AMAT 592 Assignment 1

- This assignment is done by MATLAB. **Put all your code together in one executable `.m` file and submit through Blackboard.**
- **Due: March 22, 11:59 pm. Late homeworks will not be accepted.**

1. **Linear regression with least squares**

   The data set `linreg.mat` contains the feature vector $\mathbf{x}$ and label vector $\mathbf{y}$. You can load the data by command `load linreg.mat` in MATLAB. Perform linear regression to identify the optimal coefficients w_0^* and w_1^*. Recall that you need to set up a data matrix $\mathbf{X} \in \mathbb{R}^{n\times 2}$ that contains the features and the dummy variable 1:

   $$\mathbf{X} = \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ \vdots & \vdots \\ 1 & x^{(n)} \end{bmatrix}$$

   then solve the normal equation $(\mathbf{X}^\top \mathbf{X})\mathbf{w} = \mathbf{X}^\top \mathbf{y}$, where $\mathbf{w} = [w_0, w_1]^\top$.

   (a) Compute and print out the mean squared error $\frac{1}{n}\sum_{i=1}^n (y^{(i)} - w_0 - w_1 x^{(i)})^2$.

   (b) Create a figure to plot the data points $\{(x^{(i)}, y^{(i)})\}_{i=1}^n$ and the regression line. Label the axes.

2. **Robust linear regression** The data set `linreg+outlier.mat` contains the feature vector $\mathbf{x}$ and label vector $\mathbf{y}$, but one of the point is an outlier. In this case, we consider the robust linear regression or the so-called least absolute deviation problem

   $$\min_{w_0, w_1} \sum_{i=1}^n |y^{(i)} - w_0 - w_1 x^{(i)}|$$

   (a) Solve the above optimization problem for the given data set. You may code up any computational method introduced in class. Alternatively you may simply use the MATLAB subroutine `fminsearch` (use help for its usage). Create a figure to plot the data points $\{(x^{(i)}, y^{(i)})\}_{i=1}^n$ and the regression line. Label the axes.

   (b) Solve the corresponding least squares problem. Plot the obtained regression line (in different style and color) in the same figure as in part (a).