Assignments on Logistic Regression #4 (Ref Video Lectures 13-20)

(Timely submission of assignments is essential. Copying/plagiarised submission from others will fetch fail (F) grade on this subject)

1. Our task is to build a Logistic Regression based classification model that estimates an applicant's probability of getting admission to an institution based on the scores from those two examinations whose data have been provided here (you may use 70% data for training and 30% for testing).

a) Design a Predictor with two basic features which are given using Batch Gradient Descent Algorithm, Stochastic Gradient Algorithm and mini batch Gradient Descent algorithms (determining minibatch size is your choice- here it could be 10, 20, 30 etc.) with and without **feature scaling** and compare their performances in terms of % error in prediction.(only allowed to use NumPy library of Python, no other functions/libraries are allowed).

b) Inject more features from the data set in the model (at least 6-9) and repeat (a)

c) Add regularization term and repeat (b). Submit comparative analyses of your results.

2. After gaining experience of solving problem No 1) Design a classifier using logistic regression on Cleveland Medical data set for heart disease diagnosis. The processed dataset with some 13 features have been given with a label that a patient has a heart disease (1) or not (0). This design should have a professional touch within your ML knowledge. Below the data set brief description has been provided. Also, if you want to have more knowledge about the very interesting data set, please go through the link provided.

Hints both for problem #1 and 2:

Evaluating logistic regression classifier:
1. After learning the parameters, you can use the model to predict whether a particular student will be admitted. For example a student with an Exam 1 score of 45 and an Exam 2 score of 85, you should expect to see an admission probability of 0.776. Another way to evaluate the quality of the parameters we have found is to see how well the learned model predicts on our training set. Create a predictor function which will produce "1" or "0" predictions given a dataset and a learned parameter vector W.
2. Use Confusion matrix to evaluate the performance of your classifier.

Deadline for submission: September 25, 2020, mid night.

Full marks: 50+50=100