

ECSE 4965: Program Assignment 3

Name: Li Dong

RIN: 661243168

March 17, 2017

It is noted that the image data are converted to a compact .txt file for easy manipulation. The preprocessing functions are also included in the source code.

Plot for the loss function value, average training classification accuracy and average testing classification accuracy can be seen from Figure 1. It is noted that the classification accuracy is simply defined as 1-classification error.

A mini-batch size of 50 is used for stochastic gradient descent as asked. The learning rate is set to 1. During training, weight matrix are initialized to $0.01 \times N_{trunc}(0, 1)$, where N_{trunc} is the truncated normal distribution. In this case, values outside two standard deviations are resampled. Bias are initialized to zeros as asked.

The weights are saved in nn_parameters.txt as asked.

By applying the weights trained from 50,000 training data to the 5,000 testing data, we can get the following results:

digits	average error (%)
all	4.360
0	0.797
1	2.060
2	3.018
3	7.102
4	5.165
5	6.783
6	2.887
7	2.657
8	6.019
9	7.531

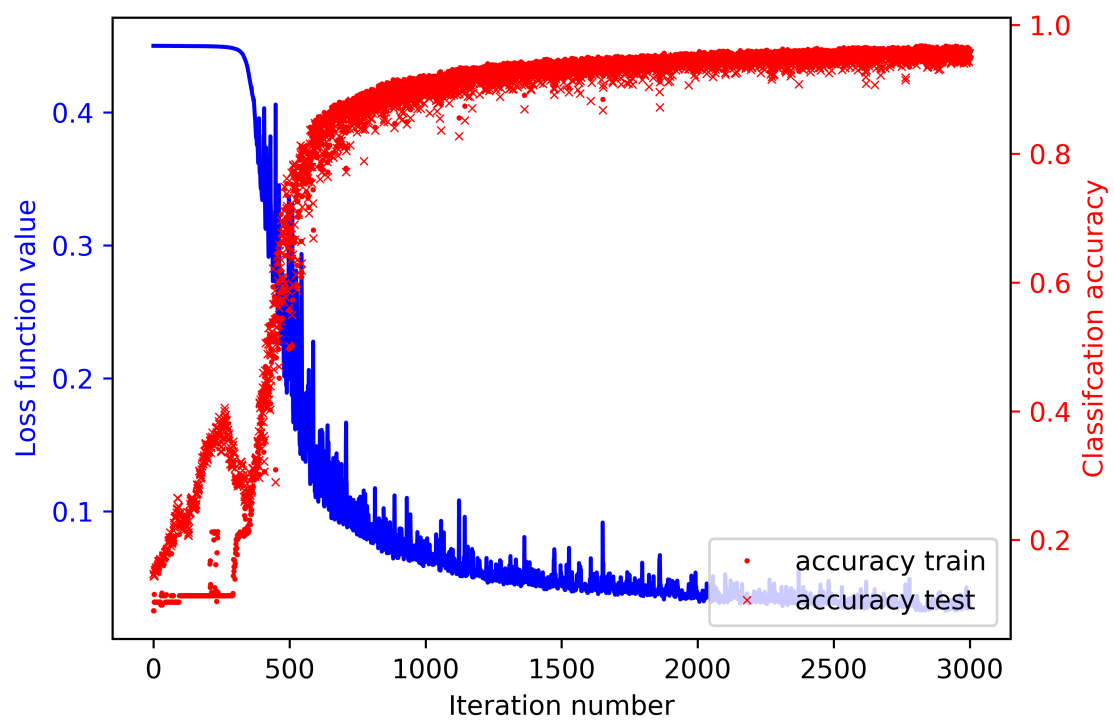


Figure 1