-mean(full_model$residuals^2) #MSE Train
full_test<-predict(full_model,test,type="response")
full_residials<-full_test-test$Crime
MSE_test_full<-mean(full_residials^2) #MSE Train

model_step_bw <- step(full_model,direction = "backward",test = "F")
model_step_both <- step(full_model,direction = "both",test = "F")

glm_model1<-glm(Crime~M + Ed + Po1 + Po2 + U1 + U2 + Ineq + Prob + Time,famil
y = gaussian,train)
MSE_train_m1<-mean(glm_model1$residuals^2) #MSE Train
confint(glm_model1) # 95% CI for the coefficients
p1_test<-predict(glm_model1,test,type="response")
p1_residials<-p1_test-test$Crime
MSE_test_m1<-mean(p1_residials^2) #MSE Train

# Fit glm model: lasso gaussian model 10-fold cross validation
```

```r
library(glmnet)
set.seed(123)
lasso_glm<-cv.glmnet(as.matrix(train[,c(1:15)]), train$Crime, family = "gauss
ian", alpha=1,
                     weights = NULL, offset = NULL, lambda = NULL,
                     type.measure = c("default", "mse", "deviance", "class",
"auc", "mae","C"),
                     nfolds = 10, foldid = NULL, alignment = c("lambda",  "fr
action"),
                     grouped = TRUE, keep = FALSE, parallel = FALSE,
                     gamma = c(0, 0.25, 0.5, 0.75, 1), relax = FALSE, trace.i
t = 0)
lasso_glm
plot(lasso_glm)

coef(lasso_glm, s = "lambda.min")
lasso_min_p<-predict(lasso_glm, as.matrix(test[,c(1:15)]), s = "lambda.min")
p2_residials<-lasso_min_p-test$Crime
MSE_test_m2<-mean(p2_residials^2) #MSE Train

coef(lasso_glm, s = "lambda.1se")
lasso_1se_p<-predict(lasso_glm, as.matrix(test[,c(1:15)]), s = "lambda.1se")
p3_residials<-lasso_1se_p-test$Crime
MSE_test_m3<-mean(p3_residials^2) #MSE Train

# Fit glm model: Elastic net model 10-fold cross validation
set.seed(123)
elnet_glm<-cv.glmnet(as.matrix(train[,c(1:15)]), train$Crime, family = "gauss
ian", alpha=0.5,
                     weights = NULL, offset = NULL, lambda = NULL,
                     type.measure = c("default", "mse", "deviance", "class",
"auc", "mae","C"),
                     nfolds = 10, foldid = NULL, alignment = c("lambda",  "fr
action"),
                     grouped = TRUE, keep = FALSE, parallel = FALSE,
                     gamma = c(0, 0.25, 0.5, 0.75, 1), relax = FALSE, trace.i
t = 0)
elnet_glm

coef(elnet_glm, s = "lambda.min")
elnet_min_p<-predict(elnet_glm, as.matrix(test[,c(1:15)]), s = "lambda.min")
p4_residials<-elnet_min_p-test$Crime
MSE_test_m4<-mean(p4_residials^2) #MSE Train

coef(elnet_glm, s = "lambda.1se")
elnet_1se_p<-predict(elnet_glm, as.matrix(test[,c(1:15)]), s = "lambda.1se")
p5_residials<-elnet_1se_p-test$Crime
MSE_test_m5<-mean(p5_residials^2) #MSE Train
```

```r
# Fit glm model: ridge gaussian model 10-fold cross validation

set.seed(123)
ridge_glm<-cv.glmnet(as.matrix(train[,c(1:15)]), train$Crime, family = "gauss
ian", alpha=0,
                     weights = NULL, offset = NULL, lambda = NULL,
                     type.measure = c("default", "mse", "deviance", "class",
"auc", "mae","C"),
                     nfolds = 10, foldid = NULL, alignment = c("lambda",  "fr
action"),
                     grouped = TRUE, keep = FALSE, parallel = FALSE,
                     gamma = c(0, 0.25, 0.5, 0.75, 1), relax = FALSE, trace.i
t = 0)
ridge_glm

coef(ridge_glm, s = "lambda.min")
ridge_min_p<-predict(ridge_glm, as.matrix(test[,c(1:15)]), s = "lambda.min")
p6_residials<-ridge_min_p-test$Crime
MSE_test_m6<-mean(p6_residials^2) #MSE Train
```