IE 534, homework 1

Shuhui Guo

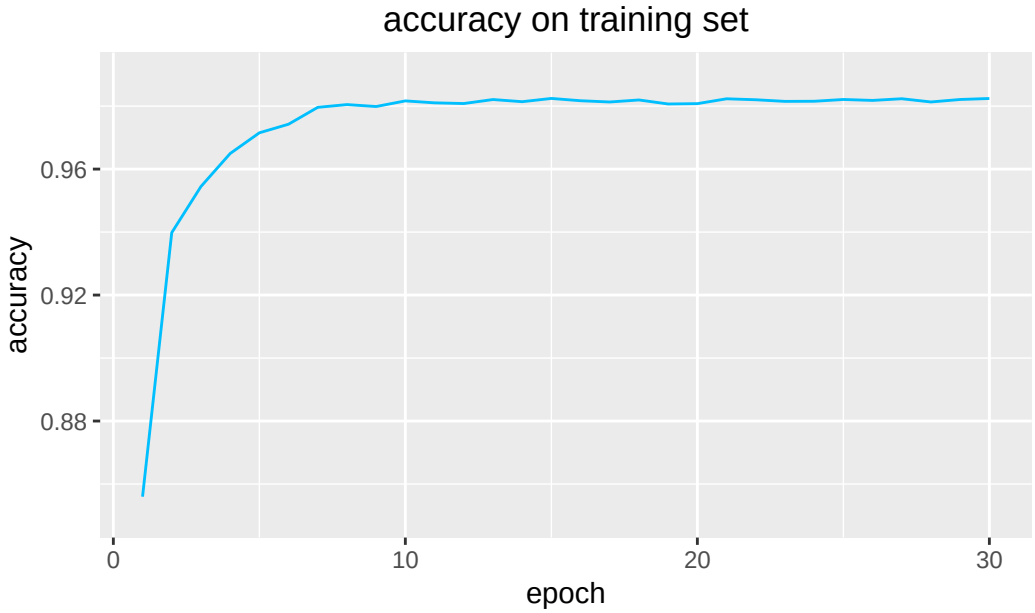
The objective of this homework is training a neural network using stochastic gradient descent to classify handwritten numbers in the MNIST dataset. The dataset is divided into training set and testing set. The neural network is built on the training set and the performance is evaluated on the testing set. The input of the neural network is a $d = 28 \times 28$ array of pixels from the grayscale image and the output is a vector of probabilities for the potential images of size $K = 10$. The most likely outcome is the outcome with the highest probability. To achieve a relatively higher accuracy, the number of units in the hidden layer is set to be $d_H = 150$.

The neural network architecture is

$$\begin{aligned} Z &= Wx + b^1 \\ H_i &= \sigma(Z_i), \quad i = 0, ..., d_H - 1 \\ U &= CH + b^2 \\ f(x;\theta) &= F_{softmax}(U) \end{aligned}$$

where Sigmoidal units $\frac{e^z}{1+e^z}$ is chosen to be the nonlinearity $\sigma(z)$. The parameters of this architecture are $C \in \mathbb{R}^{K \times d_H}$, $b^1 \in \mathbb{R}^{d_H}$, $b^2 \in \mathbb{R}^K$, and $W \in \mathbb{R}^{d_H \times d}$. These parameters are collected in $\theta = \{C, b^1, b^2, W\}$

To efficiently minimize the objective function $\rho(f(X;\theta), Y)$, stochastic gradient descent algorithm is used. The number of epochs is set to be 30. After 10 epochs, the training accuracy keeps above 0.98. The accuracies on the training set after each epoch are plotted as below:



And the accuracy on the testing set is 0.9727.