

## Lab Assignment 02

**Total marks: 30**

**Instructions:**

1. Do not use any package until it is specified to use in the problem.
2. Plots should be labeled clearly.

### 1 Nonlinear Regression [20 Marks]

Consider a classic example of throwing up a tennis ball in the air. We can predict the ball's height ( $h$ ) at any instance of time ( $t$ ) using Newton's laws of motion (ignoring air resistance). A data set (prob1data.txt) which follows similar trajectory is provided for training your model.

1. Plot the training data. Write a code in Python to perform nonlinear regression on the given data. Implement batch gradient descent algorithm for optimization. (Choose  $\alpha = 0.01$ , number of iterations = 50000) [8 Marks]
2. Implement stochastic gradient descent for optimization of weights. Plot cost history ( $J$ ) vs number of iterations for both cases batch gradient descent and stochastic gradient descent. Comment on the difference, if any. [Hint: Refer to the example code.] [2 Marks]
3. Plot the cost history ( $J$ ) vs number of iterations for different learning rates ( $\alpha = 0.1, 0.5, 0.01, 0.05$ ). Write your inferences from the plot. [2 Marks]
4. Implement line search method (Secant method) to find learning rate ( $\alpha$ ). Plot the cost history ( $J$ ) vs number of iterations. Comment on the difference between implementing line search method and choosing arbitrary  $\alpha$ . [6 Marks]
5. Plot the predicted ball's height ( $h$ ) values for a given time range (let say  $t = 0$  to  $2.5s$ ) using the trained model. Compare it with the given training data set on the same plot. [2 Marks]

### 2 Multivariate Regression [10 Marks]

Housing Price Prediction Problem. Suppose 'Mr. X' is planning to buy a house in Delhi and wants to predict the price of the house given some features like number of bedrooms, number of bathrooms, area of the house, etc. The file 'prob2data.csv' contains a training set of housing prices in Delhi.

1. Read the excel file using *pandas* and perform data cleaning. Remove 1<sup>st</sup> column 'id' which may not be necessary here. Perform mean normalization of features. [2 Marks]
2. Write a Python code to perform multivariate regression to predict the house price. Consider all 5 columns ('bedrooms', ..., 'yr built') as features. Implement batch gradient descent for optimization of weights. [5 Marks]
3. Predict the house price using the model, for 4 bedrooms, 2.5 bathrooms, 2570 sq. feet area, 2 floors, 2005 yr. built, and state the difference between the model prediction and actual value (Rs. 719000). Show in % error. [1 Mark]
4. Plot the cost history ( $J$ ) vs number of iterations for different learning rates ( $\alpha = 0.1, 0.5, 0.01, 0.05$ ). Write your inferences from the plot. [2 Marks]