Dataset Description

This dataset is taken from Northeastern University CS6140 Machine Learning course material.

The data set represents images of handwritten digits

Each image was preprocessed into a numerical feature vector of 784 dimension, each dimension is grayscale value of some pixels in the image.

The label of the image is either 3 or 5, which indicates that the handwritten digits in this dataset is either 3 or 5.

There are 6,517 images in the training set and 1,629 images in the test set, which leads to 6,517 lines of data in the "train.csv" and 1,629 lines of data in the "test.csv".

Model Training Description

This code implements a fully connected feedforward neural network with one hidden layer.

The input dimension is automatically determined by the given data.

The number of nodes in the hidden layer could be determined by the user. By default, the implementation uses 100 nodes in the hidden layer.

There is only one node in the output layer since the given dataset actually formalize a binary classification problem.

The model was trained by backpropagation algorithm. By default, after 50 epoches of iteration, the algorithm stops, which could be modified by user.

The activation function of all nodes are sigmoid.

The accuracy of trained model on the test set is 96%.