MSiA-400 Everything Starts with Data
Lab Exercise #3

Due Date: **Thursday, December 6, 12 pm**

EXERCISE INSTRUCTIONS: Please submit one report file that includes: short answer, related code and print for each problem if necessary. Push your answers to Github.

Problem 1

Short answers: Please answer the following questions in a few sentences.

Based on our labs,

a. What is the purpose of doing Markov Chain Monte Carlo (MCMC)?
b. What is the difference between the Metropolis Algorithm and the Metropolis Hastings Algorithm?
c. What is the purpose of Ridge regression? What is the purpose of LASSO regression?
d. State the IIA assumption for Multinomial Logit discrete choice model.

Problem 2

The gas_mileage.csv data set contains the response Mpg (Miles per gallon) and 11 predictors such as Displacement, Horsepower, Torque, etc. for 32 cars.

a. Fit quantile regression models for the 0.05, 0.10, 0.15, …, 0.90, 0.95th conditional quantiles for Mpg regressed on all the predictors using the `quantreg` package in `R`.
b. Plot the results using the `plot` function.
c. Interpret the results for 3 predictors of your choice.
d. Report the summary for the conditional median (0.50th conditional quantile) using the bootstrap method for computing standard errors of regression coefficients.

Problem 3

The car.csv data contains the response y = 0,1 (whether a family purchases a new car, yes = 1, no = 0) and 2 predictors: family income and age of car.

a. Fit a support vector machine to predict the response using default setting for kernels and hyper-parameters in the `svm` function in `e1071` package.
b. Plot the result using the `plot` function.
c. Predict the response for a family with income = 50, car age = 5.