Homework 2
IE 7275 Data Mining in Engineering

Task 1: Tutorial

- Practice R models presented in "R Code for Textbook Examples in Chap 6.pdf."
- Read R tutorial on "Regression." The data sets (women.R, mtcars.R and states.R) referenced in the tutorial are included

Problem 1: Concrete Slump Test Data [50 points]

Once you completed tutorial task, perform the following to complete the problem.

- Create a scatterplot matrix of "Concrete Slump Test Data" and select an initial set of predictor variables
- Build a few potential regression models using "Concrete Slump Test Data"
- Perform regression diagnostics using both typical approach and enhanced approach
- Identify unusual observations and take corrective measures
- Select the best regression model
- Fine tune the selection of predictor variables
- Interpret the prediction results

Problem 2: Forest Fire Data [50 points]

Once you completed tutorial task, perform the following to complete the problem.

- Create a scatterplot matrix of "Forest Fire Data" and select an initial set of predictor variables
- Build a few potential regression models using "Forest Fire Data"
- Perform regression diagnostics using both typical approach and enhanced approach
- Identify unusual observations and take corrective measures
- Select the best regression model
- Fine tune the selection of predictor variables
- Interpret the prediction results

Files Included in the Folder:

- Homework 4.docx
- R Code for Textbook Examples in Chap 6.pdf.
- R Tutorial on Regression.pdf
- Modeling Slump Flow Concrete.pdf
- Concrete Slump Test Data Description.pdf
- Concrete Slump Test Data.xlsx
- Data mining Approach to Predict Forest Fires.pdf
- Forest Fires Data Description.pdf
- Forest Fires Data.xlsx
- mtcars.R
- states.R
- women.R