Usama Munir Sheikh (661541154)
Deep Learning
Programming Homework 03

Problem: Implement in TensorFlow, from scratch, Neural Network with Two Hidden Layers of 100 nodes each. Train the network on the MNIST digits data for 10 classes.

Results Required:

1) Final classification error for each digit on the test data set after training the NN

And plots for the

2) average training classification error,
3) average testing classification error,
4) and value of the loss function after each parameters update.

Solution

1) This submission has my code for the backpropagation implemented in,

`pghw3_alltrain.py`

along with the corresponding plots (also shown below and the parameters given in,

`nn_parameters.txt`

To run this code, keep the file **inside** the directory, which has the subdirectories for `labels`, `test_data` and `train_data`.

To get good training accuracy and relatively fast convergence, I had to lower the learning rate as the accuracy went higher and the loss went lower. (here I use two different learning rates to get good convergence)

2) I will show the results of using just a fixed learning rate where I only achieved a training accuracy of 91% and a test accuracy of 90%. I had to stop this algorithm early since the higher value of learning rate makes the algorithm diverge as I near the convergence point.

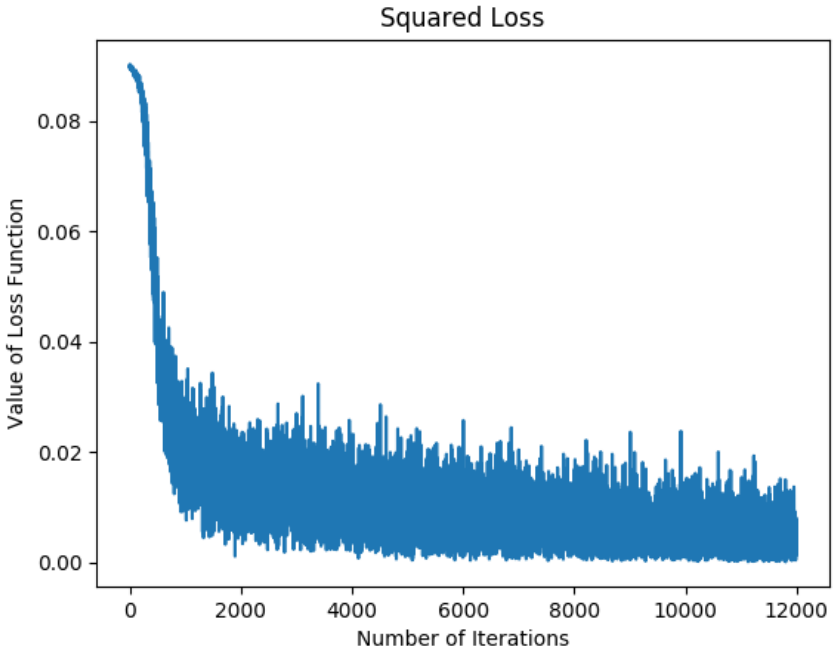
3) This submission also contains a folder by the name of

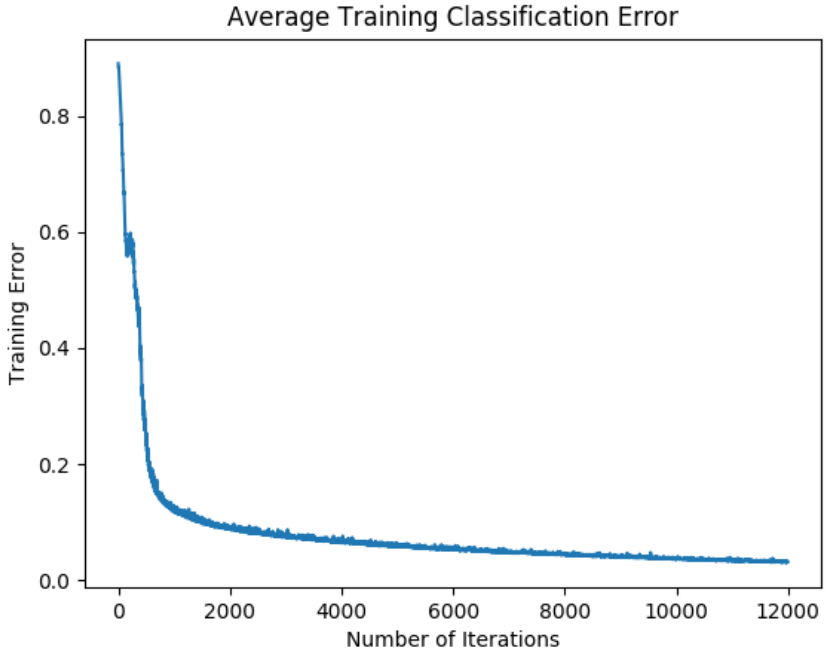
`Results_theirCode_adam`

This folder contains results found using the TensorFlow built-in back propagation method called adam optimizer for testing purposes. The results and code using the adam optimizer are given in this folder.

