CSCI 5922 Lab assignment 1 part 2

Jake Krol

February 2025

1 Experimental design

I will **study** the effect of model depth and width on test accuracy for multiclass classification on MNIST. The experiment is **conducted** by comparing how decreasing and increasing model width (hidden layer dimensions) effects test accuracy, and how decreasing and increasing model depth (number of hidden layers) effects test accuracy. No other hyperparameters will be adjusted, and I will report test accuracy after training on 25 epochs. The baseline model will be part 1 of the lab assignment: a single hidden layer with 256 dimensions.

I **expect** the model to overfit and test accuracy to decrease when hidden layers become wider and/or the model becomes deeper with more layers. When hidden layers and/or their depth increase, I **expect** the model to underfit if the parameters are reduced too much, such as a single hidden layer with relatively low dimensions.

Width	Depth
256	1
256	2
512	1
128	1
512	2
1024	1

Table 1: **Hyperparameters**. A table of hyperparameters used to measure the effect of model capacity on test accuracy in MNIST. Width is the number of neurons per hidden layer. Depth is the total number of hidden layers in the network.

2 Results

Width	Depth	Test accuracy (%)	Total parameters
256	1	70.24 (baseline)	203,530
256	2	51.83	269,322
512	1	67.91	407,050
128	1	60.81	101,770
512	2	47.25	669,706
1024	1	67.48	814,090

Table 2: **Test accuracy and parameters for models of varying width and/or depth**. Test accuracy is reported, as percentage of images correctly classified, on MNIST for various model widths and depths. Also, the total model parameters is reported for each model.

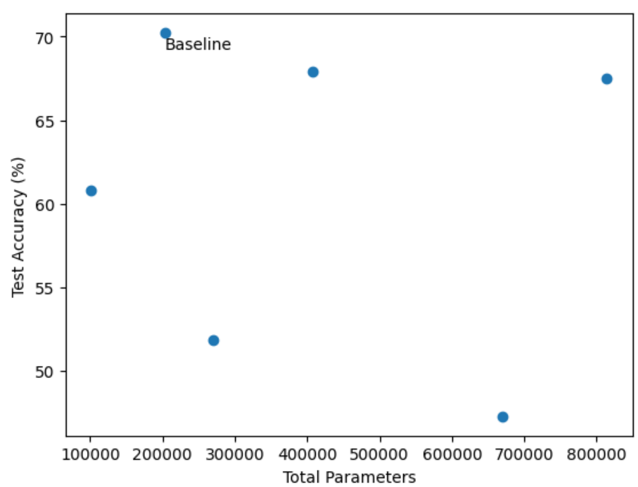

Figure 1: **Relationship between total parameters and accuracy.** The total parameters (horizontal axis) is non-linearly related to the test accuracy (vertical axis).

3 Analysis

Experiments showed the **trends** that 1) test accuracy generally decreased with more parameters and that 2) models with roughly the same number of parameters but different widths/depths can perform differently.

The **original hypothesis** that increasing and/or decreasing parameters would decrease test performance is supported by lower test accuracy performance when depth/width was increased/decreased. Table 1 shows that the highest performing model was the baseline model. Where, the parameters for the baseline model were provided by an external source which may have tuned for depth/width beforehand to achieve optimal results. Reducing parameters, (such as 1 layer, 128 width) lowered accuracy about 10%, and increasing parameters had a variable effect ranging from a 3% to 23% test accuracy decrease (Figure 1).

Figure 2: **Extra credit**: screenshot of set_hp, backward, TrainMLP, and main function modifications to support L1 and L2 regularization

4 Extra credit

I used **L1 and L2 regularization on a "wide" architecture from my experiment** (hidden layer dimension of 512), and the **test accuracy decreased dramatically**: 11.25% for L1 and 9.75% for L2 compared to 67.91% without regularization. Likely, my implementation was incorrect, or the lambda parameter (weight of regularization term) needs to be tuned to a better value. Since there are many parameters in a fully connected neural network, maybe a very low lambda value is necessary.