Arshdeep Bhangu
Prof. Kremer
CIS*3700
2025/03/28

CIS*3700 Assignment #5

Report Overview:

This report digs into three tasks involving neural network training on a dataset, all run on a T4 GPU. First, we look at mini-batch learning with 1000 mini-batches compared to iteration and epoch learning, then test different mini-batch sizes (100 to 6000) to find the best one. Second, we explore a network with a single hidden layer (784 units) and tweak the number of hidden units (1/8x to 8x) to see how it impacts accuracy and loss on training and testing sets, recommending the best size. Third, we take a network with 623,290 parameters, redesign it into three two-hidden-layer setups with similar parameter counts, and compare their training and testing loss over time to pick the best configuration. Results include accuracy, loss, and time metrics, with graphs to visualize the results better.

TASK 1:

The notebook implements a mini-batch learning approach, dividing the dataset into 1000 mini-batches (batch size 60), and compares it to iteration learning (batch size 1) and epoch learning (batch size 60,000) using a T4 GPU runtime. Mini-batch (1000) completed 25,000 updates in 309.21s, achieving a test loss of 0.0470 and accuracy of 92.13%. Iteration learning, with 1,500,000 updates, took 16,720.94s, yielding a loss of 0.0502 and accuracy of 91.50%. Epoch learning, with only 25 updates, was fastest at 21.41s but performed poorly, with a loss of 0.3201 and accuracy of 20.78%.

Additional mini-batch sizes were tested: Mini-batch (100) took 42.53s (87.44% accuracy), Mini-batch (500) took 157.03s (91.23% accuracy), Mini-batch (2000) took 579.04s (92.63% accuracy), and Mini-batch (6000) took 1693.10s (95.64% accuracy). Mini-batch learning balances speed and accuracy better than iteration (slow, marginal gains) and epoch (fast, poor performance) methods. Among mini-batch sizes, 6000 offers the best accuracy (95.64%) and lowest loss (0.0251), though it’s time-intensive. For a practical trade-off, Mini-batch (1000) is recommended, providing high accuracy (92.13%) and reasonable speed (309.21s).

Attached are figures for better comparison:

Arshdeep Bhangu
Prof. Kremer
CIS*3700
2025/03/28

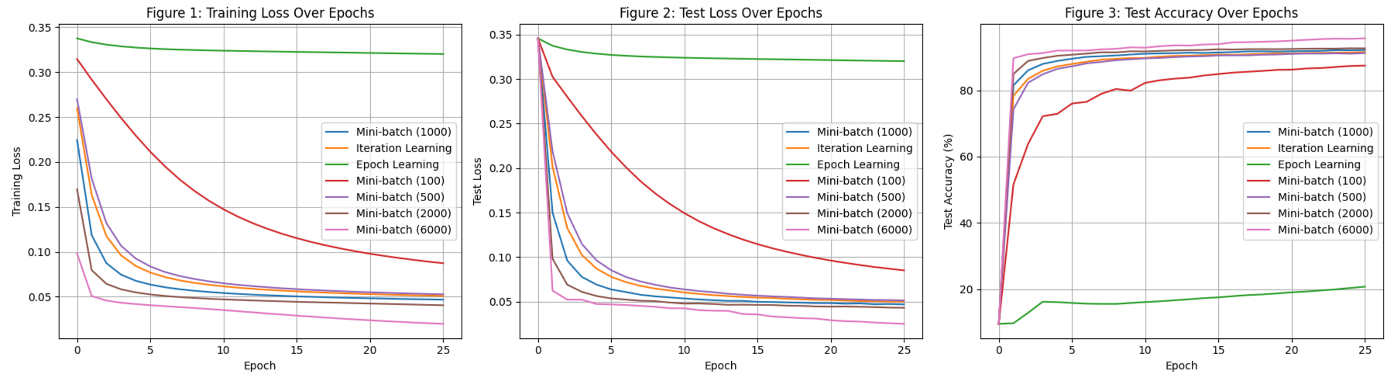


Figure 1-3: Comparing Training Loss, Test Loss, and Test Accuracy Over 25 Epochs for Mini-batch (100, 500, 1000, 2000, 6000), Iteration, and Epoch Learning Methods

---

TASK 2:

The notebook implements a neural network with a single hidden layer of 784 units (28x28) and evaluates performance across networks with 1/8x (98), 1/4x (196), 1/2x (392), 1x (784), 2x (1568), 4x (3136), and 8x (6272) hidden units. Performance is assessed via training and testing accuracy, loss over time, and final confusion matrices.

