Intro to Deep Learning Assignment 1

Tsinghua University
Allan Li (李盼加乐), 2025403473

Part 1: Objective Questions

1. Solution: **A, B**
2. Solution: **B, C**
3. Solution: **C**
4. Solution: **A, B**

Part 2: Programming Practice

The purpose of this report is to analyze the effectiveness of the **Softmax Classifier** in recognizing handwritten digits from the **MNIST dataset**. The MNIST dataset contains grayscale images of digits (0–9), each represented as a 784-dimensional vector derived from 28×28 pixel images. The task is a multiclass classification problem where each image must be correctly assigned to one of ten possible digit classes.

Using Initial Hyperparameters

The following hyperparameters were used in the initial implementation of the Softmax Classifier:

```
batch_size = 100
max_epoch = 10
learning_rate = 0.01

# For regularization
_lambda = 0.5
```

The results were as follows:

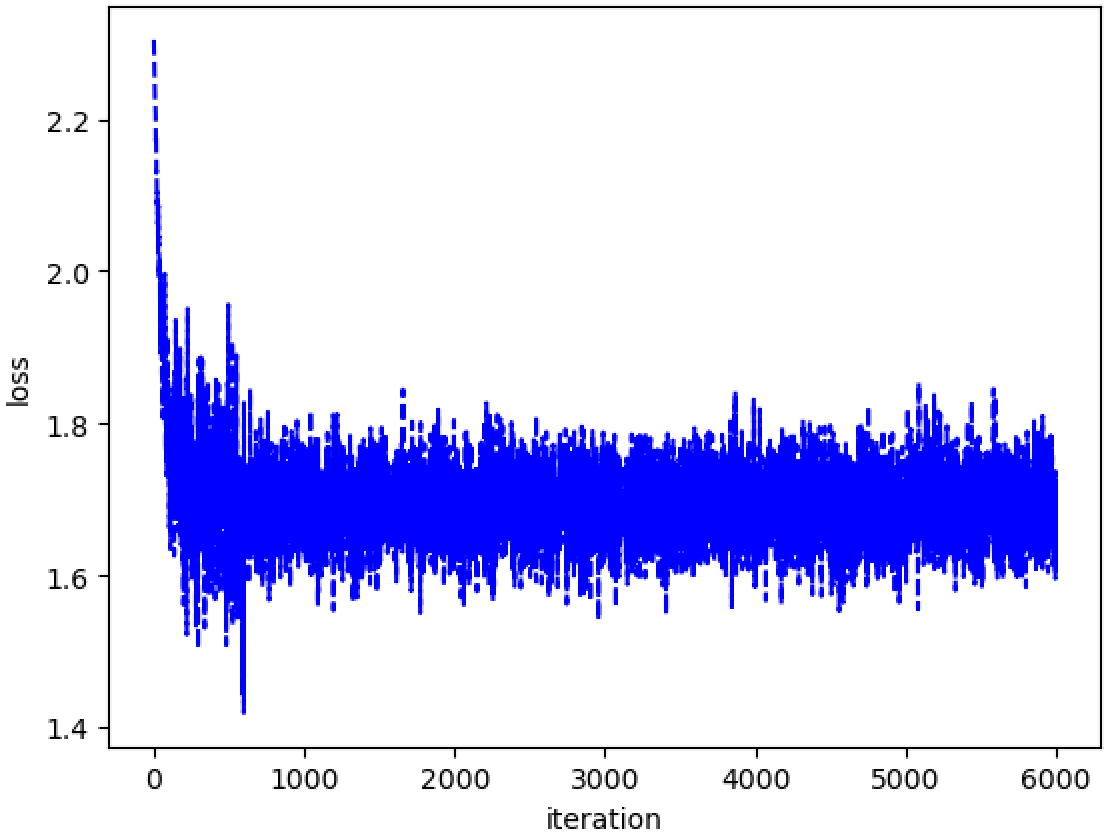

Figure 1: Loss Curve using initial hyperparameters.

