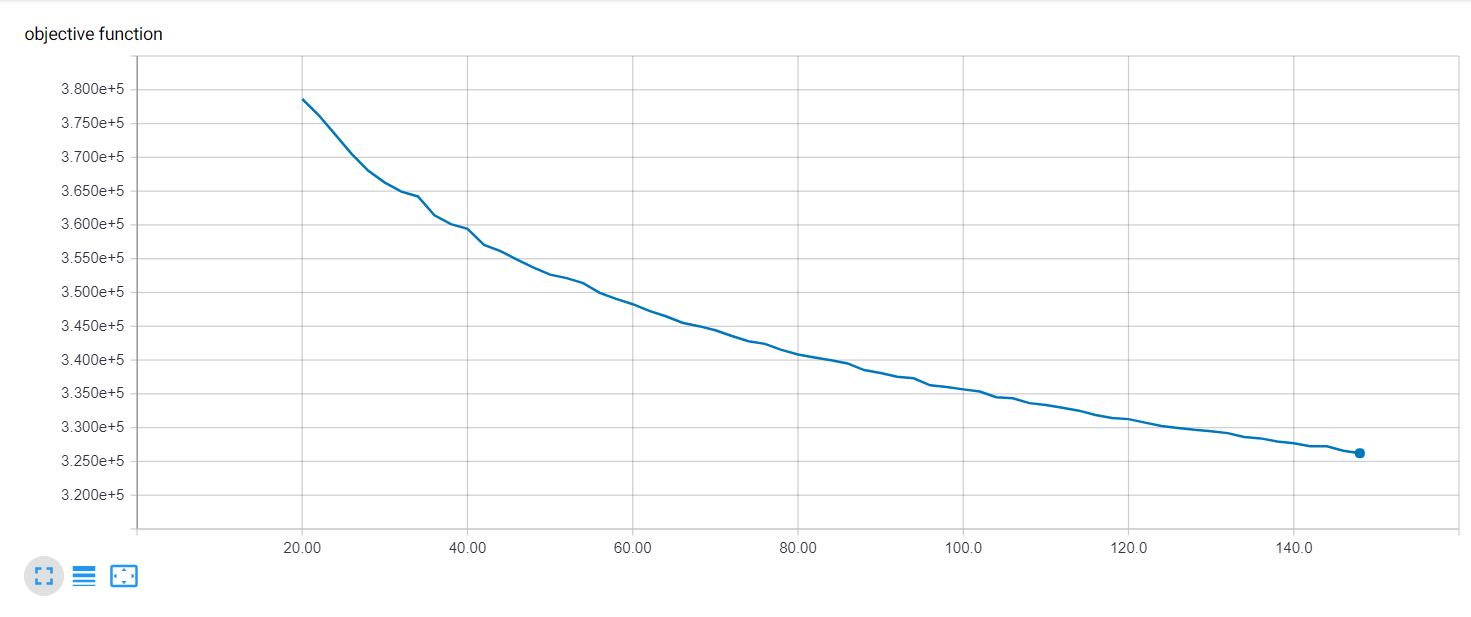
2.

a)



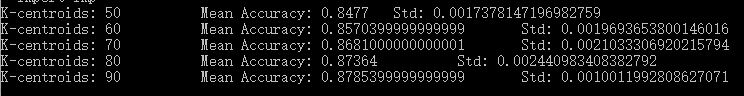
*Figure 1: Elbow finding graph*

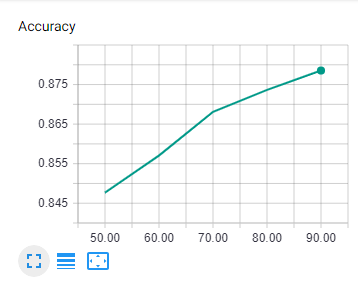
Above is a graph where the y-axis denotes the value of the objective function and the x-axis denotes k, the number of centroids. We ran the elbow finding experiment from k=20 to k=150 with increments of 2. We see that the graph is more like a curve. At the start, it is steeper and right around 60 to 70 we see that the graph slows down. So, we think it is a reasonable assumption that choosing around 70 hidden neurons for our hidden layer is an optimal choice.

b)

In our implementation of the code, you will see that we used 5-fold cross correlation when we run our experiments to investigate the performance on our neural network for different sizes of hidden layer and dropout in the hidden layer.

c) We ran a 5-fold cross validation for hidden layer sizes of 50, 60, 70, 80, and 90 for 50 epochs each due to time constraints. As such, we did not train till convergence. Another note is that we do not have a graph for standard deviation because it was not implemented at this time, however we do have a graph showing the mean accuracy over the k-fold cross validation for each hidden layer size.

