

First I defined the number of hidden layers and number of inputs and outputs. I also defined the softmax function and sigmoid function. Then I initialized the weights and biases and trained the model by using stochastic gradient descent. I used 0.03 learning rate for the first 150000 iteration and 0.003 learning rate for the next 20000 iteration and 0.0003 learning rate for the final 80000 iteration. Finally I substituted the test dataset for this model and classified the test dataset and calculated the test accuracy. The test accuracy was 97.19% for my computation but it will change a little bit because of the randomness of the initialization of the weights and biases.