EE 228 Homework 2 Report

Tianzi Luo
862251424

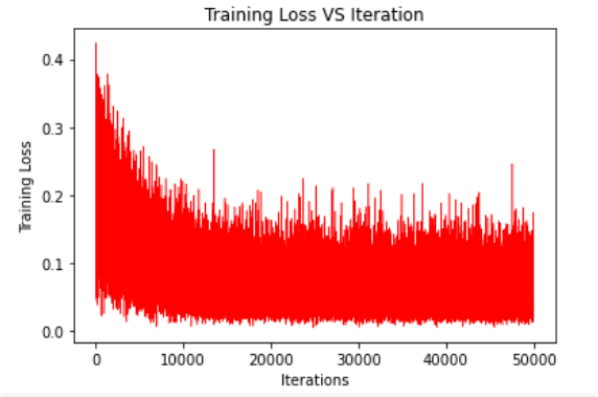
Question 1: Please see the code.

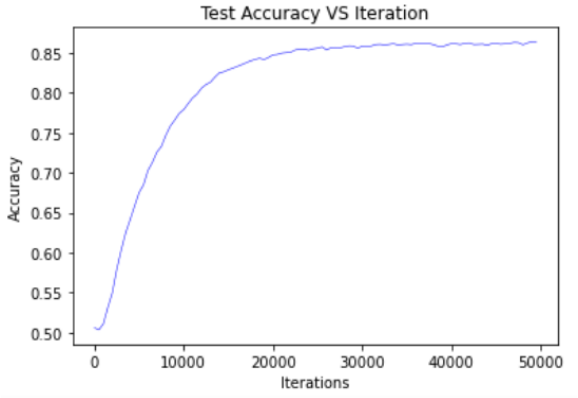
Here is the size of dataset
train set & train set label:
validation set & validation set label:
test set & test set label:

```
train set shape (50000, 785)
test image shape (10000, 785)
validation image shape (10000, 785)
```

Question 2: baseline

linear classifier and quadratic loss, similar to homework 1. We need to change the decision of the neural network.





Finally, the test accuracy is 0.864.

