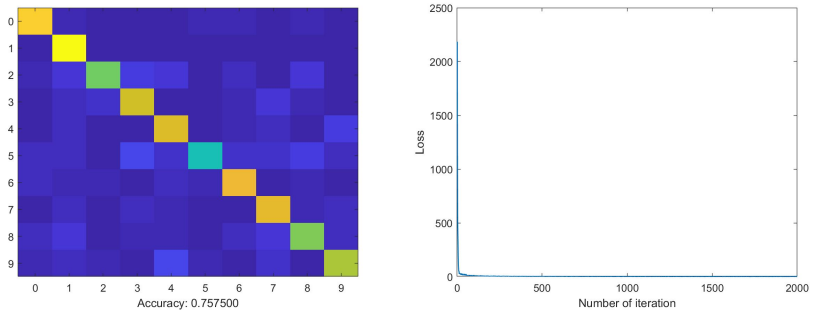
Single-layer Linear Perceptron

Single-layer linear perceptron is trained using stochastic gradient descent method and no activation function is used for digit classification. The loss function is Euclidean.

Parameters: Learning rate=0.01, Decay Rate=0.8, Accuracy: 75.75%.



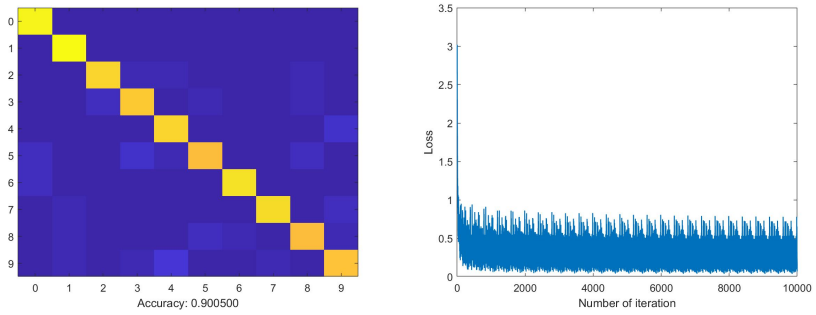
(a) (b)

Figure 1: (a) Confusion matrix for SLP_linear, (b) Loss

Single-layer Perceptron

Single-layer perceptron is trained using stochastic gradient descent method for digit classification. The activation function is Softmax and the loss function is Cross Entropy. [The loss in fig (b) looks fluctuating because the scale is small]

Parameters: Learning rate=0.01, Decay Rate=0.8, Accuracy: 90.05%.



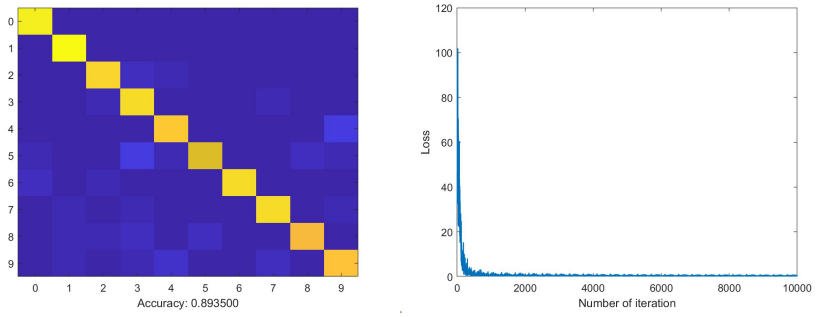
(a) (b)

Figure 2: (a) Confusion matrix for SLP, (b) Loss

Multi-layer Perceptron

Multi-layer perceptron is trained using stochastic gradient descent method for digit classification. The activation function is ReLu (except the last one, which has Softmax activation) and the loss function is Cross Entropy. There is one hidden layer with 30 nodes.

Parameters: Learning rate: 0.001, Decay Rate: 0.95, Accuracy: 89.35%.



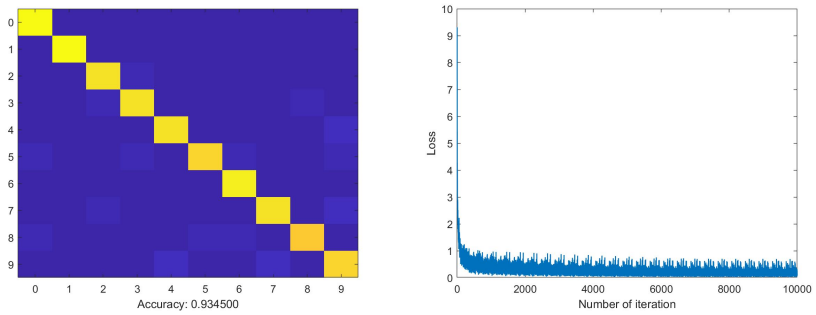
(a) (b)

Figure 3: (a) Confusion matrix for MLP, (b) Loss

Convolutional Neural Network

Convolutional Neural Network is trained using stochastic gradient descent method for digit classification. The model consists of one convolutional layer, followed by a ReLu layer, MaxPooling, and a Fully Connected Layer. The convolutional layer has a 3x3 filter size and 3 filters. The activation function is ReLu (except the last one, which has Softmax activation) and the loss function is Cross Entropy.

Parameters: Learning rate: 0.9; Decay Rate: 0.9; Accuracy: 93.4%



(a) (b)

Figure 4: (a) Confusion matrix for CNN, (b) Loss

NOTE: Number of iterations is 10000 and R is size of minibatch for training every network .