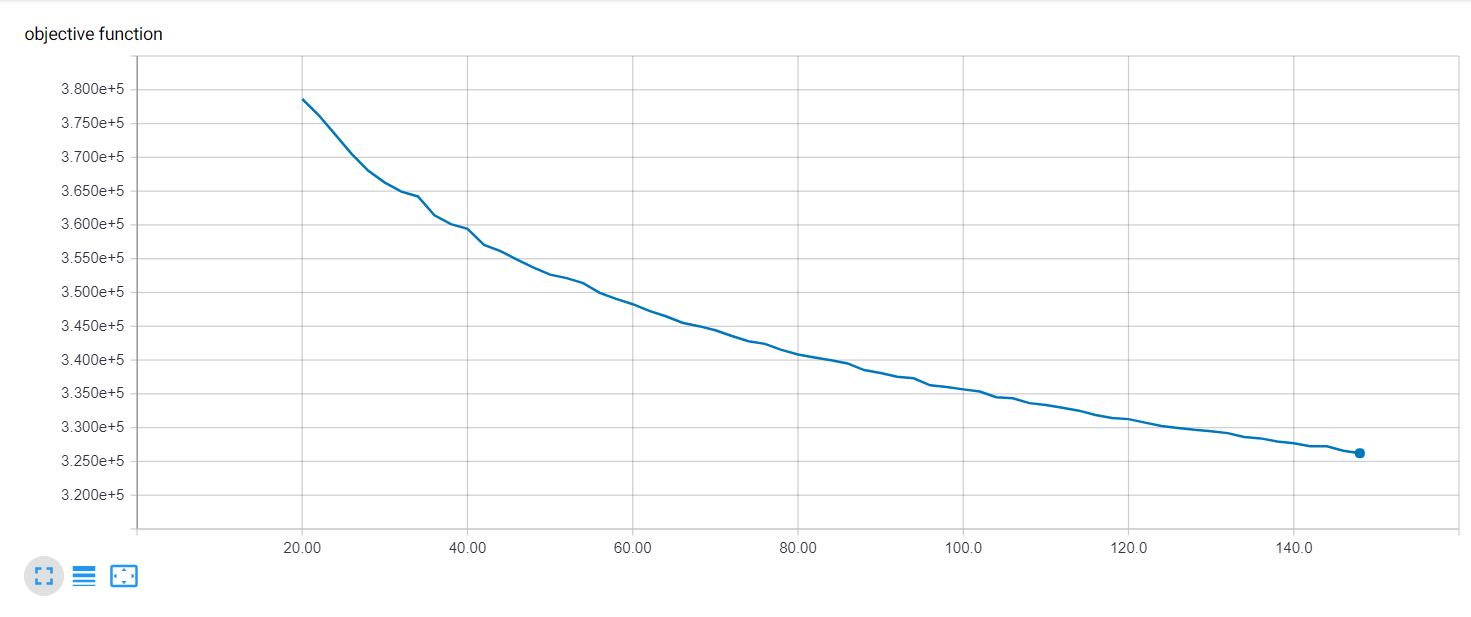
2.

