CS – 531: MACHINE LEARNING
(Assignment# 2)

Date: 3th October, 2018 **Max. Marks: 100**
Due Date: 10th October, 2018

In this assignment, you will implement linear regression algorithm **from scratch** and then use it to predict popularity of news. This is based on a dataset from the UCI data repository. The dataset contains a total of 39797 examples. Each example is described in terms of 61 attributes, i.e. "shares" (continues numbers) as a class attribute, and 60 other features. These are in comma-separated-file format with the first row in the file listing the attribute names (separated by commas), and remaining rows in the file giving the values of these attributes for single news i.e. each row contains a single example.

Implement the linear regression learning algorithm described in class. You may use any programming language you like, but **Don't Use Any Libraries like (sklearn)**. You may use numpy, scipy, pandas.

Your submission must include following things:

1. Exploratory data analysis on the dataset
2. Calculate Mean square error to evaluate your algorithm.

Once you have implemented the algorithm, You need to use it and train a linear regression model to predict "share" (each new share how much time) based on other 60 features given in dataset.

Bonus: You may use feature engineering and other similar techniques to enhance the performance of your model.

Submission:

- The file **AssignmentNo_YourName_YourRollNo.zip**. **Report and Code** should be submitted on Google Class Room.
- Submission other than Google Class Room won't not be accepted.

Note that if any single line of code matches with implementation of your classmates or with any codes available on Kaggle or Github, you will get ZERO in **TWO Assignments**.

BEST OF LUCK!!