Homework 2

Raghav Sreeram, 915163498

Part 1

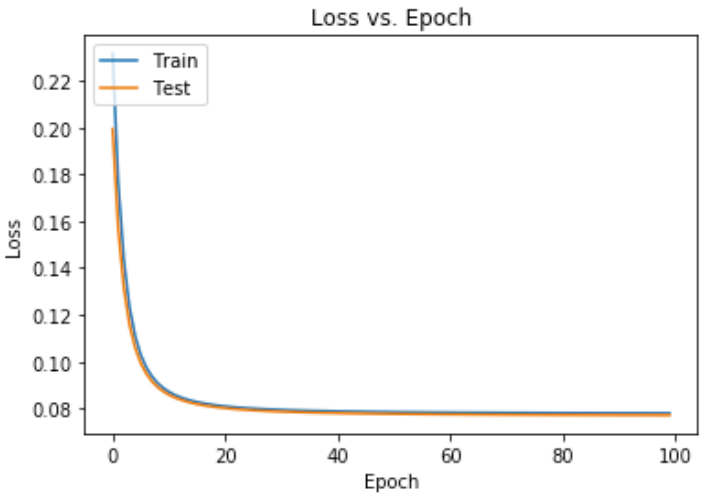
1. Used **Isolation Forest** and **LOF**
   a. Are there outliers?
      - Isolation Forest gives **149 outliers**
      - LOF gives **149 outliers**
   b. Do methods Agree and why?
      - **104 outliers agree and 90 don't agree**
      - This is largely due to the different ideas and structure of each algorithm. The Isolation Forest builds the random decision trees and picks outliers as the ones that require fewest levels. Whereas, LOF is based on density with respect to its neighbors. LOF does not does not depend on predict and is one of the key differences.
   c. What assumptions are made?
      a. Both methods have similar definition for an outlier and try to find points with different features than rest of the other ones. Isolation Forest aims to separate outliers under the assumption it takes less effort to separate them from the group. LOF works by assuming outliers have far less neighbors than non outliers.

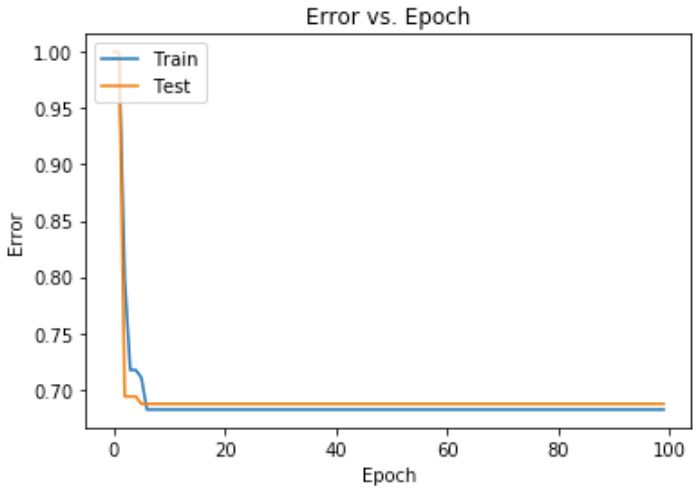
Moving forward, I used data after removing outliers from Isolation Forest.

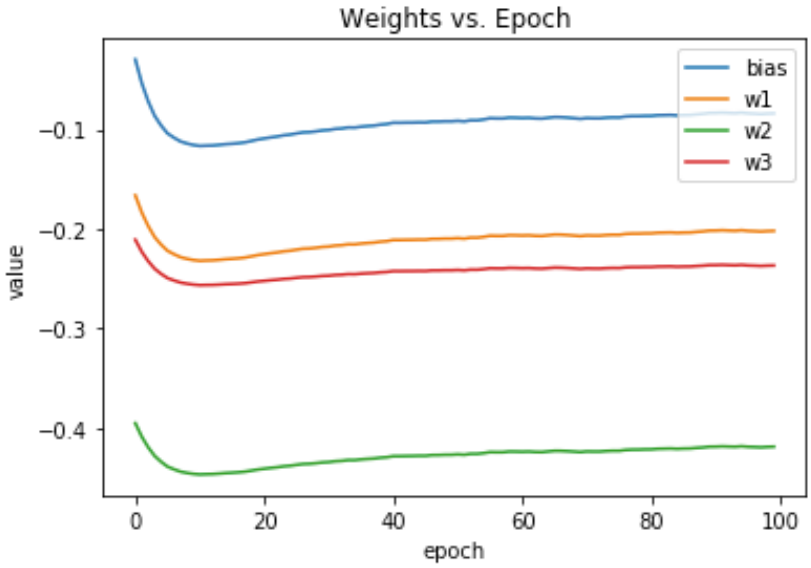
Part 2

Construct 4-layer ANN with Sigmoid Activations and MSE loss function. Two hidden layers with Three nodes each. Provide for class “CYT” weight values per iteration for last layer, and training and test error per iteration.

