MNIST Report

Ahmad Hassan | 30055847

My neural net consists of an input layer, 1 hidden layer and an output layer. I chose this structure because according to my research, having more than 1 hidden layer provides no added benefit while increasing computational cost. My hidden layer contains 397 neurons which is the mean of the input size and output size.

For my activation functions I used “ReLu” for my hidden layer as it is computationally inexpensive, is non-linear and has sparse activation making it efficient with a large number of neurons. As for my output layer I used “softmax” as it is suitable for classification making it a good fit for the desired outcome of this model.

I chose “adam” as my optimizer because it has low training error and does not tend to underfit in low epoch training scenarios. This is contrast with SGD that has problems with underfitting.

Code running:

To run my code, upload the file to Google colab and execute all the code blocks in sequence.