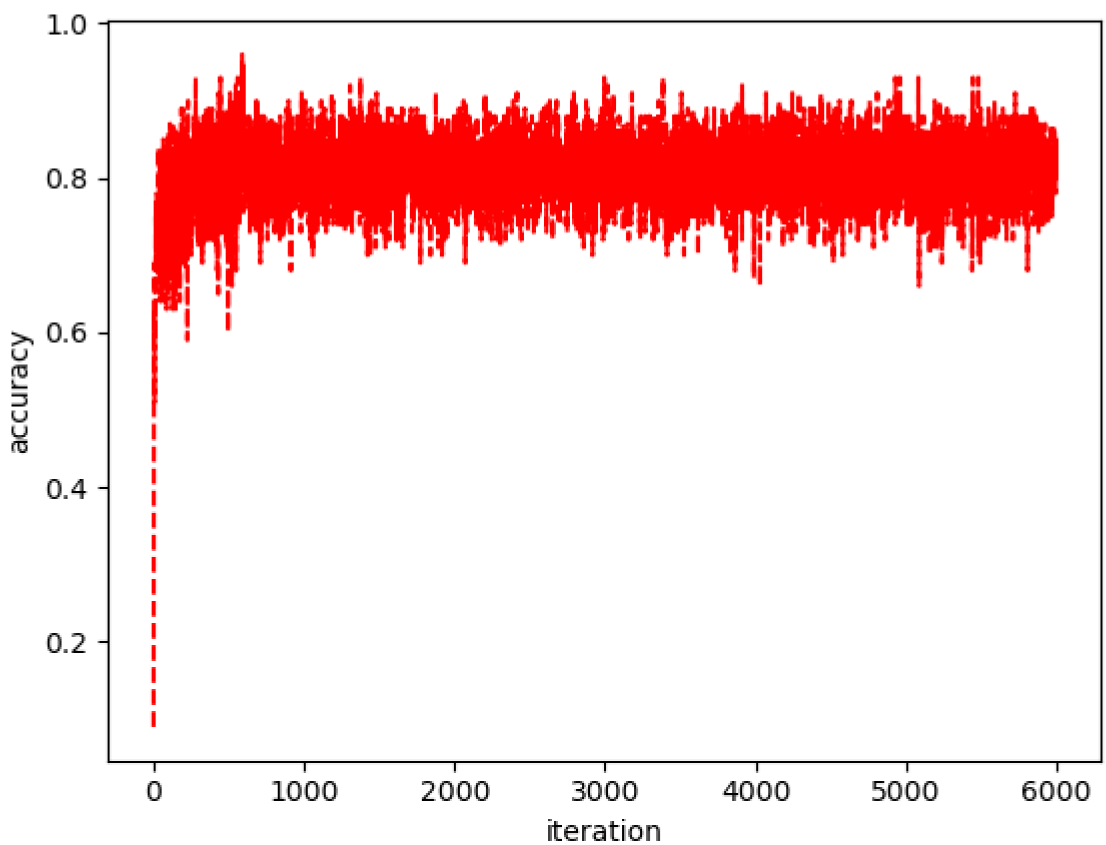

Figure 2: Accuracy curve using initial hyperparmeters.

As well, the **training accuracy** was: **0.8060**, and the **testing accuracy** was **0.8211**.

Training accuracy was calculated as the average accuracy across all batches of the last epoch. Testing accuracy was calculated as the % of correct classifications in the testing set.

Modifying Hyperparameters

Since modifying some hyperparameters might benefit the model, each of the four hyperparameters were individually and independently tested to determine the difference in training accuracy and testing accuracy.

Essentially, these were tested assuming that the only deviation from the initial set of hyperparameters is the one specified within each table (and all else is equal, according to the values above)

Batch Size

Batch Size	Training Accuracy	Testing Accuracy
32	0.8065	0.8156
64	0.8108	0.8176
100 (base)	0.8108	0.8227
256	0.8123	0.8247
512	0.8124	0.8008

Remarks: As batch size increases, training accuracy increases, but testing accuracy eventually starts to decrease as overfitting starts to occur.

Learning Rate

Learning Rate	Training Accuracy	Testing Accuracy
0.001	0.8115	0.8226
0.005	0.8127	0.8255
0.01 (base)	0.8111	0.8255
0.05	0.8014	0.8076
0.01	0.7882	0.7806

Remarks: As the learning rate increases, training accuracy and testing accuracy both decrease as underfitting starts to occur.

Regularization (`_lambda`)

`_lambda`	**Training Accuracy**	**Testing Accuracy**
0.01	0.8907	0.8994
0.1	0.8601	0.8664
0.5 (base)	0.8106	0.8225
1	0.7830	0.7897
5	0.6946	0.7489

Remarks: As lambda increases, both the training and testing accuracy decrease because the model overpenalizes large weights and underfits the data.

Epochs

Epochs	**Training Accuracy**	**Testing Accuracy**
5	0.8108	0.8287
10 (base)	0.8106	0.8237
15	0.8118	0.8234
20	0.8114	0.8163
30	0.8096	0.8212

Remarks: As the number of maximum epochs increases, the training and testing accuracies do not vary much, because the model converges quickly.