HOMEWORK 5

Problem.

default.csv is a simulated dataset which contains three features pertaining to an individual (whether the person is a student, what their income is, and what their balance is) and one outcome variable (did the individual default on their card payment). After splitting the data into a training and test set, we can train a logistic regression model on this dataset.

Using either all three or any two of the features, train a logistic regression model using Newton's method. Clearly express your gradient function and the Hessian (i.e., matrix of second partial derivatives) which you will need to compute while implementing Newton's method.

Compare the coefficients obtained with a black-box implementation of logistic regression with the coefficients obtained above using Newton's method.

Does the Newton's method converge with ease? In case of non-convergence, set your initial points to be close to the estimates obtained using the black-box implementation and check.

Bonus. Modify the stochastic gradient descent routine you have implemented in Homework 4 to implement a mini-batch stochastic gradient descent method to train a logistic regression model for the above dataset. Use mini-batch sizes of 16, 32, or 64.