ECE 595 Homework 2 Due: 5 PM, Feb. 21

Using the data file AutoData\_HW2.xlsx, perform a multivariate linear regression with four of the features, namely, No. of cylinders (column 2), displacement (column 3), horsepower (column 4) and weight (column 5) on the first 160 observations. (The remaining 40 observations will be used for cross validation – see below.) The target is the MPG (miles per gallon, column 1).

**a**. Because of the varying sizes of features, they need to be normalized individually before running your gradient descent algorithm for multivariate linear regression. Your code should perform the computation of the cost function and updating the hypothesis parameters in a vectorized form. Do NOT use *for* or *while* loops! You may stop the algorithm with a fixed number of iterations.

Cost function J, for example, can be vectorized as follows. Watch the sizes of the variables.

*ht = X\*theta; % Hypothesis: m x 1*

*J = (1/(2\*m))\*(ht - y)' \* (ht -y); % ht: m x 1; y: m x 1*

Parameter updating can similarly be carried out in vectorized form.

**b**. Use the last 40 observations as validation data to determine the predicted values for all the 40 observations - in a vectorized form - and obtain the mean squared error err\_cv between the observed (true) MPG and the predicted MPG for the validation data.

c. Using the pseudoinverse operation, obtain a closed form solution for the parameters theta for the training data.

You need to turn in the following.

1. Scatter plot of *y* vs. *x*, where *x* is still the weight (as in HW 1)
2. Plot of predicted MPG against observed MPG for the validation data
3. Plot of J vs. Iteration index for the training data
4. Minimum J and the hypothesis parameters from gradient descent code
5. Mean squared error err\_cv
6. Hypothesis parameters (theta) for the closed form
7. Your code for training using gradient descent, cross validation and closed form solution

Optional Exercise: You may use the fifth input variable, namely, acceleration (column 6), redo a 5-variable linear regression and see if it improves results for cross validation.