Machine Learning CSC 6850 Home Work 2

Name: Abbaas Alif Mohamed Nishar

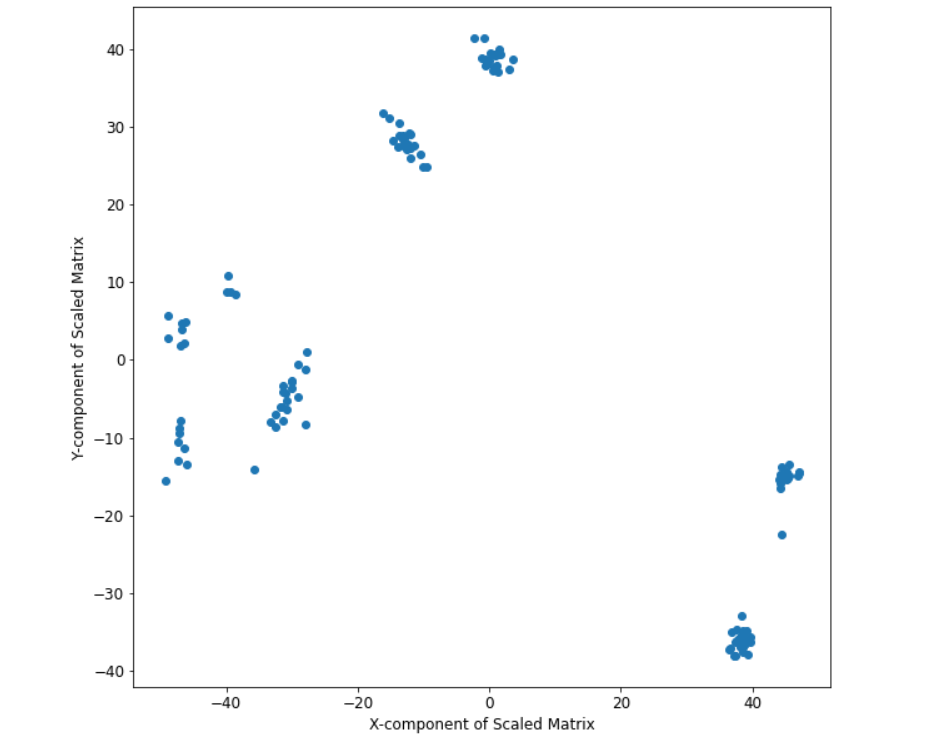
Panther ID: 002640321

**Applying MDS on RNA sequence data and clustering them using K-means clustering algorithm**

**Answers:**

For the explanation and documentation please refer to the ipython file as I have written it as markdowns. We calculate the hamming distance and then use it to compute the MDS in 2 dimensions for the RNA sequence data.

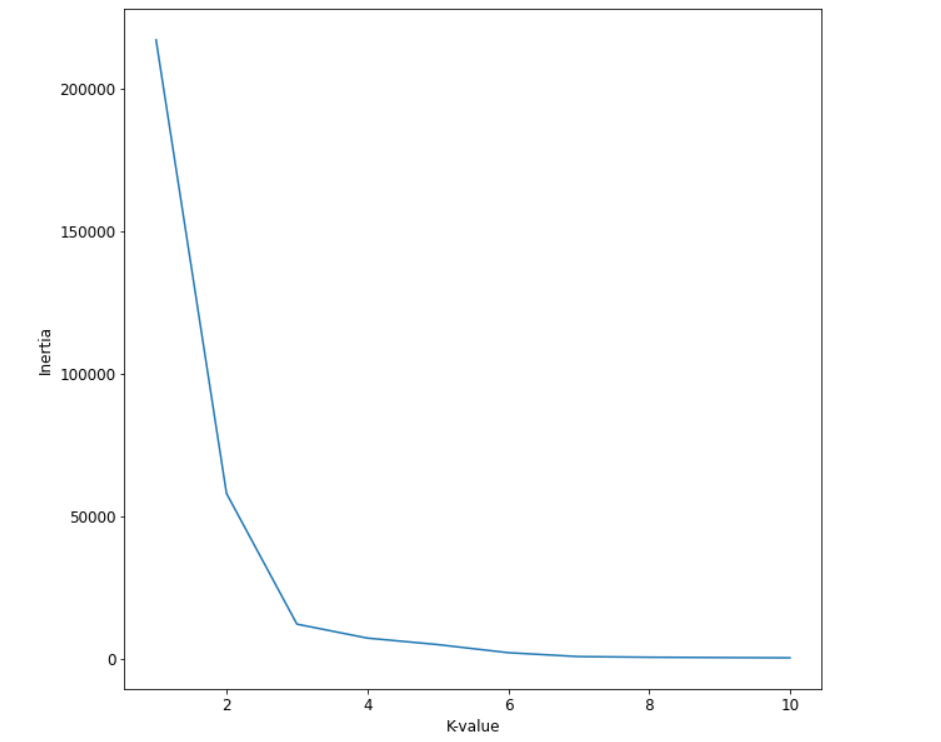
Output of MDS:



