Deep Learning - Homework 2

Brendon Boldt

Oct/23/2017

1 Results

To test the hyperparameters of the model, we varied both base learning rate and the number of hidden units. The code used to run this experiment can be found in `hw2.py`.

Table 1: The table displays the accuracies resulting from the corresponding hidden layer size and learning rate. Note that "LR" stands for "Learning Rate".

Neurons/LR	$1 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-1}$	$1 \cdot 10^{0}$
16	0.0957	0.1574	0.3724	0.7197	0.6781	0.3333
32	0.0684	0.1577	0.4410	0.8540	0.7126	0.1032
64	0.1577	0.2990	0.7568	0.8769	0.4855	0.0974
128	0.1626	0.4422	0.8126	0.8799	0.0987	0.1032
256	0.2174	0.6067	0.8362	0.8725	0.1009	0.0980
512	0.2867	0.7034	0.8654	0.6754	0.0974	0.1032
1024	0.4777	0.7645	0.8640	0.0892	0.0980	0.1028
2048	0.5950	0.8063	0.8056	0.1135	0.1009	0.0892
4096	0.6625	0.8105	0.7913	0.1009	0.0958	0.1009

In order to better visualize the data, we have included a graph (Figure 1) of accuracy vs. hidden unit count.

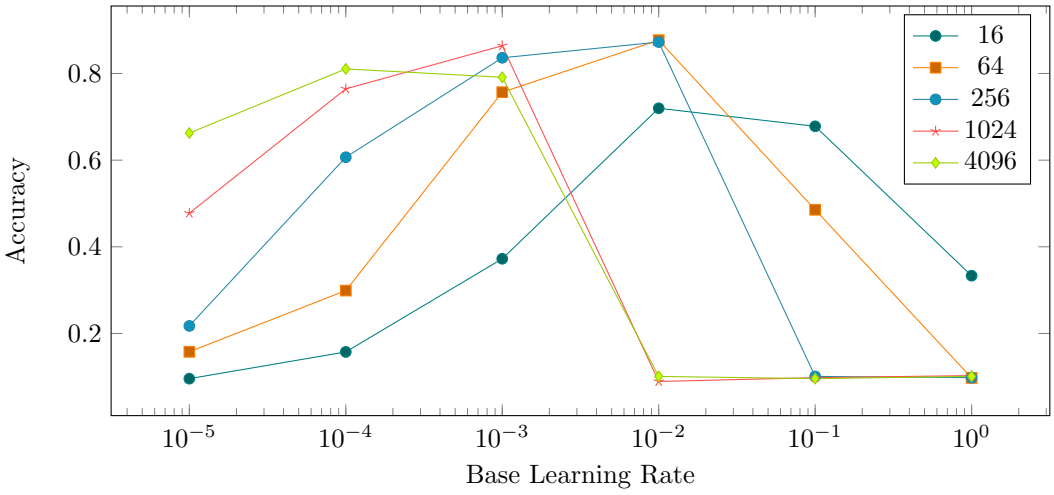

Figure 1: Accuracies achieved by hidden layer sizes for different learning rates

It is clear from the data that smaller hidden layer sizes performed better with higher learning rates than did larger networks. One explanation for this is that the smaller nets tend are simpler (due to containing

fewer hidden units); the smaller models are simpler in the sense that there is less fine tuning of connection weights required to effectively make use of the hidden units. The smaller learning rates resulted in lower accuracies likely because the simpler networks had not been trained enough.

With a larger hidden unit count, the connection weights need to be very precise in order to balance out the connections amongst the hidden units. When this precision is achieved with the lower learning rates, the models are able to perform well on the test data. An explanation to why this small amount of training was enough for the larger networks but the smaller ones might be that the shear number of neurons made it possible model the data with little training. But in this case, the network would likely still benefit from a longer training time.

In the future, it would be important to include a validation set in order to determine if the models are overfitting to the training data as this risk increases as the number of hidden unit gets larger.