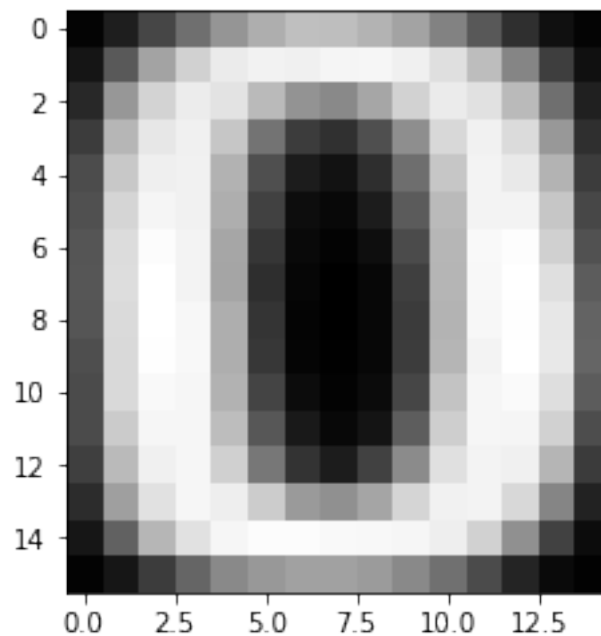Dragi Kamov
Aadil Anil Kumar

Mini-Project 1

K-Means Clustering:

