ISYE 6501, Week 8 HW

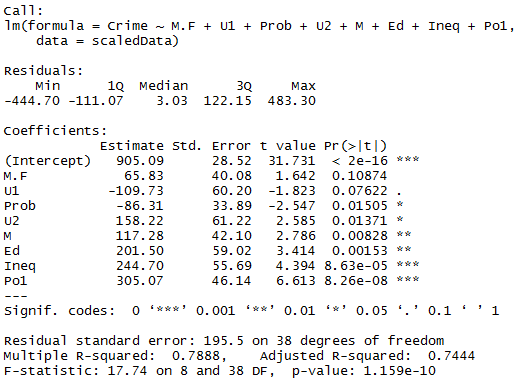
**Question 1**

Using the crime data set from Homework 5, 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.

**Response –**

1. Stepwise Regression

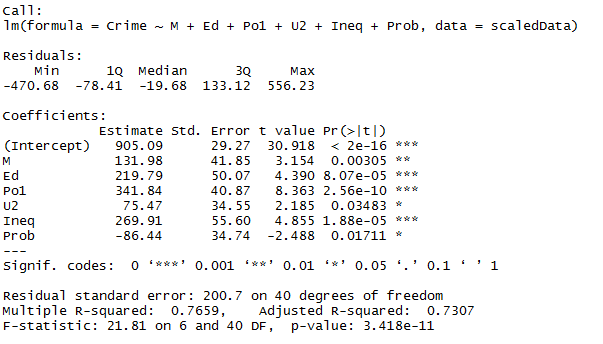
On performing stepwise regression using step function, the output model is as follows-



The R2 value for the model is 0.667

Observing the p values,  the predictors M.F and U1 have higher p values. Linear regression is re-run removing these 2 predictors.

The model output is -



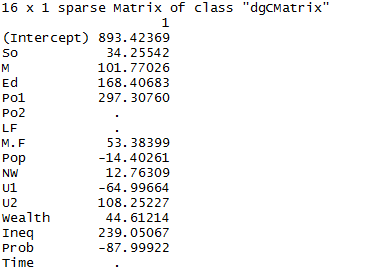
The R2 value for the model is 0.666 which is slightly lower than the earlier model. Since simpler models are preferred when there is no much difference in R2 value, the regression equation is

Crime = 905.09 + 131.98 \*M + 219.79\*Ed + 341.84\*Po1 + 75.47\*U2 + 269.41\*Ineq – 86.44\*Prob

1. Lasso Regression

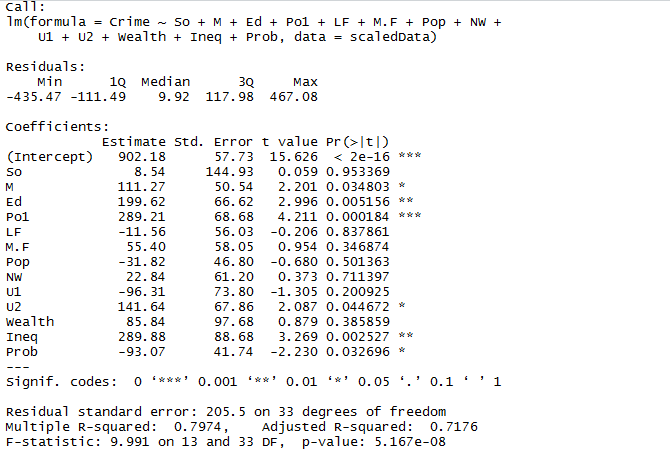
Lasso regression is performed using cv.glmnet function available in the glmnet library with alpha=1

This function gave away the minimum lambda value. Using the minimum lambda value, the significant predictors are evaluated as follows –



A linear regression model is built using all the predictors except Po2, LF and Time from the above result.

The regression model output is as follows –

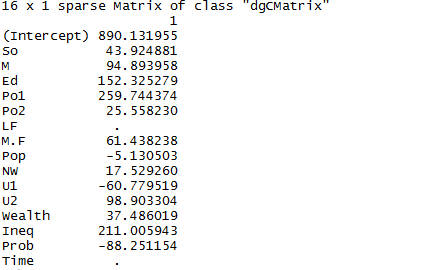


The R2 value for the model is 0.55. The reduction in R2 value can be attributed to the overfitting caused due to using 12 predictors, some of which are insignificant predictors as seen in the summary above. Using p-values, the model can be further refined, and the output model will be the same as stepwise regression model once the insignificant predictors are not considered.

1. Elastic net regression

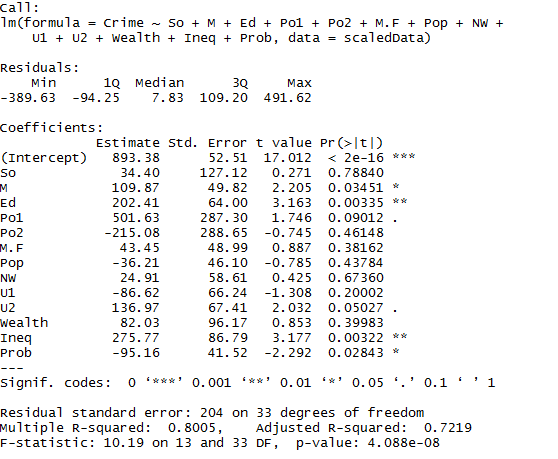
Elastic net regression is performed using cv.glmnet function available in the glmnet library with alpha=0.5

Using the min lambda calculated from the cv.glmnet output, elastic net regression significant coefficients are selected from the below output-



A linear regression model is built using all the predictors except LF and Time from the above result.

The regression model output is as follows –



The R2 value for the model is 0.53. The reduction in R2 value can be attributed to the overfitting caused due to using 13 predictors, some of which are insignificant predictors as seen in the summary above. Using p-values, the model can be further refined, and the output model will be the same as stepwise regression model once the insignificant predictors are not considered.

Overall, among the 3 regression models compared in this homework, stepwise regression provides a model with better R2 value. This may be due to the amount of data used to build the regression models. The ratio of number of records to number of predictors is very low in the input data.