

Homework 03

Brian Henderson

11/27/17

1 Part One

Using tien.py, complete a total of 10 experiments varying the number of hidden units in three layers. For these experiments, the auto-encoder uses 400 for the pre-training epochs and 50 for the fine-tuning epochs. Note** These experiments were meant to use "Our Dataset," but I was unable to get the auto-encoder to use it, so these experiments were run using the MNIST dataset.

1.1 Results

Table 1: Testing accuracies of various hidden unit combinations

Neuron Combination	Final Testing Accuracy
$32 \times 32 \times 32$	0.8023
$64 \times 32 \times 64$	0.8906
$32 \times 64 \times 128$	0.8553
$64 \times 128 \times 256$	0.8275
$32 \times 128 \times 256$	0.9024
$64 \times 256 \times 32$	0.8911
$32 \times 256 \times 512$	0.8497
$64 \times 256 \times 512$	0.9184
$256 \times 128 \times 512$	0.9402
$512 \times 128 \times 64$	0.9300

1.2 Analysis

The neuron combination of 256x128x512 had the highest final testing accuracy of 0.9402. There is a clear trend that the experiments that had more neurons for each layer had higher final testing accuracies. This is expected from my point of view, since the experiments did take longer which (in a proper architecture) typically indicates higher quality results. Since there is a noticeable jump in testing accuracy with the larger neurons (experiment 00: 0.8032 vs. experiment 08: 0.9402, difference: .1379), this leads me to think that this architecture does very well with larger neuron combinations.

2 Part Two

The neuron combination that had the best results from the previous set of experiments was 256x128x512. The learning rate used for part one was 1.0. The following experiments will modify that to 0.01, 0.001, and 0.0001.

2.1 Results

Table 2: Testing accuracies for various learning rates using a constant neuron combination of 256x128x512

Learning Rate	Final Testing Accuracy
0.01	.6615
0.001	.2259
0.0001	.1123

2.2 Analysis

There is a significant decrease in the final accuracy rating with the smaller learning rates. This could be because of a number of different reasons, however since this architecture worked very well with a learning rate of 1.0, this leads me to possibly believe that this architecture was designed with that in respect. Also with a lower learning rate, this indicates that the neural network would take longer to train.

3 Input on "Our Dataset"

Although I was not able to use "Our Dataset" for these experiments, I do have some input on what I believe the results would have been if I did. If the experiments were run using "Our Dataset", I would hypothesize that the testing accuracies and ratings would be significantly lower than the ones that were produced in these experiments. Since the MNIST dataset is very professionally equipped and prepared, the auto-encoder had a much more successful time with it. "Our Dataset" was also much smaller than the MNIST dataset, leaving me to believe that the auto-encoder would not have been able to produce as quality output. I believe that the lower learning rate trend with the lower testing accuracy would still have occurred, and that the neuron combination of 256x128x512 would still have been the highest quality.