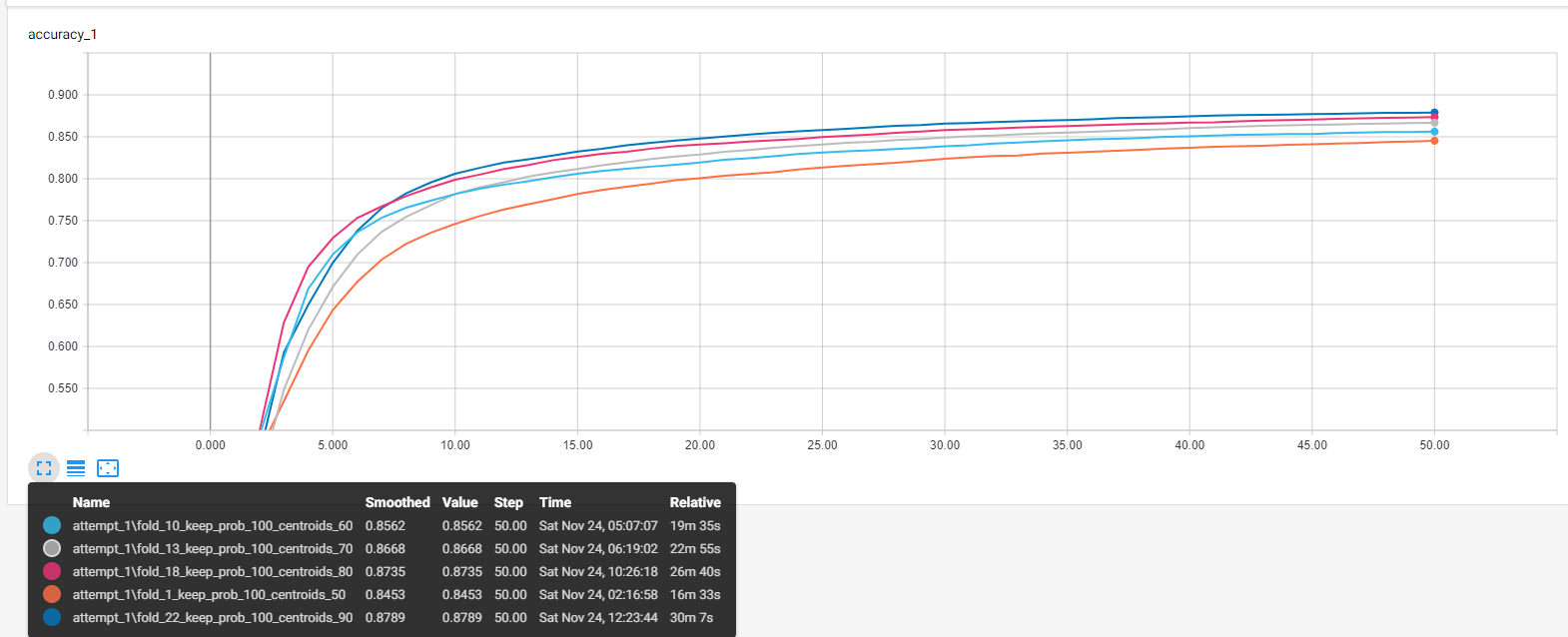




*Figure 2: Mean accuracy on 5-fold Cross Validation over Different Hidden Layer Sizes.*

*Y-axis denotes accuracy, X-axis denotes number of hidden neurons in layer*

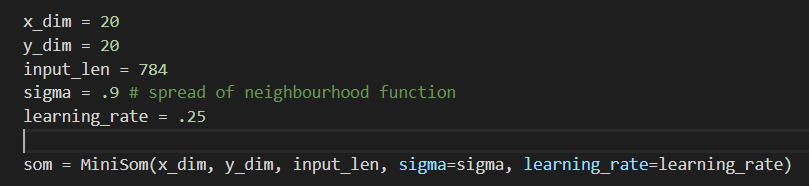
As you can see for figure 2, with each increase in the size of the hidden layer, the accuracy also increases. This makes sense as you have more centroids to differentiate the different classes.



