CS 547 MP1 Report

Hongpeng Guo hg5@illinois.edu

In this MP, I implemented a neural network with one hidden layer of dimension 128. The input layer is of dimension 784 and the output layer is of dimension 10.

I chose the activation function to be the sigmoid function.

I implemented the mini-batch SGD algorithm with batch size to be 5.

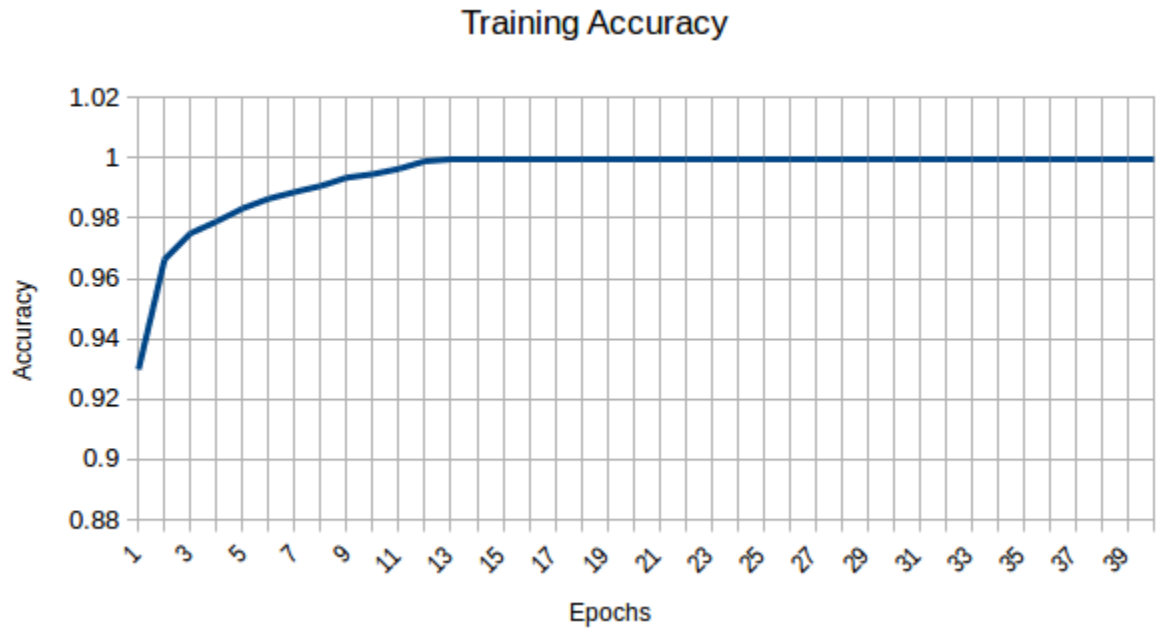
To achieve better results, we normalize the matrix W and C following $N(0, \frac{1}{d_H})$.

The training process contains 40 epochs, with gradually decreasing learning rates as below:

0 <= epoch < 10	LR = 0.1
10 <= epoch < 20	LR = 0.01
20 <= epoch < 30	LR = 0.001
30 <= epoch < 40	LR = 0.0001

The final accuracy of our model on the test data set is: 97.99%

The training accuracies of each epochs are plotted and listed as below:



0.9298333333333333
0.9663333333333334
0.975
0.9789
0.9832666666666666
0.9864666666666667
0.9887
0.9907166666666667
0.9934833333333334
0.9946333333333334
0.9963833333333333
0.9989333333333333
0.99955
0.99965
0.99965
0.9996666666666667
0.9997166666666667
0.9997333333333334
0.9997666666666667
0.9997666666666667
0.9997833333333334
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985
0.99985