Final Exam, 2024
100 pt

Download the MNIST database: a dataset of 60k training images and 10k test images, each of size 28×28 (grayscale), illustrating handwritten digits from 0 to 9. Each pixel is an integer in the range {0, 1, ... , 255}; these are usually rescaled to [0,1] to represent pixel intensity.

1. (40 pt) Train a fully connected neural network (NN) classifier with two 50-node hidden layers: $784 \times 50 \times 50 \times 10$ on the MNIST dataset. Experiment with various hyperparameter settings and architecture choices such as hidden layer activation functions and the optimizer type to maximize the classification accuracy. Report all the relevant details of your final NN model, including the weight and bias initialization scheme. Plot the training and test set classification accuracy as a function of the number of epochs for two typical runs. If early stopping was used, indicate it with straight vertical lines on the graph.

2. (40 pt) Train a convolutional neural network (CNN) classifier on the MNIST dataset. Take the LeNet5 architecture (see lecture notes) as the starting point, and tinker with it to maximize the classification accuracy. Report all the relevant details of your final CNN model, including the weight and bias initialization scheme. Plot the training and test set classification accuracy as a function of the number of epochs for two typical runs. If early stopping was used, indicate it with straight vertical lines on the graph.

3. (20 pt) Discuss the relative advantages and drawbacks of the fully connected NN and CNN. Which model was easier to train and more robust to the hyperparameter and architecture choices? Which model required fewer parameters? Which model yielded superior performance on the test set (averaged over the two runs)? Note: you should be able to obtain >90% test classification accuracy in both models.

You are welcome to use any software environment to carry out this assignment.