Results show that 1/8x (98 units) achieves the highest testing accuracy (92.79%) and lowest testing loss (0.0428), with training accuracy at 92.72% and loss at 0.0433. As hidden units increase, performance plateaus or declines: 1x (784) yields 92.05% testing accuracy and 0.0474 loss, while 8x (6272) drops to 91.10% accuracy and 0.0514 loss, indicating overfitting or decreasing returns. Loss trends over time likely converge fastest for smaller networks, with larger ones showing slower improvement and higher final losses on testing data. Confusion matrices would reflect better class distinction with fewer units, degrading as complexity increases.

For this dataset, 98 hidden units (1/8x) are recommended, offering the best balance of high accuracy, low loss, and computational efficiency.

Figures below provide a better understanding:

Arshdeep Bhangu
Prof. Kremer
CIS*3700
2025/03/28

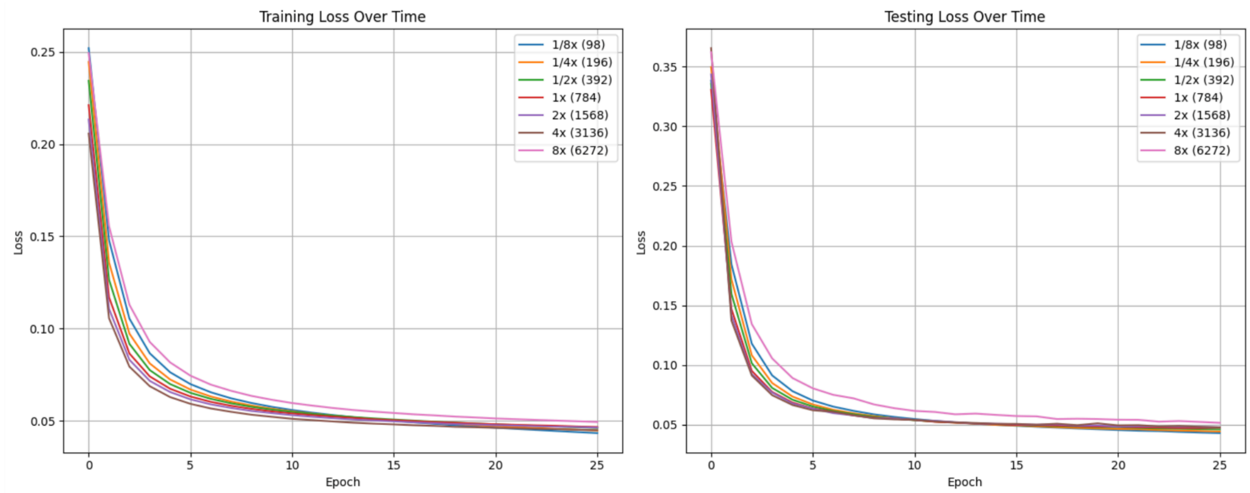
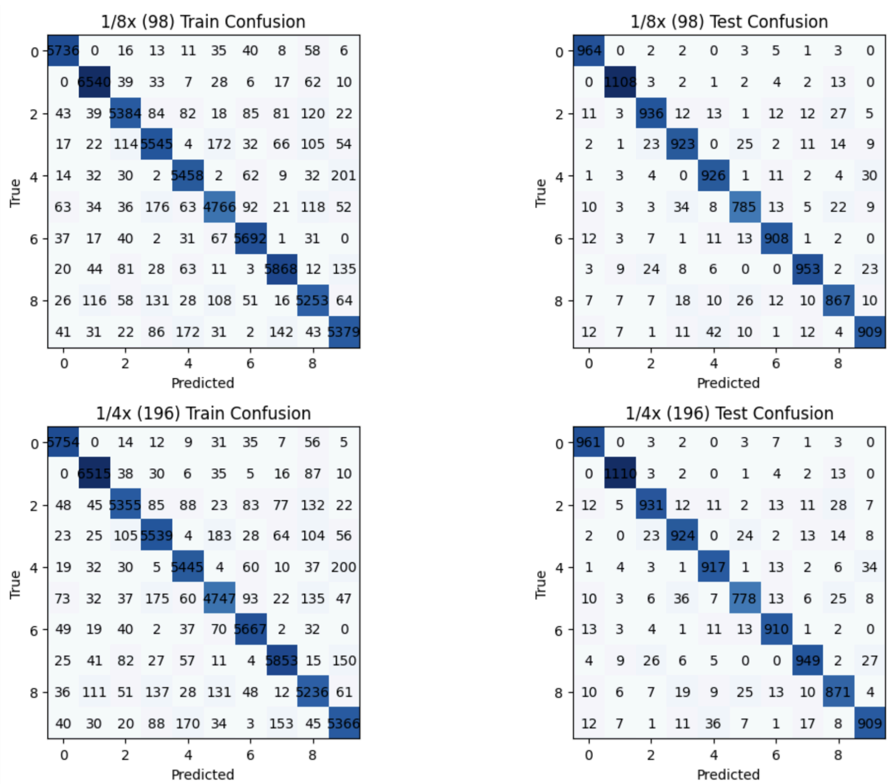
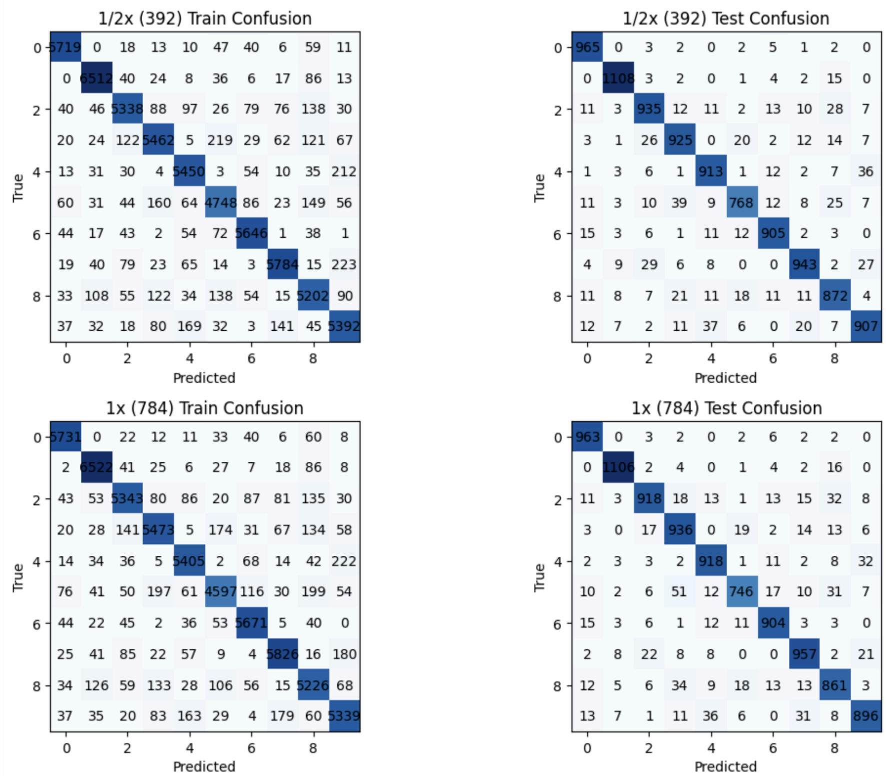
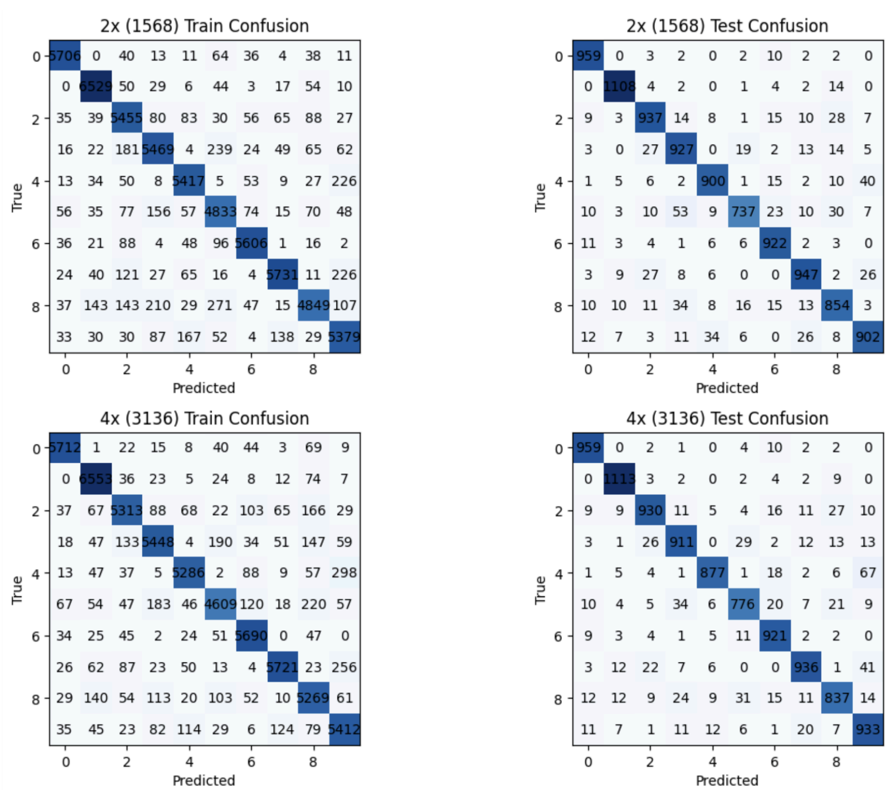