Model uses batch size 1 for 100 epochs, and uses sgd, mse.
Included Loss vs. Epoch because it shows model is learning and converges around 20 epochs.





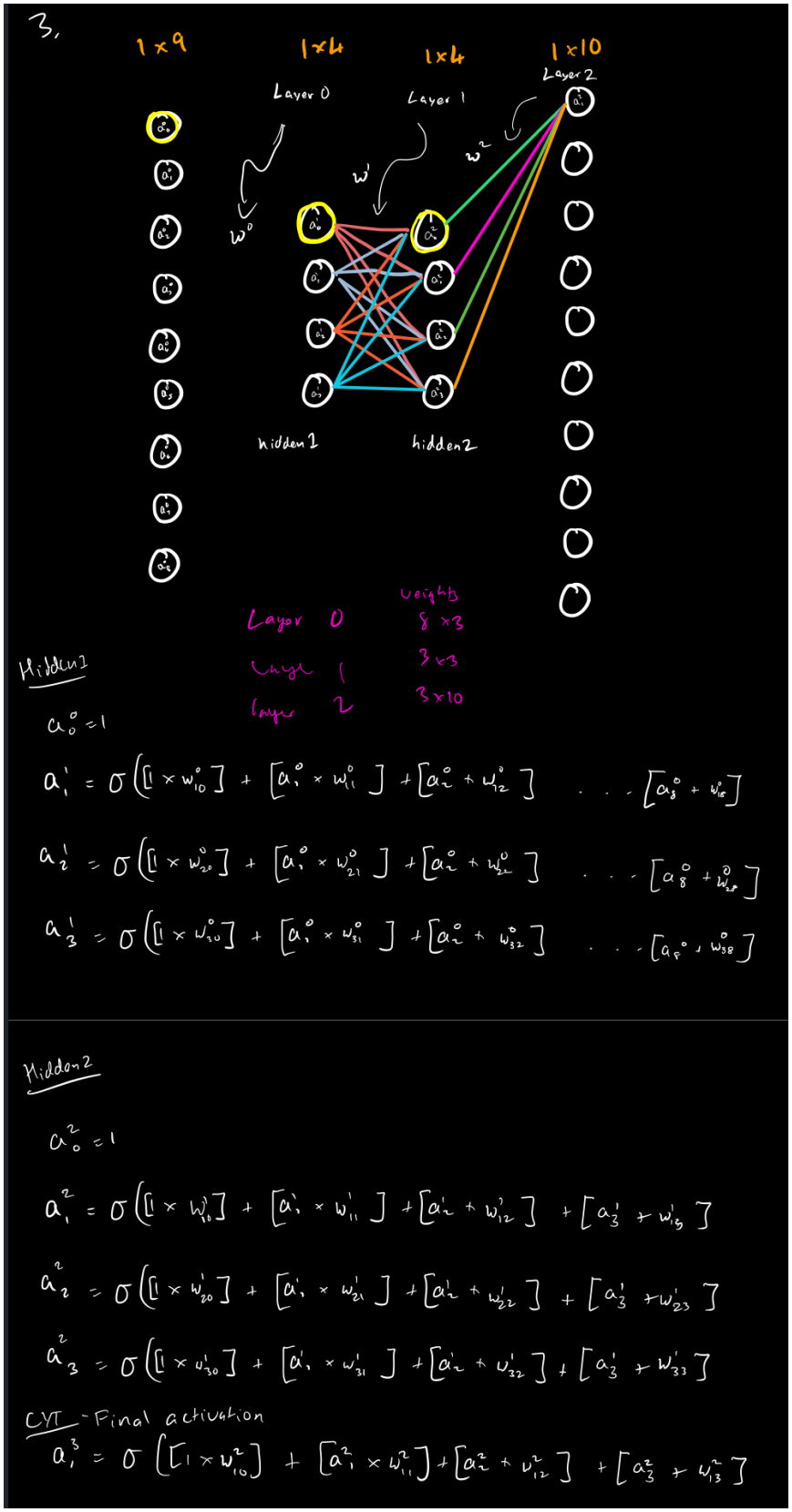


Part 3

ANN with all 1484 samples.