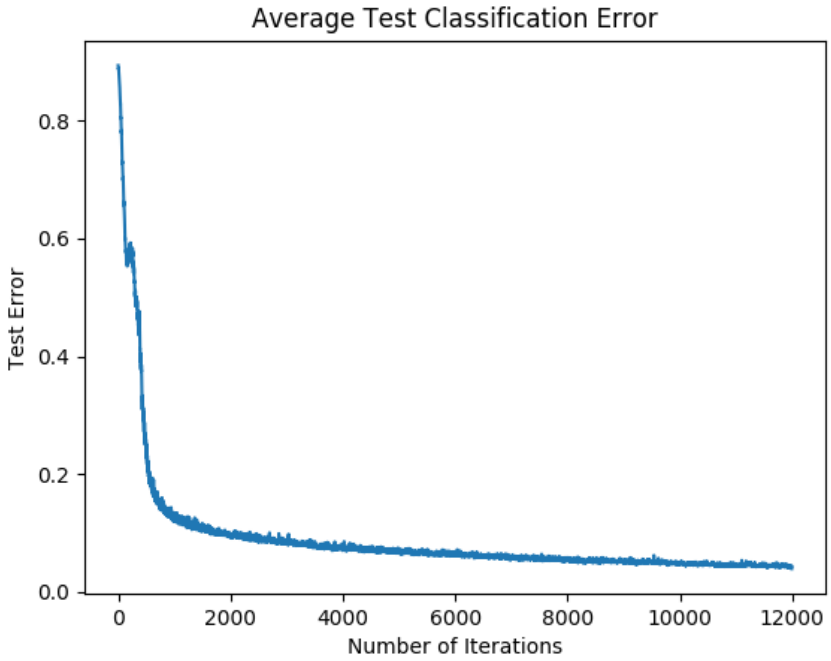
Results – Final (My Code Using Gradient Descent but by changing learning rate as loss becomes lower)