Figure 1-2: Training and Testing Loss Over 25 Epochs for Networks with Different Hidden Units (1/8x: 98, 1/4x: 196, 1/2x: 392, 1x: 784, 2x: 1568, 4x: 3136, 8x: 6272)

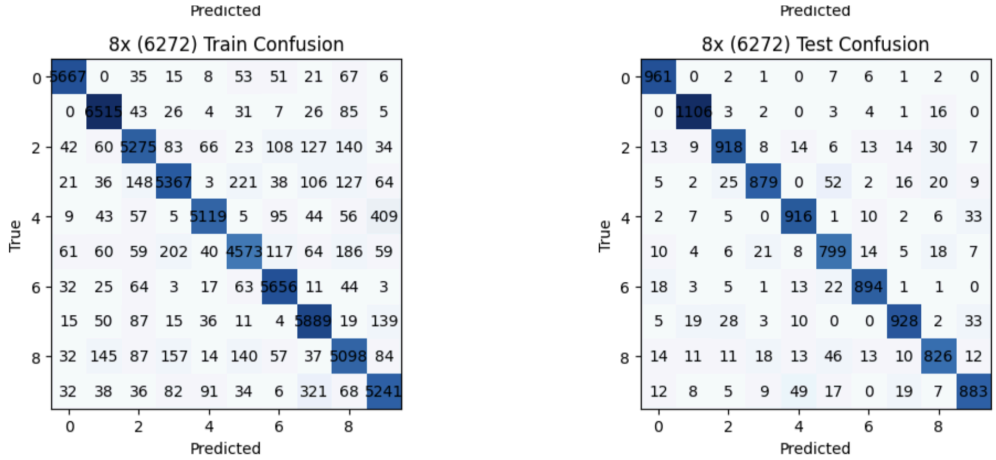


Arshdeep Bhangu
Prof. Kremer
CIS*3700
2025/03/28





Arshdeep Bhangu
Prof. Kremer
CIS*3700
2025/03/28



Training and Testing Confusion Matrices for Networks with Different Hidden Units (1/8x: 98, 1/4x: 196, 1/2x: 392, 1x: 784, 2x: 1568, 4x: 3136, 8x: 6272)

