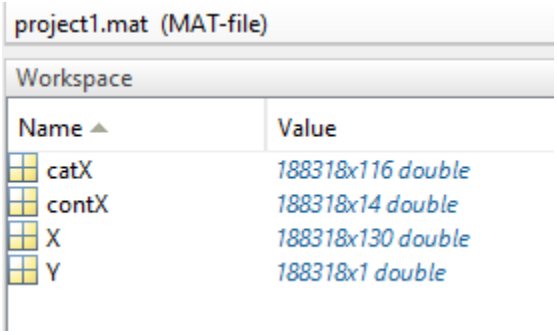
PREDICTION OF INSURANCE CLAIM SEVERITY LOSS USING REGRESSION MODELS

A. This framework predicts insurance claim severity loss using regression models. We ran four regression models for this study as stated below. This framework was implemented using MATLAB. All functions and libraries used are built in the MATLAB software package. Datasets and source code used for this study can be downloaded from the link below
http://github.com/ruthogunnaike/losspredictor

   i. Linear Regression
   ii. Support Vector Regression
   iii. Random Forest Regression
   iv. Feed Forward Neural Network

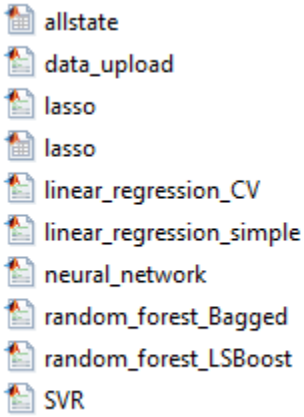
B. The source code folder contains the data set used for this study. Also contained in the folder are the source code files for each models used for this study. It is necessary to change certain values in the source codes to run each model. E.g. entering the kernel function to run on a support vector machine learning model or specifying the input for a neural network model.

   i. Allstate.mat – contains the response variable and the predictor variables. contX-is the continuous predictor variables, catX- is the categorical predictor variables. X – is the total predictor predictors and Y- is the target/response variable (loss)



   ii. Data_upload.m – is for uploading the data from excel files. The transformed data can be access in allstate.mat.

   iii. The matlab scripts are to run the models as depicted in the file names. Lasso.mat file is an already executed lasso regularization result which was done to avoid overfitting



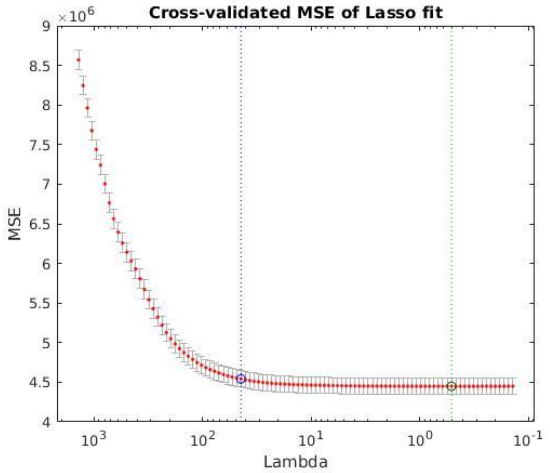
C. Follow the steps below to run each model.

