ISYE 6501 HW8

March 2021

1 Question 11.1.1 Stepwise

To use Stepwise selection, I don't need to scale data, and I will use the Stepwise Regression to make the model selection. The original linear regression model uses all 15 variables and has an Adjusted R-squared value of 0.7078. Using Stepwise Regression selection and AIC to select models that performs best. The selected model contains variables "M Ed Po1 M.F U1 U2 Ineq Prob", and has an Adjusted R-squared value of 0.7444. Which is a big improvement from the original model.

1.1 Original model:

```
Call:
lm(formula = Crime ~ ., data = df)
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
---
Multiple R-squared:  0.8031,       Adjusted R-squared:  0.7078
```

1.2 Stepwise Regression Selected model:

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

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 *
---
Multiple R-squared:  0.7888,        Adjusted R-squared:  0.7444
```

1.3   My R code for Question 11.1.1 is:

```
library(MASS)
set.seed(416)

df <- read.table("uscrime.txt", sep = '\t', stringsAsFactors = FALSE, header = TRUE)
head(df)

#fit
fit <- lm(Crime~., data=df)
summary(fit)
# Stepwise regression model
AIC_fit <- stepAIC(fit, direction = "both", trace = FALSE)
summary(AIC_fit)
```

2   Question 11.1.2 LASSO

Using LASSO selection, I found that variables "M So Ed Po1 M.F Pop NW U1 U2 Wealth Ineq Prob" are recommended. The R-squared value is 0.7848662, which is better than model selected using Stepwise regression selection. The features selected by LASSO selection are different from the model selected by Stepwise regression selection.

2.1   LASSO selected model:

```
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           .
```

2.2 My R code for Question 11.1.2 is (continued from the 11.1.1 part):

```
# Scale data
scale1 <- scale(df[,1], center = TRUE, scale = TRUE)
scale2 <- scale(df[,3:15], center = TRUE, scale = TRUE)
scaledf<- cbind(scale1, df[,2], scale2, df[,16])
colnames(scaledf) <- c("M","So","Ed","Po1","Po2","LF","M.F","Pop","NW",
                        "U1","U2","Wealth","Ineq","Prob","Time","Crime")
head(scaledf)

# LASSO
library(glmnet)
X <- as.matrix(scaledf[, 1:15])
Y <- as.matrix(scaledf[,16])
LASSO <- cv.glmnet(x=X,y=scaledf[,16],nfolds=5,alpha=1, family = 'gaussian', type.measure='mse')
c<-coef(LASSO,s='lambda.min',exact=TRUE)
c
LASSO$glmnet.fit$dev.ratio[which(LASSO$glmnet.fit$lambda == LASSO$lambda.min)]
```

3 Question 11.1.3 Elastic Net

To use Elastic Net approach, R has a built in function to find the best Alpha Lambda, the result I found are 0.9 & 3.461984 respectively. And the coefficients kept are listed below. R-squared value is 0.7950193. This value is slightly better than the model selected by LASSO regression.

```
16 x 1 sparse Matrix of class "dgCMatrix"
                    1
(Intercept) 894.14821
M           104.41548
So           32.12715
Ed          175.32605
Po1         295.12270
Po2              .
LF               .
M.F          52.79654
Pop         -19.19797
NW           15.07595
U1          -72.07934
U2          117.20963
Wealth       54.96388
Ineq        250.13697
Prob        -89.55149
Time             .
```

3.1 My R code for Question 11.1.3 is:

```
# Elastic Net
library(caret)
Al = seq(0, 1, by = .01)
set.seed(416)
model <- train(Crime~., data = scaledf, method = "glmnet",
               trControl = trainControl("cv", number = 10),tuneLength = 10)
# Best tuning parameter
model$bestTune
```

```
coef(model$finalModel, model$bestTune$lambda)
predictions <- predict(model, X)
R2 <- cor(Y, predictions)^2
R2
```

4 Conclusion:

For this data set, it's hard to say that which model is absolutely the best since their R squared value are so close and consider the size of data is small.

While Elastic Net seems performed better than LASSO selection and Stepwise Regression. But model selected by Stepwise Regression is the simplest in all 3 models. Consider the small size of the data, I would suggest to test the 3 models in larger data set and make better decision.