

Name: Serena Yang  
Course: CSC 578  
Assignment: HW 3

### Part A

Complete and the output matched the one posted.

### Part B

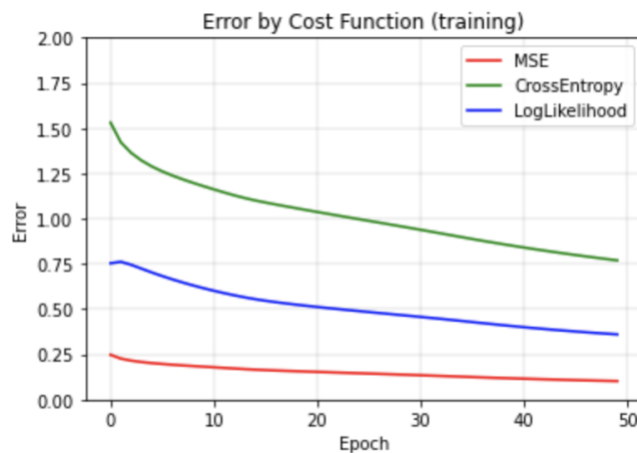
Complete and the output matched the one posted, except the decimal are different, the results are the same.

### Part C

Complete and the output matched the one posted, except the decimal are different, the results are the same.

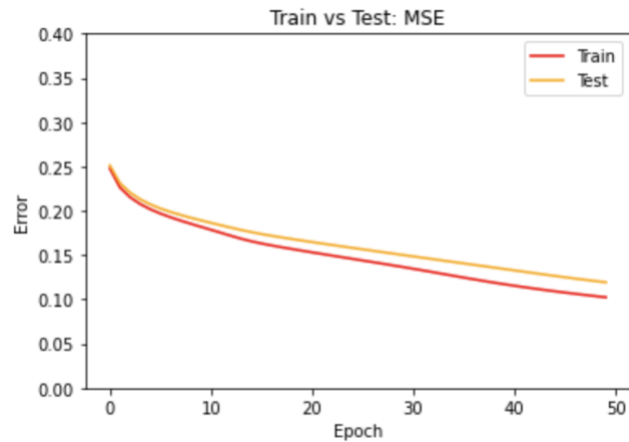
### Part D

Using matplotlib library to visualize the error from each cost function of the “net4” training network of 3 layers and the size [4,20,3], and 50 epochs with eta = 0.1 and the minibatch size = 5.

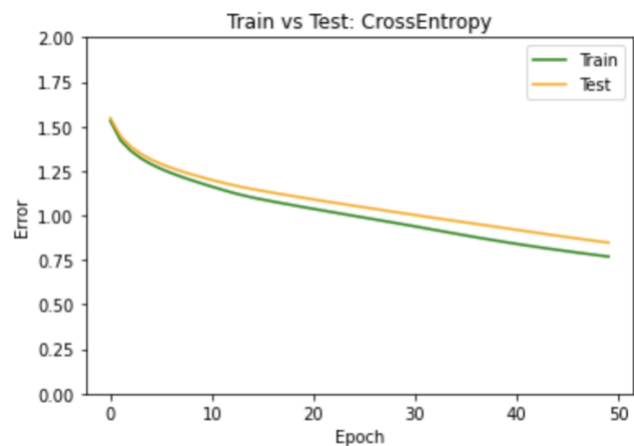


This plot shows the model errors on 70% of the original dataset and 50 epochs. The MSE performs the best from 0.25 for epoch 0 to about 0.1 for epoch 49. The Cross-Entropy drops rapidly between epoch 0 to epoch and ends with epoch 49 with a 0.75 error. The log-likelihood

has a little increase from epoch 0 to epoch 1; after that, the error result decreases slowly.

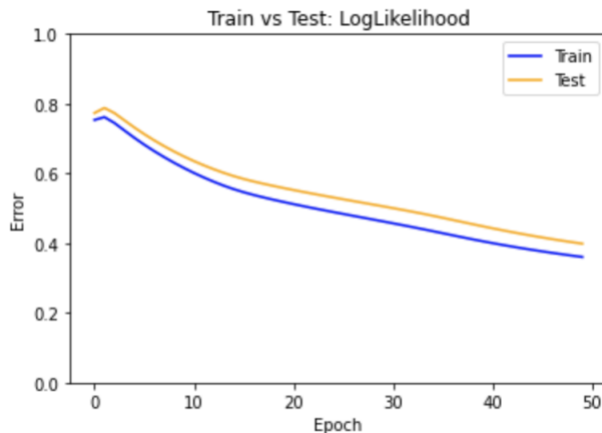


This plot shows the mean squared error(MSE) results of the network across samples calculated as the average squared differences between target values and activation outputs on the training dataset and test dataset from epoch 0 to epoch 49. The trend on the training dataset was from 0.25 to around 0.1, and the error for the test dataset was from 0.25 to 0.13. There is a overfit problem from epoch 20, and the overfitting problem will impact the result with the epoch number increase.



This plot shows the cross-entropy results on the training and test datasets from epoch 0 to epoch 49. The cross-entropy error result is the highest among these three evaluated functions. The difference between the training and test dataset gets bigger with the epoch increase. The

cross-entropy error result pattern is similar to MSE's.



This plot shows the Log-Likelihood results on the training dataset and the test dataset from epoch 0 to epoch 49. The difference between the training dataset and the test dataset has a very slight change with the epoch increase. From epoch 0 to epoch 20, the gap increased. After epoch 20, the gap stays the same with a slight difference.

## Part E

After the first two weeks of learning formula derivation and theories, this week's homework required me to apply what I learned to the assignment. I enjoy the way I see how the neural networks implement. I think the difficulty level is average, even though a few problems happened during this assignment. The challenge I met was during part 2, which was mortifying the functions. After I read the default codes, it took me a little while to tried myself to work on the neural networks, especially for the `backprop()` function and activations part. In future work, I need to spend more time on books to learn deep about the theories and how neural networks work.