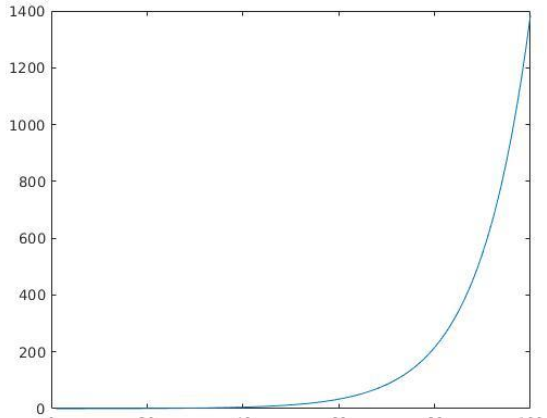
i. Upload the .mat files into MATLAB

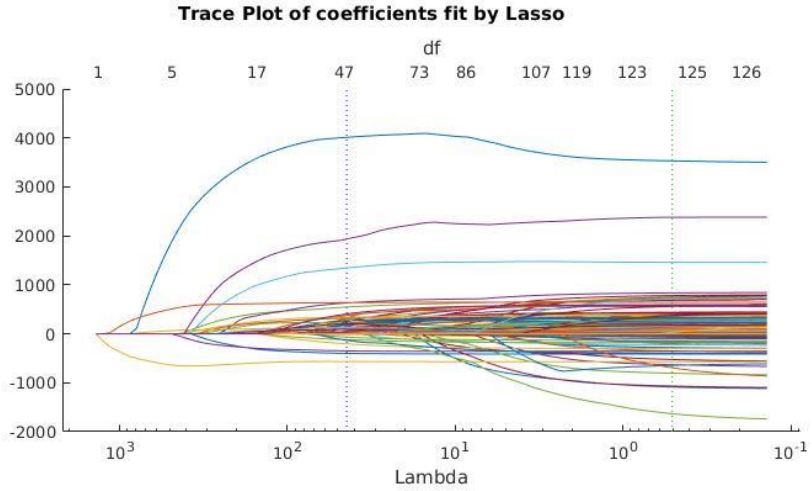
ii. Run the script for specific models.

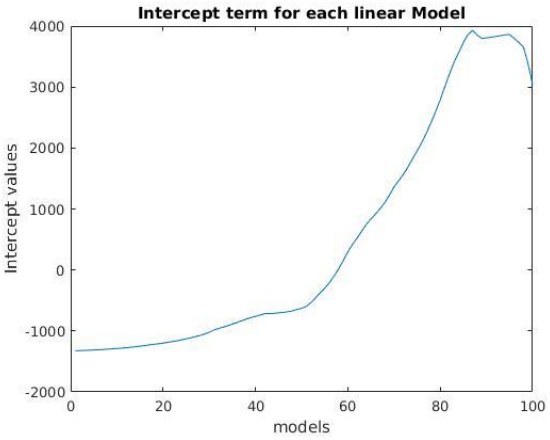
iii. Results for each model will be displayed on the matlab workspace.

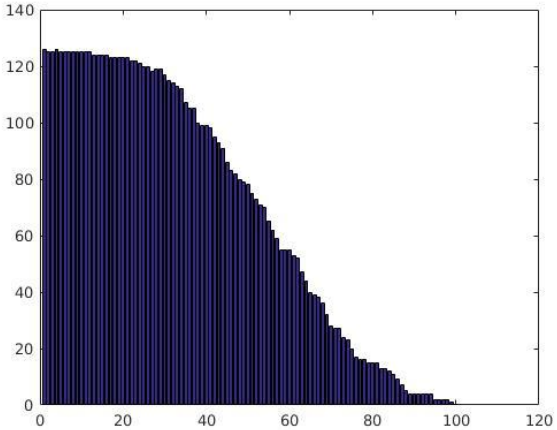
Sample Results for lasso

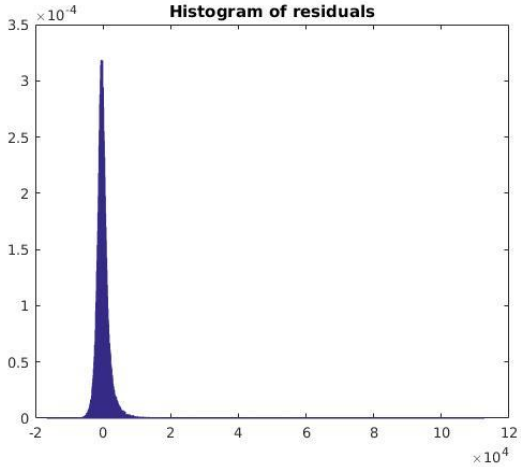


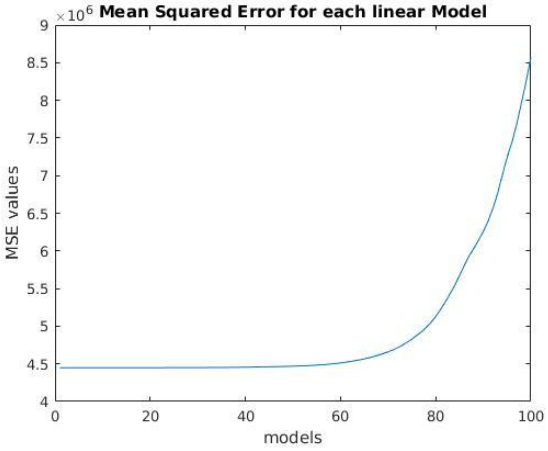





Intercept term for each linear Model
Intercept values
models
4000
3000
2000
1000
0
-1000
-2000
0
20
40
60
80
100


140
120
100
80
60
40
20
0
0
20
40
60
80
100
120


Histogram of residuals
×10^-4
3.5
3
2.5
2
1.5
1
0.5
0
-2
0
2
4
6
8
10
12
×10^4


×10^6 Mean Squared Error for each linear Model
MSE values
models
9
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
0
20
40
60
80
100

Samples for random_forest_bagged.m