Homework 1 (100 points)

Logistic Regression Implementation.

In this homework, you will be implementing Logistic Regression on a binary classification task. All your responses must be saved in a jupyter notebook and pushed to your git repository. Upload a .txt file with a link to your file as your submission on Submitty.

You need to perform the following tasks for this homework:

Task 1 (10 points): Describe a machine learning problem that you would like to solve using Logistic Regression. Clearly state why Logistic regression is the best choice for solving this problem.

Task 2 (20 points): Pick a publicly available dataset that you will use to solve this problem. You must provide a link to the dataset and perform necessary Exploratory Data Analysis (EDA). Clearly demonstrate the steps you follow for your EDA with a justification of why these were required. For example, if the dataset has lot of missing values, then why did you use a specific technique when handling missing data. This task may include data visualization (Check this link : https://www.geeksforgeeks.org/top-8-python-libraries-for-data-visualization/).

Task 3 (35 points): Implement, Logistic Regression in this step. Clearly write your cost function and derivatives before implementing gradient descent. Do not use any built-in packages for this step. You can use the vectorization techniques demonstrated in class. Implement any 2 variants of gradient descent in their original form. (Refer to the research paper discussed in class).

Task 4 (35 points): Pick any 2 optimization algorithms that are used to optimize the 'vanilla' gradient descent. Implement both. You may implement these algorithms yourself OR use a package. In your conclusion, compare both optimization techniques/algorithms with respect to the results you achieve. Also compare these results with the original implementation of gradient descent (Task 3 above). Describe why or why not should we use optimization algorithms for the task at hand.