Question 3: identity activation + quadratic loss

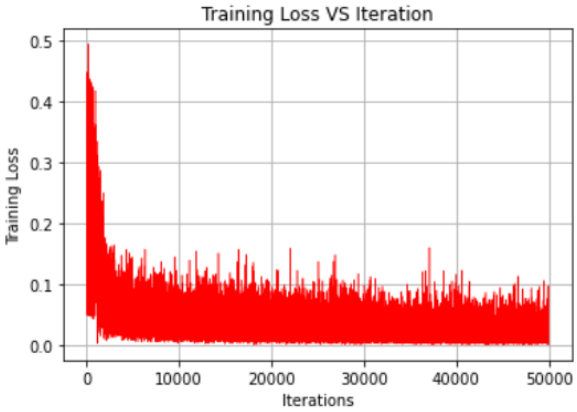
- k=5

Finally, the test accuracy is 0.9363

- k=40

Finally, the test accuracy is 0.9635

- k=200

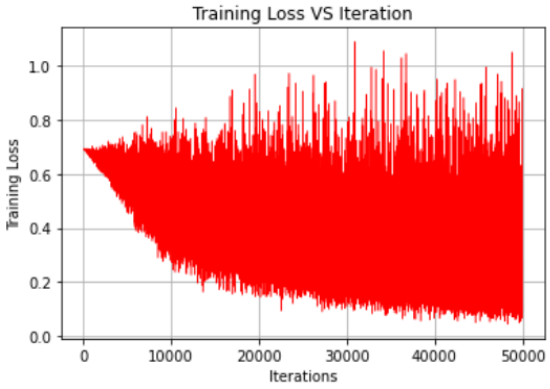


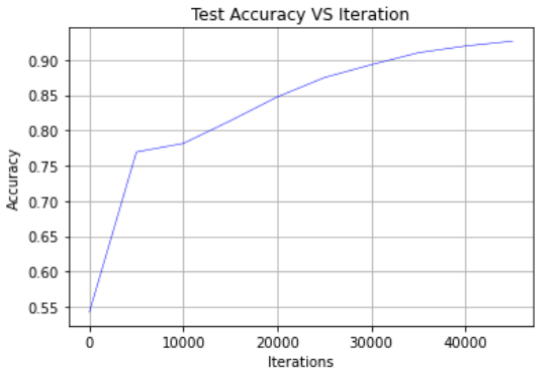
Finally, the test accuracy is 0.9642

As the K increased, the test accuracy increased simultaneously, and the result became stable around 0.96.

Question 4: sigmoid activation + logistic loss

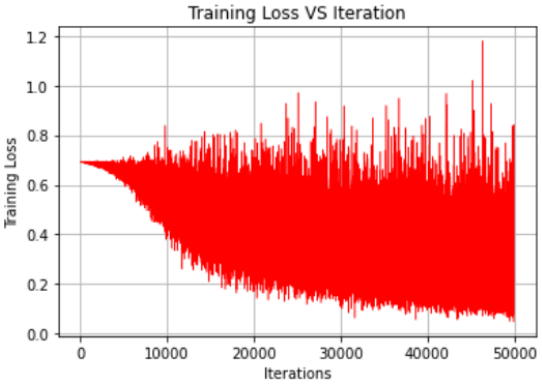
- k=5





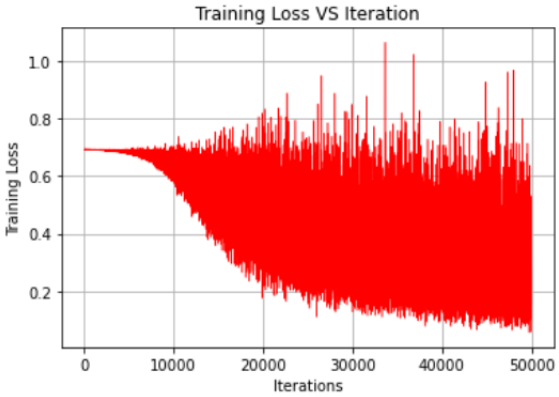
Finally, the test accuracy is 0.9197.

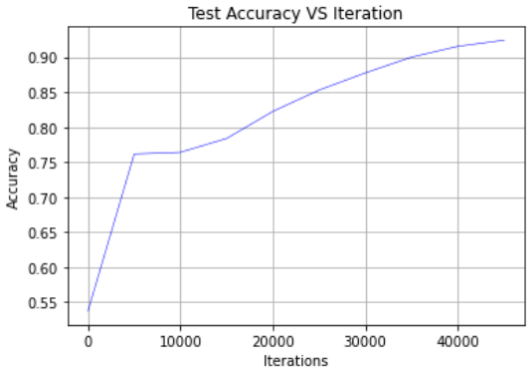
- k=40



Finally, the test accuracy is 0.9262

- k=200





Finally, the test accuracy is 0.9242.

As the K increased, the test accuracy increased simultaneously, but not significant. And the result became stable around 0.92

Question 5:

Difference:

Compared with a neural net, the linear model does not need activation function and its output layer is the same as the input layer.

Neural networks can deal with nonlinearities, we can use activation functions (like ReLU,sigmoid) to introduce a nonlinear factor to the neurons. So if the data has some nonlinear dependencies, neural networks should perform better than linear models.

logistic and quadratic loss:

The accuracy of 3 and 4 are better than 2, so in this scenario, neural networks work better than linear models.

The quadratic loss function gives a measure of how accurate a predictive model is. It works by taking the difference between the predicted probability and the actual value.

The accuracy of 3 is better than 4 in the same condition, so I think quadratic loss is more suitable than logistic loss in this problem.