

CLASS EXERCISE 1:

Instructions

The aim of this assignment is creating a model that is able to predict the miles per gallon (mpg) a car has based on other variables such as number of cylinders in the engine (cyl), displacement (disp, measure in cubic inches), gross horsepower (hp), rear axle ratio (drat), weight (wt, in 1000lbs), quarter mile time (qsec), v engine or straight engine (vs), transmission type (am, 0 = automatic and 1 = manual), number of forward gears (gear), number of carburetors (carb).

Follow through the 5-step process you went through in the Linear regression class notes to complete this assignment.

STEP 1: COLLECTING THE DATA

Get the data set from R called "mtcars", a default dataset stored in R, which is what you will use for this assignment. (HINT: Use the command `data(mtcars)` to get the data)

STEP 2: EXPLORING AND PREPARING THE DATA

1. Use the `attach()` function to be able to use the data collected.
2. Explore the structure of the data
3. Conduct summary statistics for the data
4. Run a plot for every predictor variable with our response variable, MPG.
5. Draw a histogram of the response variable
6. Conduct a correlation analysis of the variables
7. Visualize the relationship between the features using a scatterplot matrix function
8. Create an enhanced scatterplot matrix can be created with the `pairs.panels()` function

STEP 3: TRAINING THE MODEL

1. Fit a linear regression model on the data
2. Interpret the coefficients of the model.

STEP 4: EVALUATING MODEL PERFORMANCE

1. Evaluate the performance of the model by doing a `summary()` of the model. Comment on the results.

STEP 5: IMPROVING MODEL PERFORMANCE

1. Add nonlinear relationships to any 2 variables of your choosing
2. Conduct a binary transformation to any one variable of your choosing and explain why you have conducted that transformation.
3. Add an interaction effect and explain why you have chosen the variables to conduct the interactions effect on.
4. Train your model by taking into account these 3 changes in points 1 to 3.

5. Evaluate the performance of the model by doing the summary of the model. Comment on the results.

This should be submitted as either an R script, R Markdown or copy your code on Notebook and submit it. Any comments or information should be commented on your code appropriately. (Any explanation required of you from the questions SHOULD NOT be done on a separate document. Use the right format to include your explanation as part of the code)