Assignment 3
Submission deadline: Friday Oct 7, 2016 by 8 PM
Submission format: upload document in Canvas

1. **(10 points)** Load the **mtcars** dataset in R, and describe the dataset in your own words, in 2-3 lines.
   Feel free to use help(mtcars) or ?mtcars for this problem and subsequent problems.

2. **(50 points) Linear regression.** Consider **mpg** to be the response variable, and all other variables as features.

(a) Compute the correlation coefficient between **mpg** and all other features in the dataset. What are the two features most strongly correlated with **mpg**?
    (Hint: a strong correlation can be either positive or negative, use abs(x) to obtain the absolute value of a number x.)
(b) Fit two simple linear regression models: model 1 using the strongest feature from (a) and model 2 using the second strongest feature from (a). Report the linear regression formula (i.e., report the line equation) and the value of R^2 from the two models. If you had to choose between these two models, which one would you choose and why?
(c) Fit a multiple linear regression model with all features. Which features are significant in this model? What is the value of R^2 in this model?
(d) Using **stepAIC**, identify the best subset of features. Fit a multiple linear regression model using the best subset of features. Write down the regression formula and R^2 for this model. Are any of the features from (a) included in this model? Do they have the same coefficients as they had in model 1 or model 2 from (b)? If the coefficient values have changed, explain why.

3. **(40 points) Logistic regression.** Consider **am** to be the response variable, and all other variables as features.

(a) Describe the variable **am** in one sentence.
(b) Construct a plot of **hp** (x-axis) and **wt** (y-axis), with different colors for automatic and manual transmission. From the plot, do you think automatic and manual transmission can be distinguished by weight and horsepower?
(c) Fit a logistic regression model with **wt** as the only feature. Using this model, explain whether heavier cars are more likely or less likely to have manual transmission. If weight increases by 1000 lbs, what is the change in odds of a car having manual transmission?
(d) Fit a logistic regression model with **hp** as the only feature. Using this model, explain whether cars with higher horsepower are more likely or less likely to have manual transmission. If horsepower increases by 100, what is the change in odds of a car having manual transmission?
(e) If you had to choose between these two models, which one would you choose and why?

Assignment instructions:

1. **Honor code:** The Virginia Tech honor pledge for assignments is as follows:
   "I have neither given nor received unauthorized assistance on this assignment."

   The pledge is to be written out on all graded assignments at the university and signed by the student. Type up your name to sign.

2. Submit your assignment as a document (word, pdf or similar) to Canvas, clearly marked with student's name and assignment number, eg. Sengupta_Srijan_HW3.pdf. Your submission should include R code and answers to problems.

3. Late assignments will not be accepted. Check Canvas regularly for assignments and submission dates.

4. You are free to discuss assignment problems with your classmates, but submitted work (answers and codes) **must** be your own work. Students are not allowed to copy computer codes or answers from each other, and must write their own codes and answers.