1. Randomly define K centroids

2. Assign data points to the closest centroid using standard Euclidean distance

3. Calculate the mean values of all points belonging to a centroid, this is the new value of the centroid

4. Repeat steps 2 and 3 for specified iterations, causing convergence
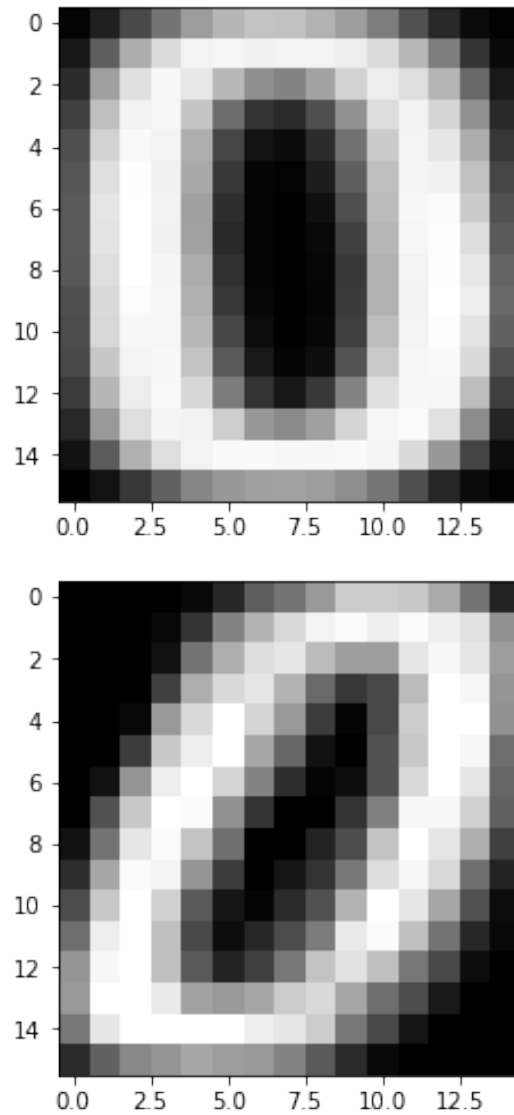
K = 1

For the case where K = 1, there is only one cluster that is created. Hence, running many iterations does not make a difference in the convergence of the centroids as it just the mean value of the set. We can see this by taking the mean of the zeros vector and then computing the Euclidean Distance from the mean to the centroid of the cluster. (See Juptyer Notebook)
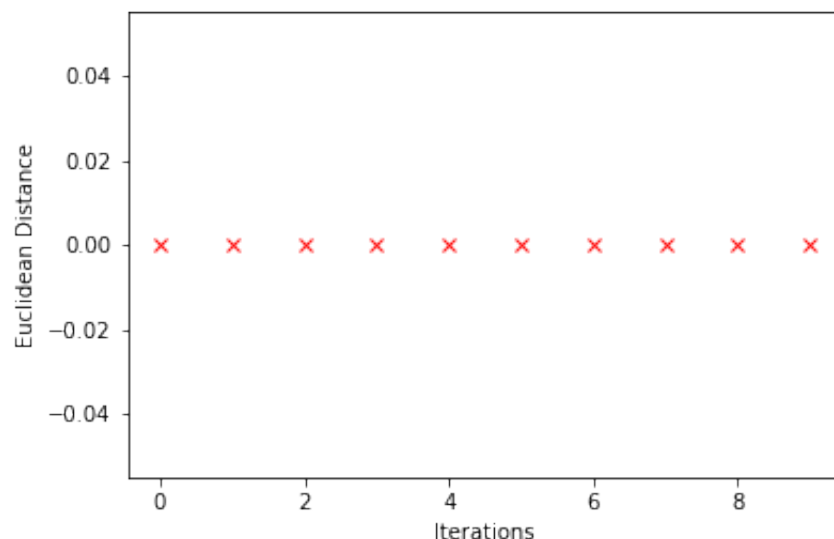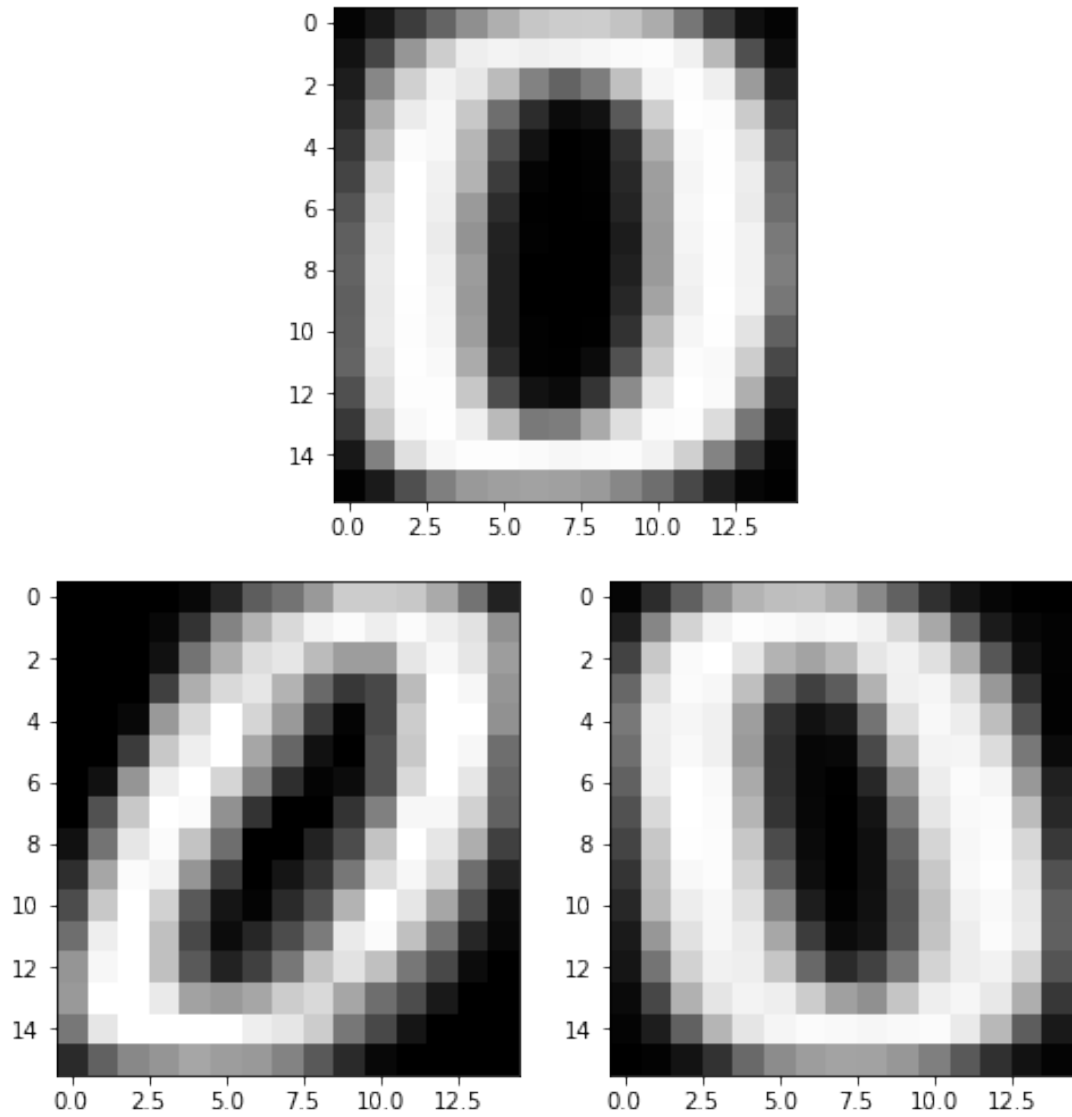
K = 2