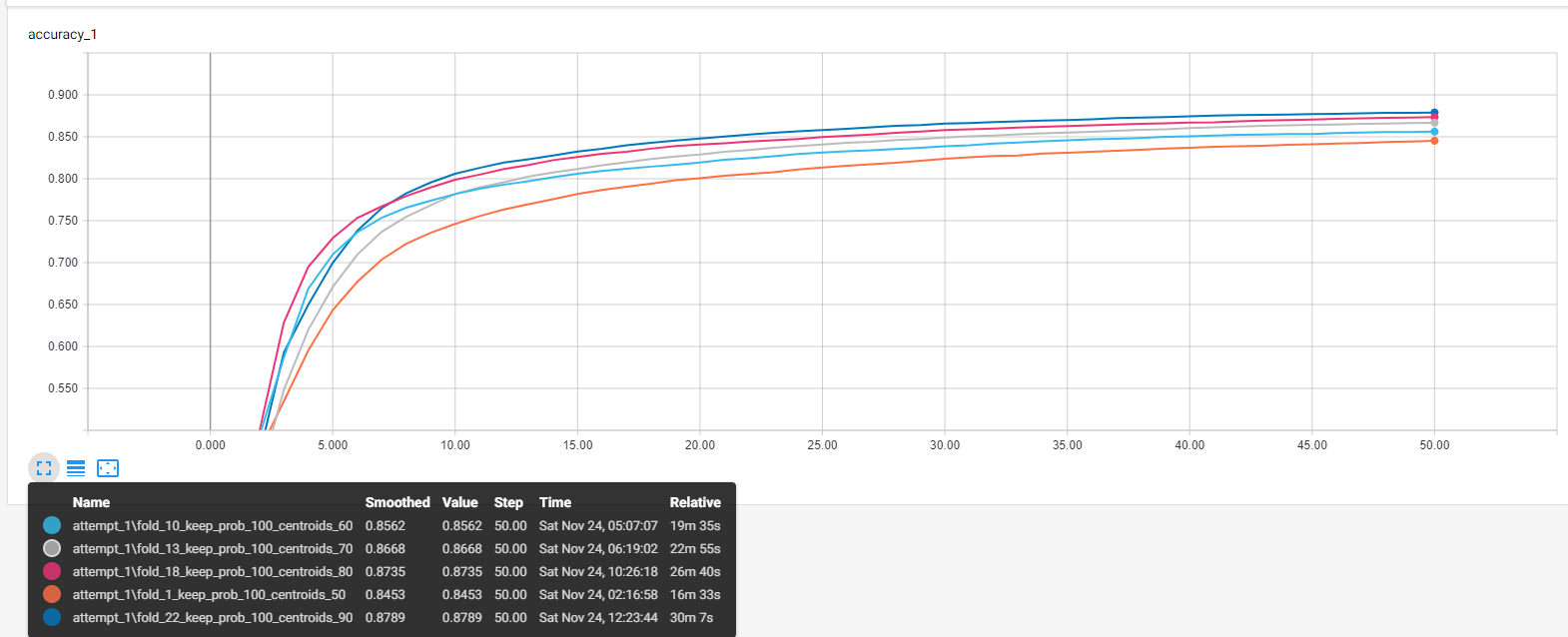
**a)**



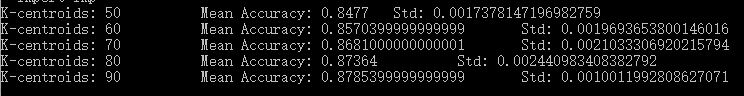
*Figure 1: Elbow finding graph, y-axis denotes value of objective function, x-axis denotes number of centroids*

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 validation (CV) when we run our experiments to investigate the performance on our neural network for different hidden layer sizes and different dropout percentages.

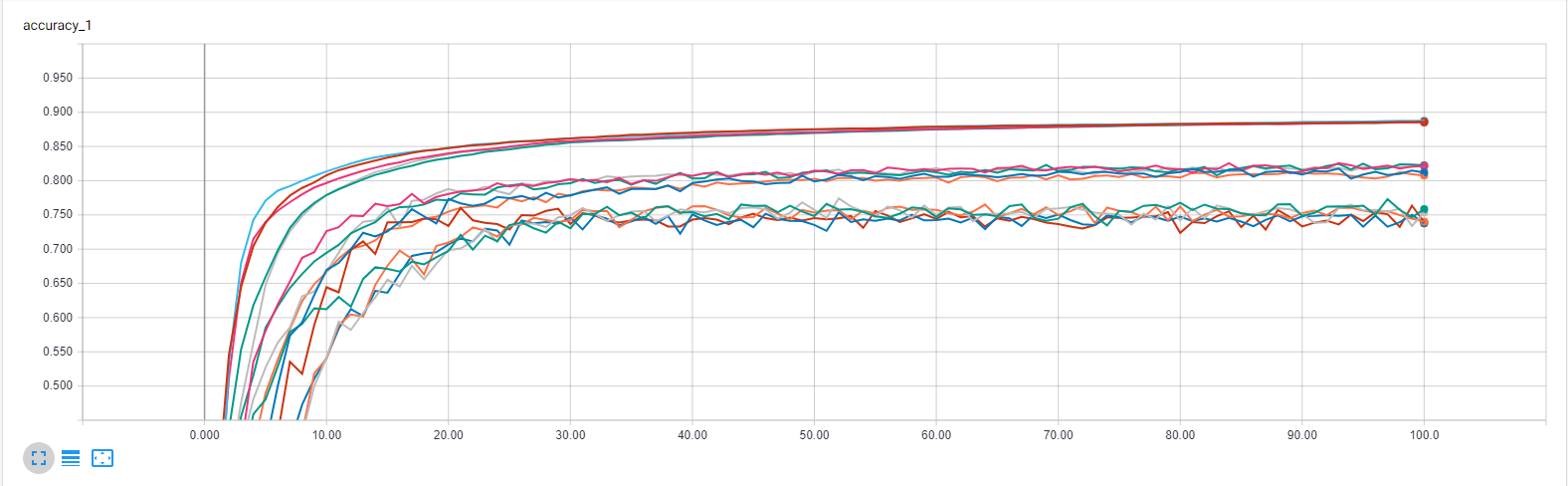
**c)** We ran a 5-fold CV for hidden layer sizes of 50, 60, 70, 80, and 90 for 50 epochs each due to time constraints. As such, we did not train till convergence. 