*Figure 3: Comparing Testing Accuracy at each epoch for different hidden layer sizes*

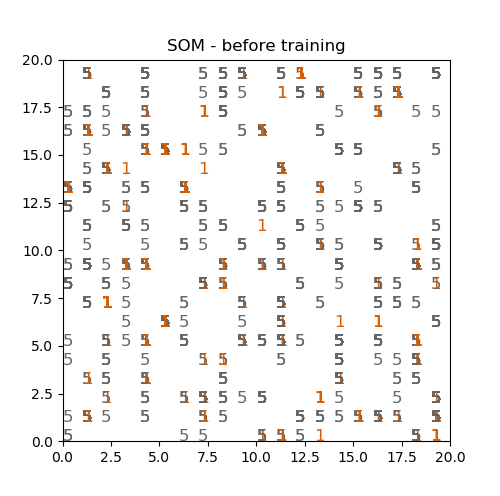
Figure 3 is just another graph visualization of the accuracy improvements when increasing hidden layer size. This graph only shows the training accuracies for a single fold for each hidden layer size of 50, 60, 70, 80 and 90.

3.

a) 

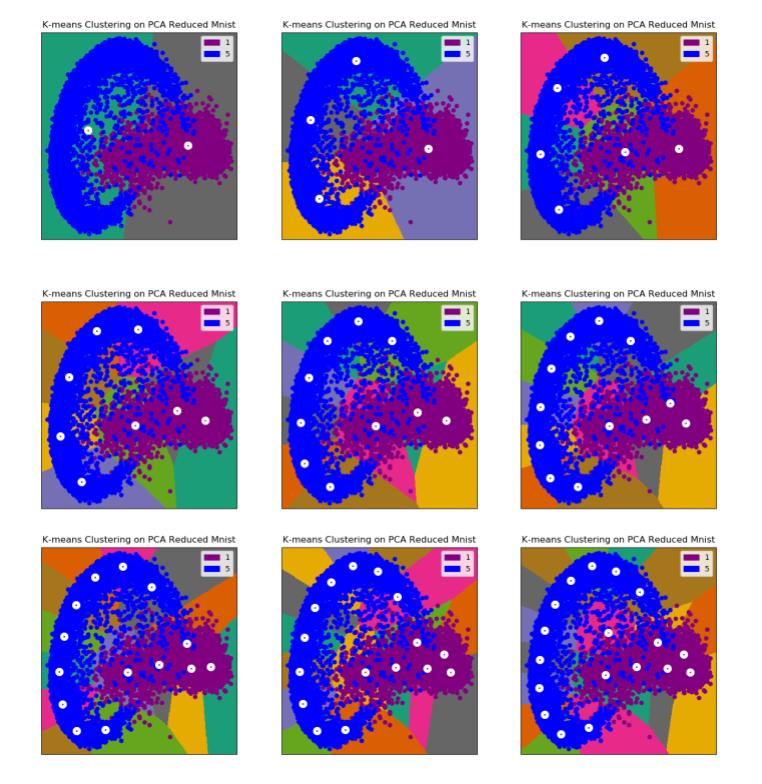
In our implementation, we used a dimension of (20,20) for our SOM, 0.9 for sigma and a learning rate of 0.25.

Plots for before training SOM and after training SOM.



As you can see from the before and after shots of the SOM, the 1’s and 5’s are mostly organized and clustered with each other.

Below are Voronoi graphs for ‘1’ and ‘5’ for K-means solution. We tried from k=2 to k=20 with increments of 2. In the graph, each centroid is denoted as a white circle and the background colors surrounding each centroid show the area that is closest to that centroid.

From a qualitative perspective, we see that the network will obviously have a low accuracy if k=2 as seen in the first graph, but with k=10 the graph is well separated and should have a much higher accuracy.