CS189-hw6 Report

Eileen Li, Minyoon Jung

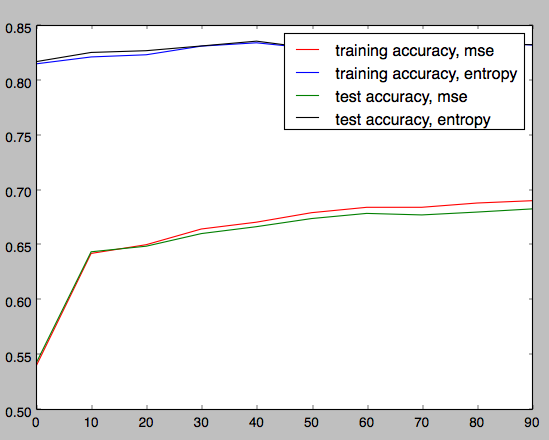
See README for more information.

Resources:

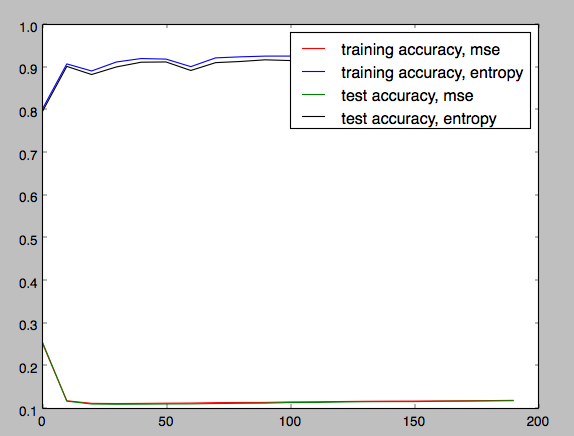
Relied heavily on the Caltech lectures as well as class notes. Probably read some articles online too, but none that really stood out.

Single-layered:

It was difficult for us to lower our error rate for the mean squared error. We played with the learning rate, with the weight initialization, with the preprocessing methods, and the best we could manage was around 30% error. Below is a graph of the accuracy. There are four lines, representing the test and training sets using mean squared or entropy.



For cross-entropy loss function, we got the error rate down to around 7% error on the training data and 8% on the test data using a learning rate of 0.01, and decreasing this rate by a factor of 1/n^0.5 where n is the number of epochs. If we had more time, we could’ve ran more epochs and achieved lower error rates. The error rates for cross-entropy were still decreasing (very slowly) when we stopped at 200 epochs. The running time for 200 epochs was 2.4 hours. Below is a graph of the accuracy. Notice that the mean squared had really results using this learning rate. Please just ignore that and refer to the first graph for mean-squared.



Multi layered:

We implemented multi-layered neutral networks with the Caltech lectures as reference. I think we still need to tweak with the parameters more, because our error rates are not low enough (not even close). But trust me when I say we spent countless hours trying to fix this. Due to time constraints, we were only able to run about 30 epochs and generate a graph from this. Even though we were not successful in our implementation, both of us learned a tremendous amount. It’s a shame our effort could not be translated into results.

