Classification Performance of Individual Networks vs Ensemble

Jer Moore a02082167

September 2019

1 Architecture and Results

Each individual net was trained initially for 20 epochs, with the learning rate of 3.0, and the mini-batch size of 20. I did this because I was mostly interested in if the size of each layer, or the number of hidden layers, played a part in how successful a particular net was. Each initial layer consisted of 784 nodes and end with 10, but from there the individual nets began to change. From the most layers to least and with their final success percent:

$784 \times 500 \times 300 \times 60 \times 30 \times 10$	95%
$784 \times 500 \times 60 \times 30 \times 10$	93%
$784 \times 800 \times 10$	64%
$784 \times 80 \times 10$	86%
$784 \times 60 \times 10$	86%
Ensemble	97%

2 Interpretation

The best individual network was the one with the greatest number of hidden layers, and it appears that for every net with fewer layers the accuracy of the network dropped. The size of a large individual hidden layer doesn't seem to have a desirable effect on the network, see the large dip at the 784,800,10 network accuracy vs the 784,80,10 and 784,60,10 networks. I was under the impression from the paper we read with this assignment that some kind of internal representation would help the overall performance of the network, but that does not seem to be the case here.
The ensemble did better than any individual network, but not by much. The overall gain from 95% to 97% may not be worth the trouble of training four more networks. I did not consider any measure of how long it took to train each individual network in this assignment. It is of no use to have a higher accuracy that comes at to great a computational cost.