Usama Munir Sheikh (661541154)
Deep Learning
Programming Homework 03





Usama Munir Sheikh (661541154)
Deep Learning
Programming Homework 03



Final Training Accuracy: 96.9%

Final Test Accuracy: 96.1%

Learning Rate: Changed from 0.4 to 0.001

Total Number of Iterations: 11990

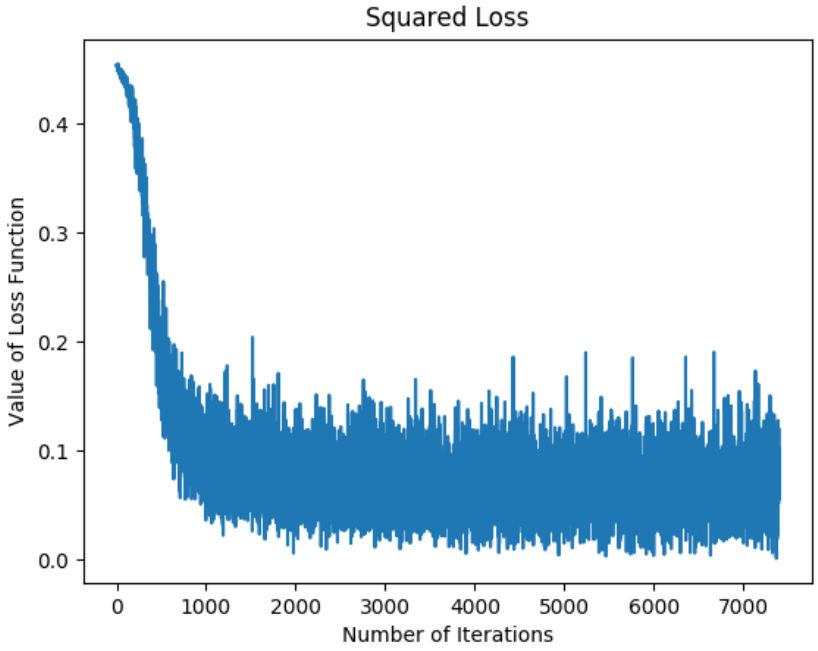
Final Value of Squared Loss: 0.004044

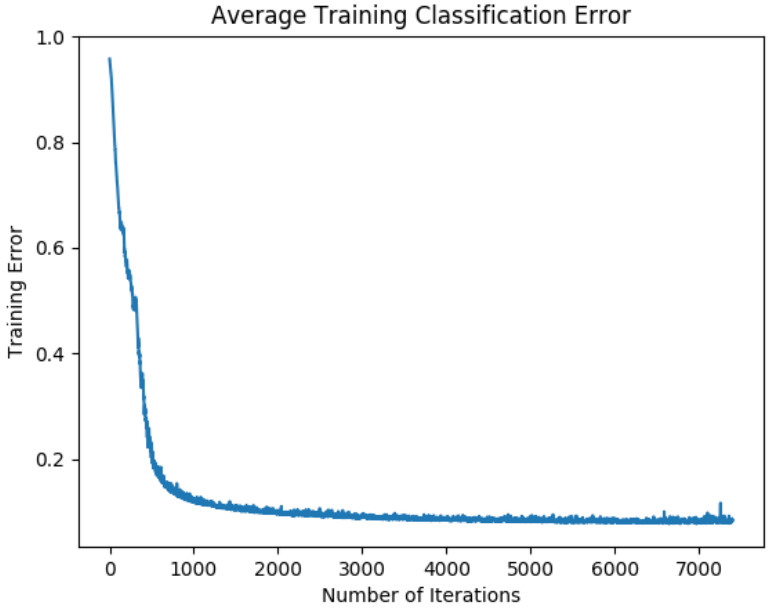
Classification Error For each Digit:

$$[0.96 \quad 0.00 \quad 0.00 \quad 0.00 \quad 0.00 \quad 0.00 \quad 0.00 \quad 0.00 \quad 0.00 \quad 0.97]$$

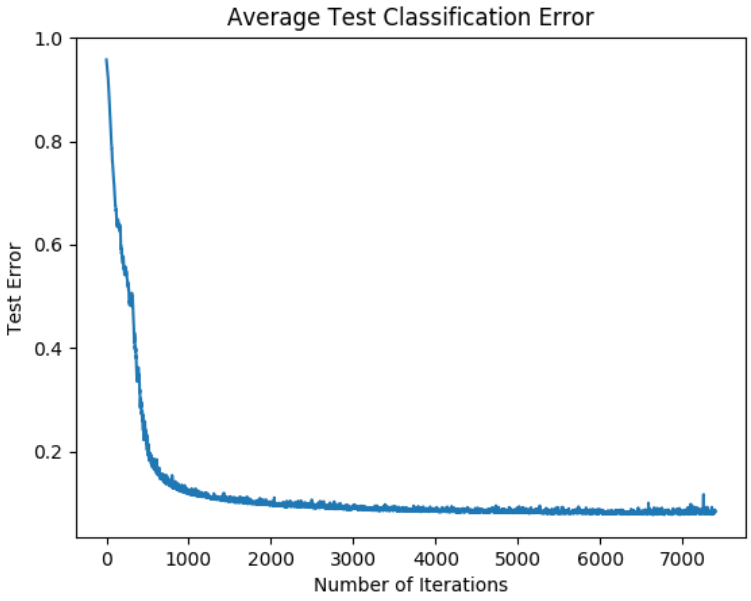
Results – First Try (My Code Using Gradient Descent but with early stopping Did not find Individual errors for this because I stopped early.)

Usama Munir Sheikh (661541154)
Deep Learning
Programming Homework 03





Usama Munir Sheikh (661541154)
Deep Learning
Programming Homework 03



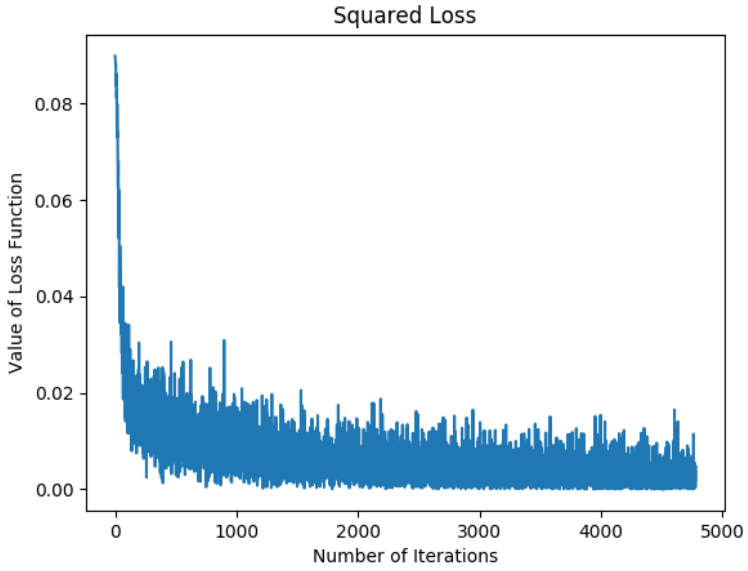
Final Training Accuracy: 91.2%

Final Test Accuracy: 90.1%

Learning Rate: Changed from 0.1

Total Number of Iterations: 7411

Results – Testing the Adam Optimizer built into TensorFlow



Usama Munir Sheikh (661541154)
Deep Learning
Programming Homework 03