Naka Assoumatine
CSCI 315 Assignment #4

PyTorch MNIST Classification

This assignment focused on implementing and experimenting with a Multi-Layer Perceptron (MLP) and a Single-Layer Perceptron (SLP) using PyTorch to classify the MNIST dataset. The goal was to explore hyperparameters, evaluate performance, and compare CPU vs. GPU execution times.

Results

1. Training with Different Epochs

Epochs = 20: Training loss decreased steadily, finishing in 30.40 seconds.
Epochs = 40: Training loss continued to decrease, finishing in 59.51 seconds.
Epochs = 60: Training loss showed further improvement, finishing in 90.46 seconds.
Observation: Increasing the number of epochs improved the training loss, but the improvement diminished after 40 epochs.

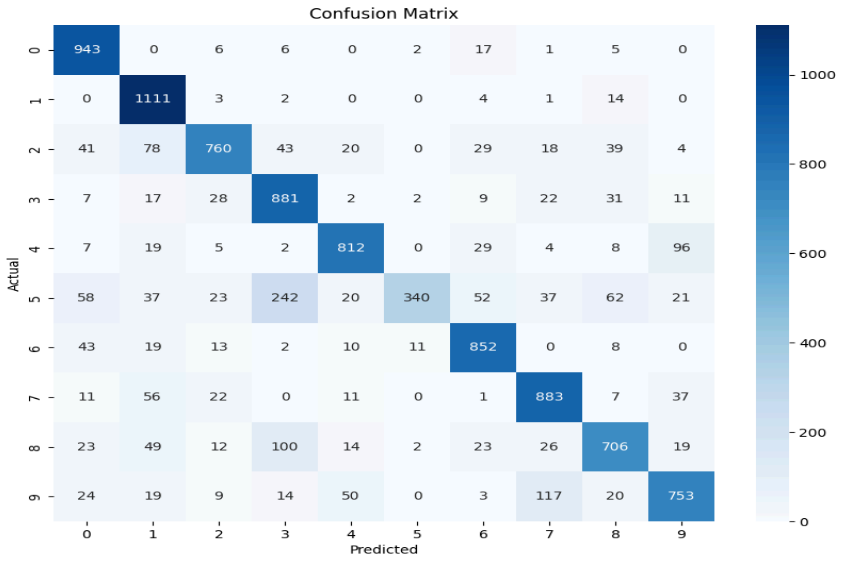
2. Varying Hidden Layer Size

Hidden Units = 128: Training loss decreased to 0.340 in 55.06 seconds.
Hidden Units = 256: Training loss decreased to 0.336 in 61.02 seconds.
Hidden Units = 512: Training loss decreased to 0.332 in 69.42 seconds.
Observation: Larger hidden layers led to slightly better training loss but required more time to train.

3. Experimenting with Learning Rates

Learning Rate = 1e-2: Training loss decreased rapidly, finishing in 66.29 seconds.
Learning Rate = 1e-3: Training loss decreased steadily, finishing in 74.50 seconds.
Learning Rate = 1e-4: Training loss decreased very slowly, finishing in 64.89 seconds.
Observation: A higher learning rate (1e-2) led to faster convergence but may risk overshooting the optimal solution. A lower learning rate (1e-4) was too slow to converge.

4. Confusion Matrix



	0	1	2	3	4	5	6	7	8	9
0	943	0	6	6	0	2	17	1	5	0
1	0	1111	3	2	0	0	4	1	14	0
2	41	78	760	43	20	0	29	18	39	4
3	7	17	28	881	2	2	9	22	31	11
4	7	19	5	2	812	0	29	4	8	96
5	58	37	23	242	20	340	52	37	62	21
6	43	19	13	2	10	11	852	0	8	0
7	11	56	22	0	11	0	1	883	7	37
8	23	49	12	100	14	2	23	26	706	19
9	24	19	9	14	50	0	3	117	20	753

The confusion matrix showed an 80.41% test accuracy for the MLP model.

5. Single-Layer Perceptron (SLP)

The SLP achieved a training loss of 0.3855 after 60 epochs, finishing in 81.04 seconds.
Performance was inferior to the MLP, as expected, due to the lack of hidden layers.

6. CPU vs. GPU Performance

CPU Execution Time: ~60-90 seconds per training run.
GPU Execution Time: ~40-75 seconds per training run.
Observation: For small models like the ones used in this assignment, the CPU performed adequately. GPU acceleration would be more beneficial for larger models.

Conclusion

The MLP achieved better performance than the SLP, as expected. Hyperparameter tuning (epochs, hidden layer size, learning rate) significantly impacted training time and model performance. The confusion matrix provided insights into the model's strengths and weaknesses.