Programming Exercises Problem 01 Report

Deep Learning for Engineers

Name: Jen-Hung Ho   **Andrew ID:** jenhungh

(1) Structure of the Multi-layer Perceptron:

There are four activation functions that users could choose to apply to each layer - Identity, Sigmoid, Tanh, and ReLU. Each of the activation function contains forward and backward propagation. The most normal Criterion of multi-class classification is Softmax function and Cross-Entropy Loss. In the SoftmaxCrossEntropy Class, there are three methods – softmax, forward, and derivative. The first two are used to do forward propagation, while the derivative is used to do backward propagation.

My Multi-layer Perceptron has four layers – input layer (784 neurons), two hidden layers (128 and 64 neurons), and output layer (10 neurons since there are 10 classes). The activation function for two hidden layers are Sigmond functions. When training the MLP model, we implement forward propagation, backward propagation, and step sequentially to update the weights and bias.

(2) Hyper-parameters

Number of neurons in each layer: [784, 128, 64, 10]
Batch Size of the training/test dataset: 8
Number of Epochs: 100
Learning Rate: 0.01

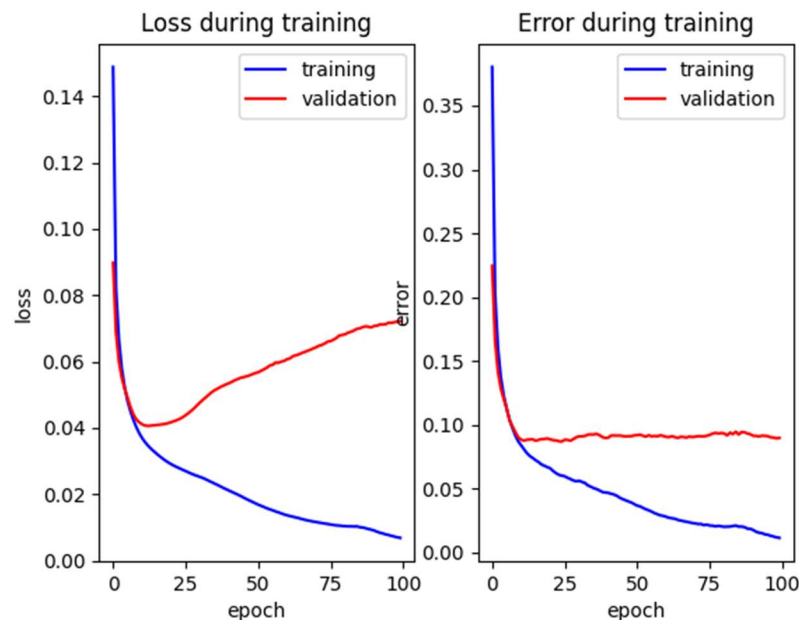### (3) Visualization of the Training Process

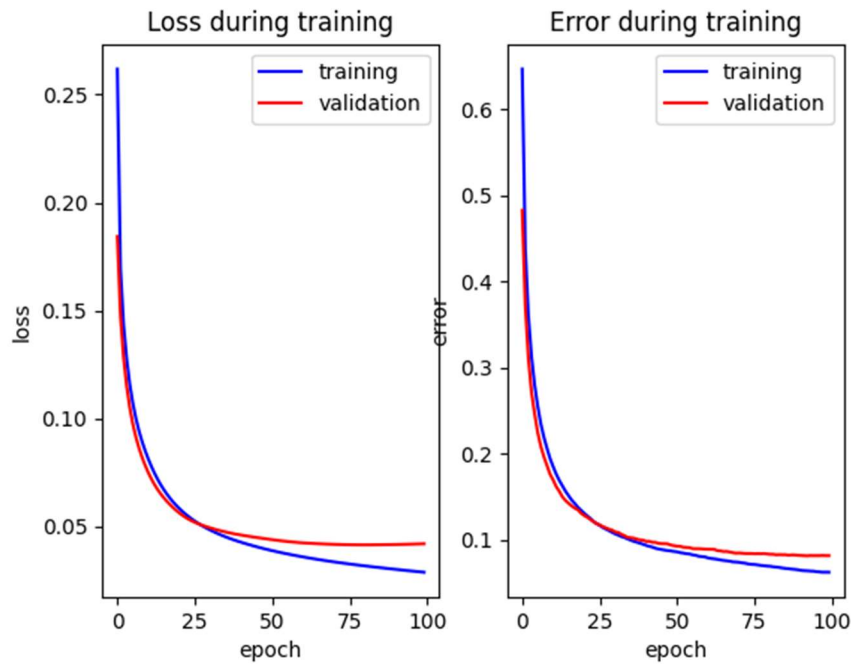Figure 01. Loss and Error of Training and Validation Dataset (lr = 0.05)

Figure 02. Loss and Error of Training and Validation Dataset (lr $= 0.01$)

When training my MLP model, I initially set the learning rate to 0.05 and found that the model is overfitting (as Figure 01 shown). Therefore, I decreased the learning rate to 0.01 and trained the model successfully (as Figure 02 shown). The training process is shown in the attached Training_Process.txt file.

There are 60000 input images and they are split into 50000 training images and 10000 validation images. The batch size is 8, which means the weight and bias will update 6250 times in a single epoch during the training process.

(4) Final Performance
Test Loss of the MLP on the 10000 test images: 0.044075658928084294
Test Error of the MLP on the 10000 test images: 8.89 %