TASK 3

The original network has 784x784 input-to-hidden connections, 784x10 hidden-to-output connections, 784 hidden biases, and 10 output biases, totaling 623,290 parameters. Three alternative networks with two hidden layers (H1 and H2) were designed to have a similar number of parameters: Network 1 (H1 = 490, H2 = 510, ~640,170 params), Network 2 (H1 = 400, H2 = 600, ~560,610 params), and Network 3 (H1 = 600, H2 = 400, ~615,410 params).

Training and testing loss over 25 epochs were analyzed. All networks show a sharp initial drop in training loss, stabilizing below 0.05 by epoch 25, with Network 3 converging slightly faster. Testing loss follows a similar trend, starting higher (0.30 – 0.35) and dropping to around 0.05, though Network 1 exhibits more variance initially. Loss trends suggest all networks achieve comparable convergence, with Network 3 showing the smoothest testing loss curve.

Given the similar parameter counts and loss behavior, Network 3 (H1 = 600, H2 = 400) is recommended for its stable convergence and slightly fewer parameters (615,410), balancing efficiency and performance.

Arshdeep Bhangu
Prof. Kremer
CIS*3700
2025/03/28

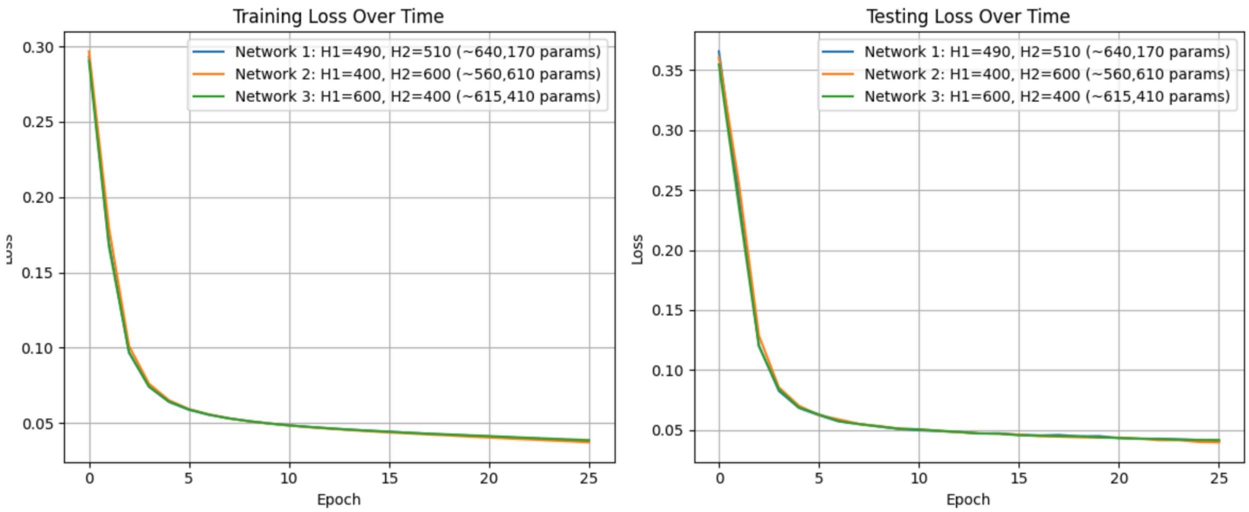


Figure 1-2: Training and Testing Loss Over 25 Epochs for Two-Hidden-Layer Networks (Network 1: H1=490, H2=510; Network 2: H1=400, H2=600; Network 3: H1=600, H2=400)