We have used a random seed to 6998 in the MDS in order to maintain repeatability. From the above figure we can clearly make out that there are three clusters so we choose k=3.

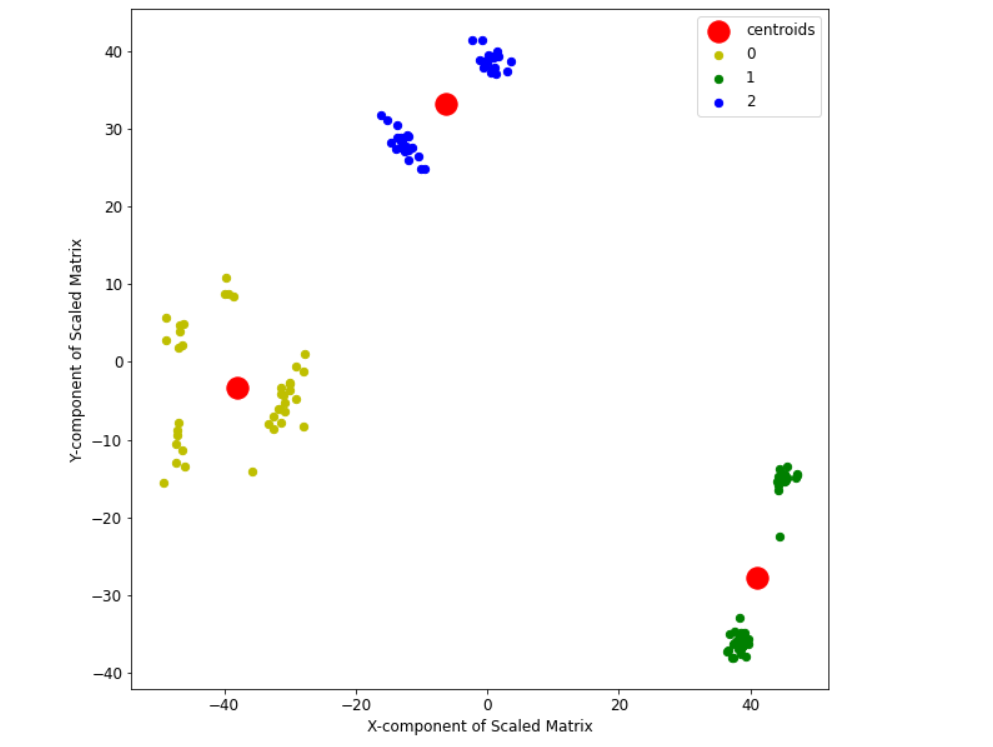
Output of Elbow Plot

Even though Elbow plot was not a part of the exercise we have included it to see that whether our intuition of choosing k=3 is validated.



Here we can see that the slope of the inertia vs K-value curve has reduced drastically after the point k=3. So that is our elbow point and our intuition through visual inspection is correct.

Output of K-means



Here we have plotted the points belonging to the same class as yellow, blue and green respectively and plotted their centroids using a bigger marker size in the red dot