CARNEGIE MELLON UNIVERSITY 07-280

HOMEWORK 7

DUE: Thursday, Mar. 12, 2026

https://www.cs.cmu.edu/~07280

INSTRUCTIONS

- **Format:** Use the provided LaTeX template to write your answers in the appropriate locations within the *.tex files and then compile a pdf for submission. We try to mark these areas with STUDENT SOLUTION HERE comments. Make sure that you don't change the size or location of any of the answer boxes and that your answers are within the dedicated regions for each question/part. If you do not follow this format, we may deduct points.

  You may also digitally annotate the pdf. Illegible handwriting will lead to lost points. However, we suggest that try to do at least some of your work directly in LaTeX.

- **How to submit written component:** Submit to Gradescope a pdf with your answers. Again, make sure your answer boxes are aligned with the original pdf template.

- **How to submit programming component:** See Programming combonent for details on how to submit to the Gradescope autograder.

- **Policy:** See the course website for homework policies, including late policy, and academic integrity policies.

- **Proofs and Derivations:** For full credit, you must clearly show all nontrivial steps. Explicit reasons for each step are not required but can be helpful, especially if the step is not obvious.

Name	
Andrew ID	
Hours to complete all components (nearest hour)	

1 Programming Experiments

This part of the written assignment should be done after you complete the programming.

For all of the following experiments use the following baseline configuration:

- Batch size: 64
- Training data size: 30,000 points
- Validation data size: 5,000 points
- Network: Only one hidden layer (size to be specified below)
- Number of epochs: 20
- Learning rate: (specified below)

1. **[4 pts]** Display the validation data confusion matrix after training the 64-neuron network with learning rate 1.0 for **20 epochs**.

   Confusion matrix

   Which digit is misclassified the most? (Break any ties arbitrarily.)

   For that digit, what digit is it most commonly confused as? (Break any ties arbitrarily.)

   What do the diagonal elements represent?

   Diagonal elements?

2. **[9 pts] Tune learning rate**

   Train your network with 64 neurons in the hidden layer and try learning rates 10.0, 1.0, and 0.1.

   On the following pages, include plots of your train/val loss and accuracy over epochs for each learning rate. Also include the visualized first layer weights for each learning rate.

   The `plot_training_history` function in `util.py` will be helpful.

   For each learning rate, does it make sense to invest time in continuing to train for more epochs, e.g. epochs 21-40? (Assume we don't change anything before we continue.) Why or why not? Be sure to comment on the results that you are providing on the following pages.

More epochs? Why or why not?: Learning rate 10

◯ Yes ◯ No

More epochs? Why or why not?: Learning rate 1.0

◯ Yes ◯ No

More epochs? Why or why not?: Learning rate 0.1

◯ Yes ◯ No

Loss and Accuracy: Learning rate 10.0

Loss and Accuracy: Learning rate 1.0

Loss and Accuracy: Learning rate 0.1

Weights: Learning rate 10.0

Weights: Learning rate 1.0

Weights: Learning rate 0.1

3. **[6 pts] Experiment with width of hidden layer**

   Train your network with learning rate 0.1 and experiment with different numbers of neurons in the hidden layer. (Still just one hidden layer.) Specifically, try 4, 16, and 64 neurons.

   On the following pages, include plots of your train/val loss and accuracy over epochs for each network architecture. Also include the visualized first layer weights for each.

   Given these results after training for 20 epochs, which architecture performs the best and which performs the worst?

   Best: ◯ 4 hidden neurons ◯ 16 hidden neurons ◯ 64 hidden neurons

   Worst: ◯ 4 hidden neurons ◯ 16 hidden neurons ◯ 64 hidden neurons

   What specifically did you use to determine which architecture is best?

   Specific criteria

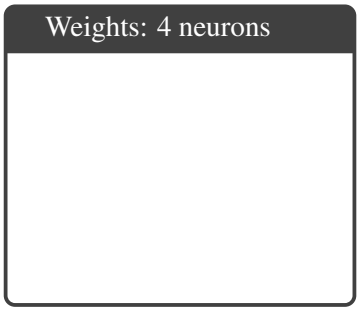
   Comment on how visualizing these first layer weights can help us understand the effect of the number of neurons on performance.

   Comment on weight visualization and number of neurons

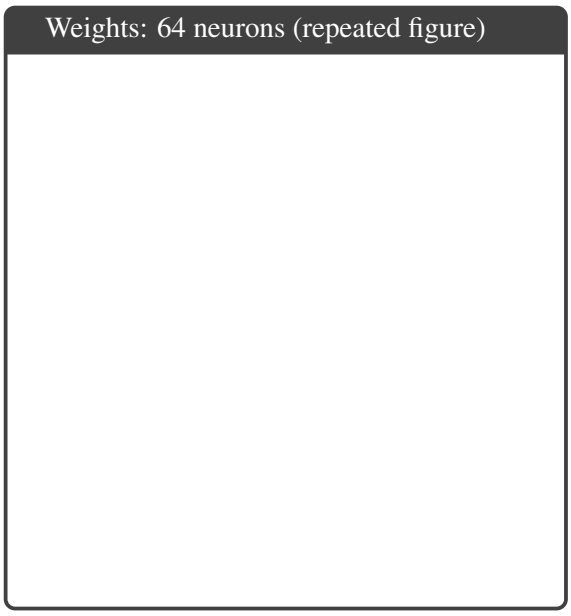
Loss and Accuracy: 4 neurons

Loss and Accuracy: 16 neurons

Loss and Accuracy: 64 neurons (repeated figure)


Weights: 4 neurons


Weights: 16 neurons


Weights: 64 neurons (repeated figure)

2 Collaboration Policy

After you have completed all other components of this assignment, report your answers to the following collaboration questions.

1. Did you receive any help whatsoever from anyone in solving this assignment? If so, include full details including names of people who helped you and the exact nature of help you received.

Your Answer

2. Did you give any help whatsoever to anyone in solving this assignment? If so, include full details including names of people you helped and the exact nature of help you offered.

Your Answer

3. Did you find or come across code that implements any part of this assignment? If so, include full details including the source of the code and how you used it in the assignment.

Your Answer