096260 - Deep Learning Course
Exercise #1 - Report
Sergey Voldman, 309572519
Noam Heimann ,300972692

Files Included

The following files are included in this submission:

- `mnist_model.net` - The trained network (with trained weights) in Torch format.

- `load_and_test.lua` - Contains a single function *loadAndTest* for testing the trained model. Calling it with no parameters, it loads the model from file `mnist_model.net`, loads the test set from mnist library in torch, and computes the error.

- `classification_mnist.lua` - File to reproduce the training of the model.

Model Description and Architecture
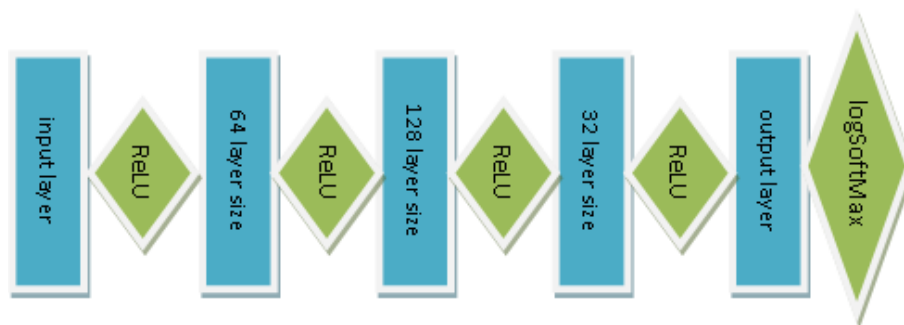
The architecture of our model is as follows:

Figure 1: Model architecture.

We used the negative log likelihood criterion (ClassNLLCriterion) for the loss function, batch size of 64 and trained over 100 epochs. The optimization parameters we applied are learningRate and momentum, fixed to 0.2 and 0.1 respectively.

Training Procedure

We normalized the data to have mean 0 and standard deviation 1, and reshuffled the order within the training data each epoch. We trained the network for 100 training cycles and return the weights (w) which achieved the smallest test error.

We tested several parameter tunings, including

1. layer architecture

2. criteria (loss functions)

3. optimization parameters - learning rate, momentum, weight decay rate

4. batch size

We found the architecture to be the most influential factor in achieving better results, with the restriction of appropriate batch size to accompany it. With this setting we were able to achieve our best accuracy, which is 98.45% on the test set and it is above the minimal 90% required. The total amount of parameters used in this setting is 63018 which is below the 65000 required. Figures 2 and 3 depict the error and loss versus time(epochs) respectively.
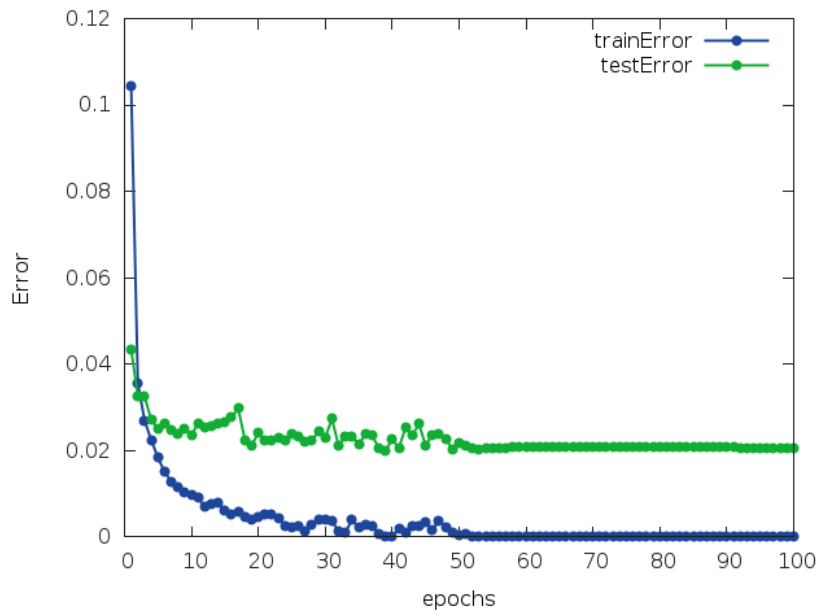
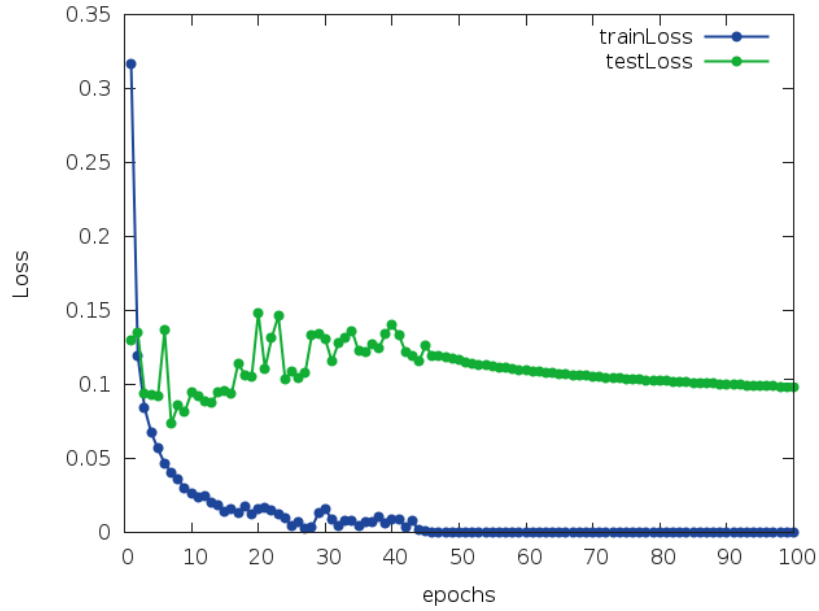Figure 2: Classification error of train vs. test over time.

Figure 3: Value of loss function of train vs. test over time.