*Figure 2: Comparing Testing Accuracy at each epoch for different hidden layer sizes with 5-fold CV*



*Table 1: Accuracy mean and standard deviation of 5-fold CV for k-centroids*

Figure 2 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. From Table 1, we can see that the accuracy increases as we increase the hidden layer size and this makes sense because with more centroids, the network will naturally be able to separate the inputs as it uses Euclidean distance as the metric for predicting labels.



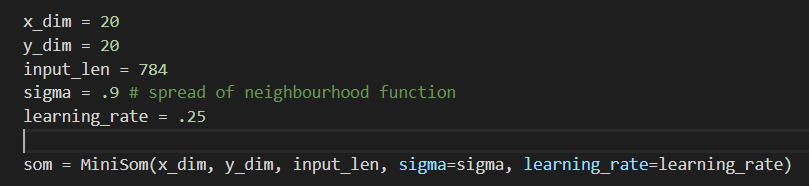
*Figure 3: Comparing Testing Accuracy at different dropouts for a hidden layer size of 70 with 5-fold CV*

https://scontent.fykz1-1.fna.fbcdn.net/v/t1.15752-9/46712067_265463364165962_4848437736907997184_n.png?_nc_cat=111&_nc_ht=scontent.fykz1-1.fna&oh=9175a4f6c22f38c4cf37ffe9378da33e&oe=5C6825DE

*Table 2: Accuracy mean and standard deviation of 5-fold CV for dropouts*

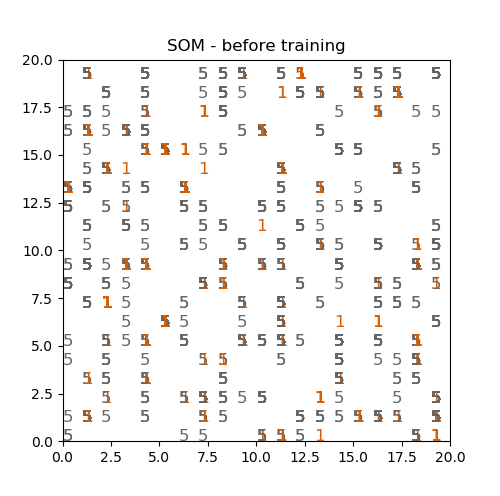
Figure 3 shows the accuracies of each 5-fold cross validation for the different dropout levels at k=70. We see from Table 2 that the lower keep probability or the more neurons we dropout at the hidden layer, the lower the accuracy becomes. This makes sense because for an RBF network, each neuron is an expert already and so the number of clusters, which is also the number of hidden layer neurons, is already indicative of the features in the underlying data. Each cluster is important because the metric for classifying the input image is based upon the Euclidean distance between the input image with each cluster. By dropping out these neurons, we lose that expert and thus we have a lower accuracy. From this we can see that dropout is not necessarily a good technique to add into an RBF network.

3.

a) 

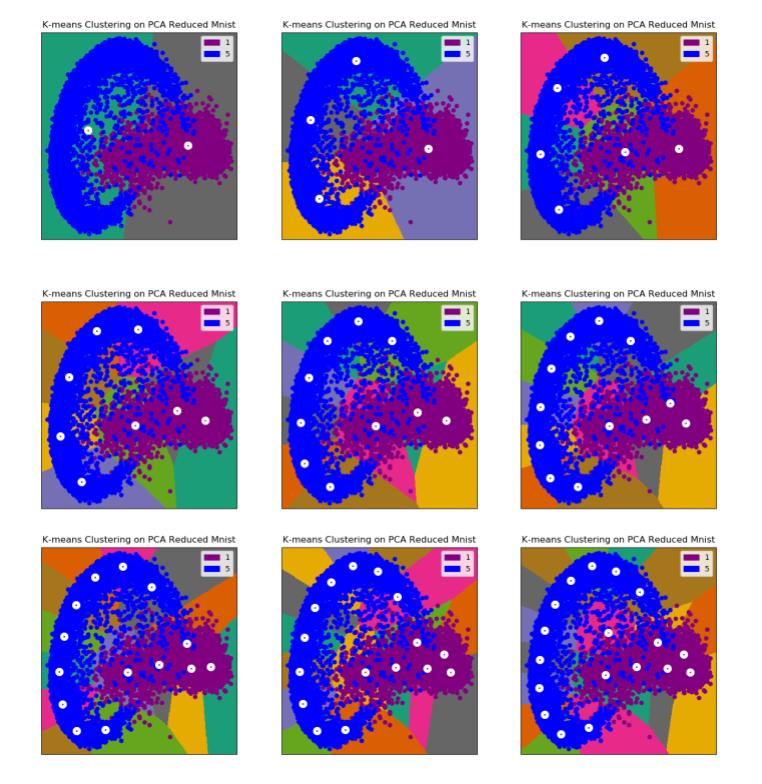
In our implementation, we used a dimension of (20,20) for our SOM, 0.9 for sigma and a learning rate of 0.25.

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.