BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY
## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CSE 218 (Numerical Analysis Sessional), January 2018 Term

Assignment 1:

You have to implement multiple linear regression for a given dataset. There will be two files, train.csv and test.csv. Each files contains 800 and 200 rows (records) respectively. Now, you will use train file only to fit the data using multiple linear regression. The first line in both file denotes the header. There are six input values (you may consider them x1,x2,x3,x4,x5,x6): longitude, latitude, total_rooms, total_bedrooms, population, and households for a record. There is only one output value(y) in each record: the median household value. Now you have to do the following tasks.

1. Plot a graph to show the data visually. In short you will output a graph where x axis denotes the longitude, y axis denotes latitude, each value is represented by a circle where the radius of the circle is proportionate to population. And use different color to denotes the output (median household value). A sample output is provided below.
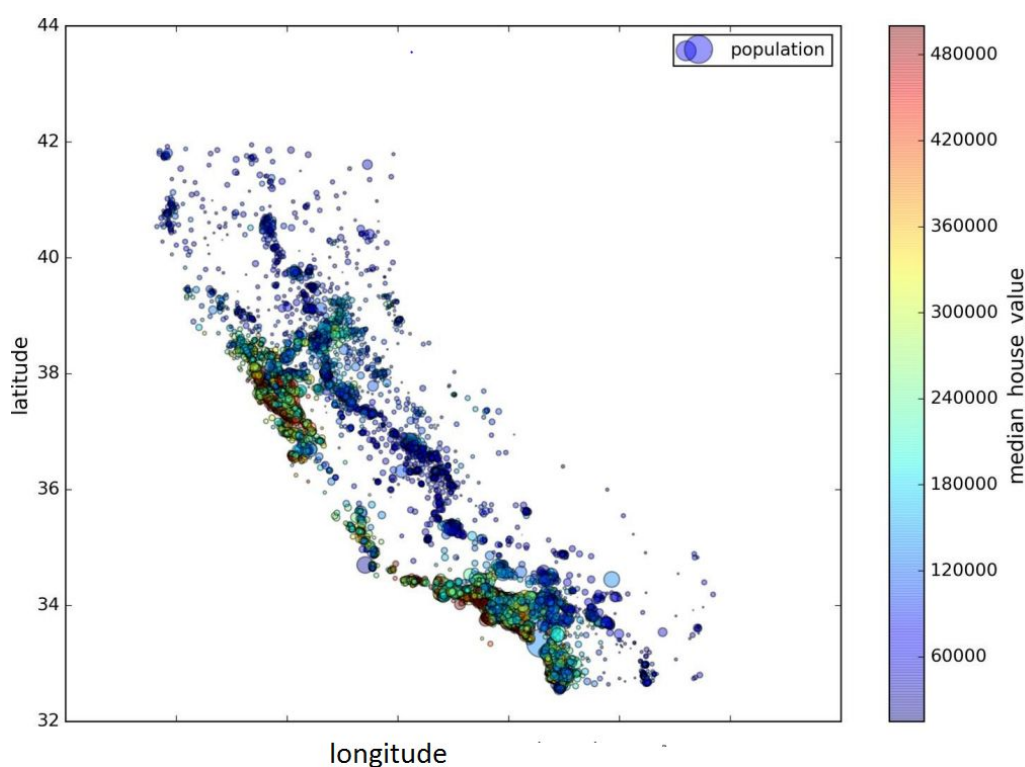
2. Use the train data to fit a multiple linear regression of the form, y = a0 + a1x1+ a2x2+ …… + a6x6 ____ (i). Show the equation.

3. Use the test data to evaluate your built model. For each record in test data, plug the value of x1,x2,....x6 and get the predicted value of output y'. If y is the actual output for that record then, the error of this record = | y- y'|. You need to calculate root means square error for both train data and test data separately. Root mean squared error is defined by,s

$$RMSE = sqrt(\sum_{i=0}^{n}(y_i - y'_i)^2/n)$$

You also need to provide the same graph as discussed earlier for test data.

Instructions:
- You whole task should be evaluated by running a single .m file. However you may use additional file for function implementations.
- You must take input from the given files. No static or runtime input will be accepted.
- You program should output two graphs, the equation and two RMSE error values

Assignment 2:

In this, task you have to use interpolation to fit missing values of a time series dataset. You will be provided a dataset of shampoo sales over three years periods (values given at each month). There are some missing values in dataset (**Note that the missing time points will not be same while evaluating, ie you might be given a dataset where missing time will be different. However total time points will be 36 always**). You have to predict these values from interpolating data for various interpolation methods taught in class (linear spline, newton's polynomial, lagrange's polynomial, quardic spline). Now you have to do the following tasks.

1. Plot the graph to show data visually. X axis will denote the time(month) and Y axis will denote the amount of shampoo sold. You have to use circle to represent each data point. Simple plt(x,y) is not allowed.
2. Fit the data using above mentioned interpolation method. Then again show the graph separately connecting each point smoothly.
3. For missing values, predict the shampoo sales values and show them on previous graphs in different color circle. You also need to output the values for each method in console.

Instructions:
- You whole task should be evaluated by running a single .m file. However you may use additional file for function implementations.
- You must take input from the given files. No static or runtime input will be accepted.

- You program should output five graphs in total (one for initial data plotting and rest four for each method showing the interpolation function and output values of missing data)

Submission instruction*:

You must have to use the following format to submit the offline in moodle.

```
1605xxx
        --Task 1
                -- All your files for assignment 1
        --Task 2
                -- All your files for assignment 2
```

So in short you will have a folder in your student id. Inside that you will have two sub folder for each task. Then you will put all your files in these sub folders as necessary. Zip the folder 1605xxx.zip (Not rar, not 7z, not tgz) and upload it in the moodle. **Any other submission will be penalized in marks reduction greatly.**