

Ex3 Report
Caleb Shere
493276919

My model:

I designed a neural network with one hidden layer as the assignment requested. The input layer consists of the 784 “pixels” of each example in the training data. The hidden layer consists of H nodes (I’ll go into the value of H and how I chose it in the next section), and the output layer consists of 10 nodes, corresponding to the 10 possible classifications for each picture. The activation of each function in the hidden layer is the sigmoid function, and I also use the softmax function on the output layer to normalize the output vector into a list of confidence percentages between 0 and 1. There is also a bias vector at both the input layer and the hidden layer. After a forward pass through the network, the weights and biases are updated via back-propagation.

Hyper-parameters:

H = the size of the hidden layer. I found that a larger H generally allowed for more accurate learning while a smaller H allowed for faster runtime. I compromised with 300, which works quite well. In addition, the larger H values are more prone to over-fitting so a value of 300 helps to avoid that.

Learning rate. With this parameter too, I chose compromise. A high learning rate can lead to over-fitting and a sporadic model, while a low learning rate can lead to not enough learning before seeing the test set. After testing a few values, I chose 0.01.

Number of epochs. I wanted to maximize learning while not getting severely impatient, so I chose 50 epochs.

In addition, I normalized the inputs (by dividing all of the values by 255), which led to drastic improvements after my initial results. And last, after fiddling with the range for the initial weights and biases and after consulting with other students, I settled on $[-0.3, 0.3]$ for the biases and $[-0.08, 0.08]$ for the weights.