

ILS 604 Big Data Analytics for Web and Text

Homework 1 (Due Feb 23, 2016 Tue midnight)

Data Set

Auto MPG <https://archive.ics.uci.edu/ml/datasets/Auto+MPG>

Data set and description available at <https://archive.ics.uci.edu/ml/machine-learning-databases/auto-mpg/>

Goal

Build regression models for predicting auto mpg (miles per gallon) from the independent variables. Please build regression models in R and report the following aspects.

To Report

1. What is your independent variable, what are your dependent variables given this analysis goal?
2. Describe the data by reporting means and standard deviation of each variable; plot pairs of variables (in a plot matrix) and report observations from the plot.
3. Build a linear regression model, and report its summary.

And what does the hypothesis testing (i.e. t-test results) tell you about the linear model coefficients?

What does R square of this model tell you?

Can you reduce any independent variables to obtain a better model?

Can you create some other variable(s) to enhance the regression model? (Tip: You could transform some existing variable to a different, more useful, form.)

Assignment due Tue Feb 23, 2016 midnight. Please submit through Canvas.

Can refer to the R code for regressions on Github Wiki

https://github.com/chenmiao/Big_Data_Analytics_Web_Text/wiki/Linear%20Regression%20with%20R