Training error: 31.19%

Final Activation function formula for “CYT”

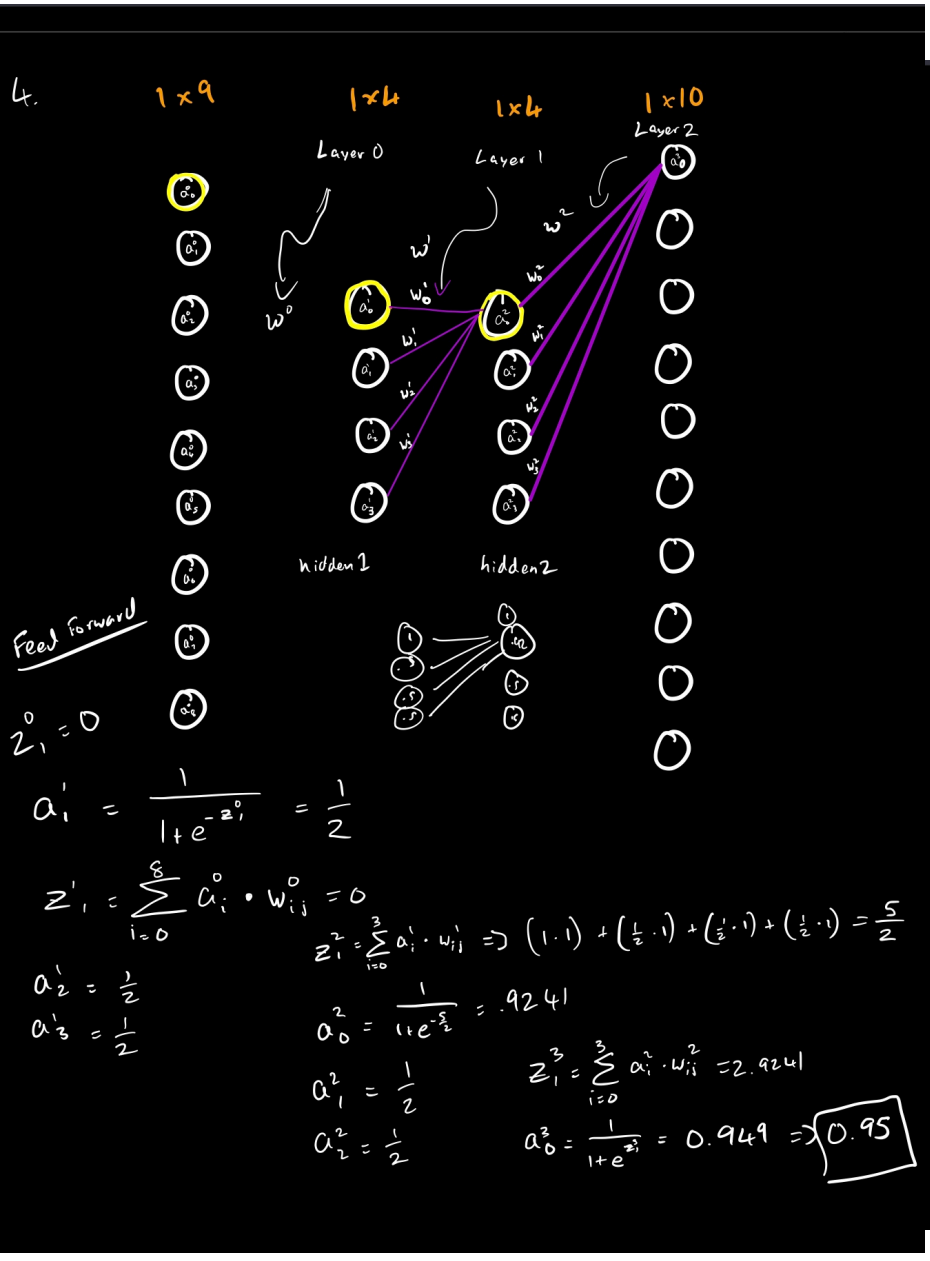


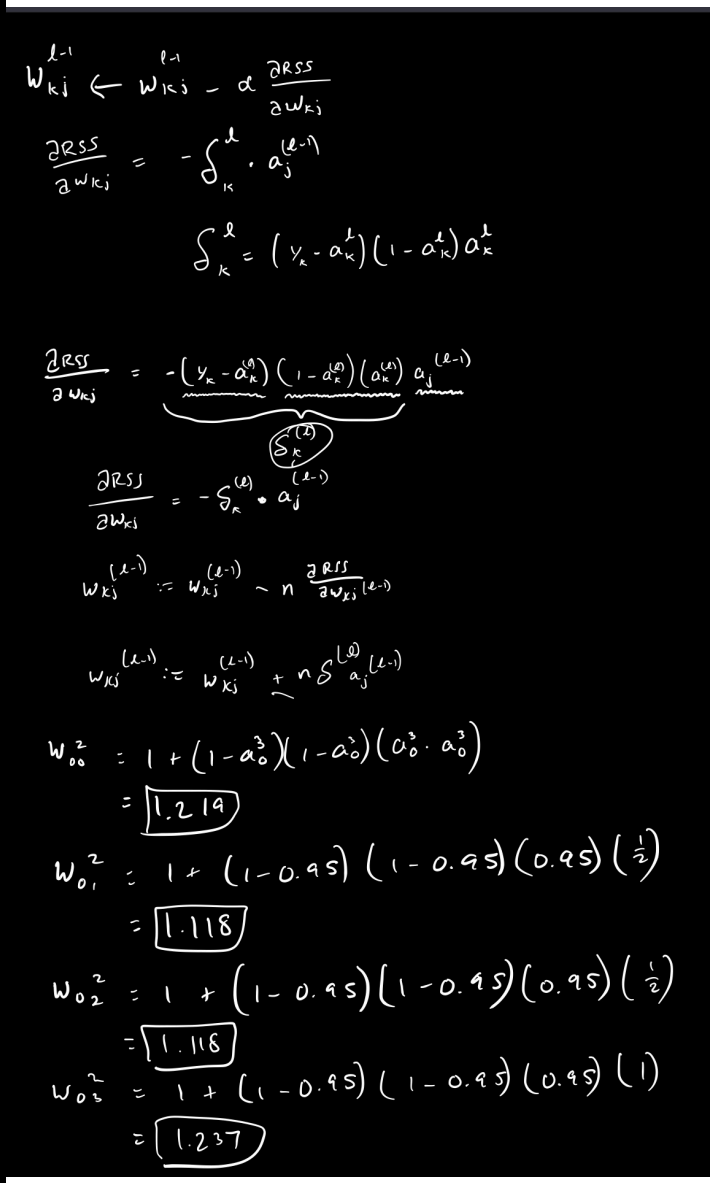
Part 4

Code is at end of Part 3

Calculate first round of weight updates with back-propagation for two final layers.

Set learning rate to 1. Didn't have time to compare with





Part 5

Grid Search

	3 Nodes	6 Nodes	9 Nodes	12 Nodes
1 Layer	71.15%	68.8%	68.73%	67.72%
2 Layer	68.8%	68.8%	68.8%	68.8%
3 Layer	68.8%	71.09%	68.8%	68.8%

Optimal configuration seems to be 1 layer 12 Nodes with an error of 67.72%
When Layer is constant, increasing Nodes seems to be reducing error and being being beneficial.
When number of Nodes is constant, increasing layers seems to reduce error for 3 Nodes but for others it is relatively arbitrary.

Part 6

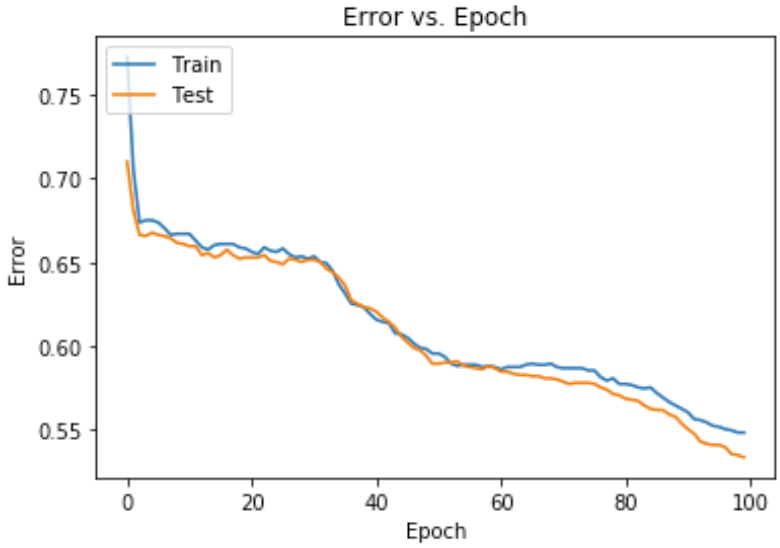
code for this is in end of Part5
Model predicts CYT.

Part 7

Changed to ReLU, Softmax, and Cross-Entropy

	3 Nodes	6 Nodes	9 Nodes	12 Nodes
1 Layer	51.01%	46.29%	43.86%	47.64%
2 Layer	55.25%	44.14%	45.55%	44.67%
3 Layer	57.344%	46.14%	45.41%	41.71%

This model works a lot better than previous models and can be seen with lower errors in the graph. The optimal configuration for this model seems to be 1 layer with 6 nodes.


Error vs. Epoch
Train
Test
0.75
0.70
0.65
0.60
0.55
Error
0
20
40
60
80
100
Epoch