For the case of K = 2, there are two clusters that are created. At this point the amount of iterations that are performed matters, as with each iteration the centroids get closer to convergence. However, it converges very quick and the distance from the mean of the vector and centroids stays the same for the rest of the iterations.
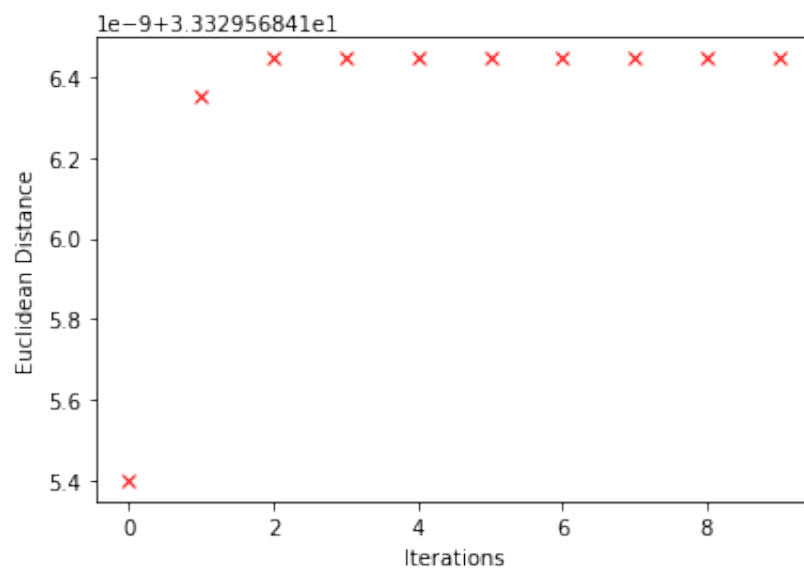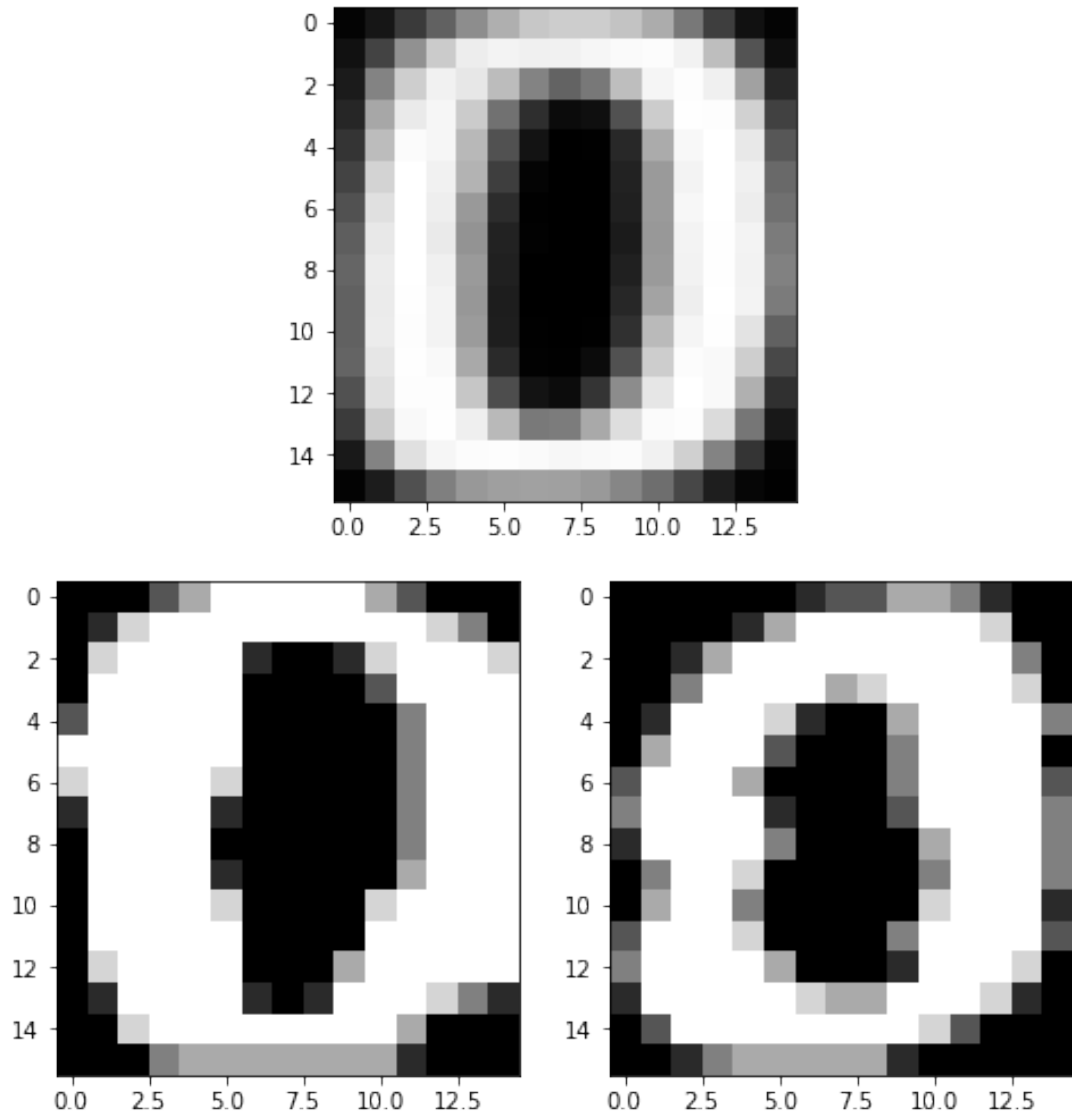
K = 3

For the case of K = 3, there are three clusters that are created. Again as with K = 2, the number of iterations definetly increases the rate of convergence.
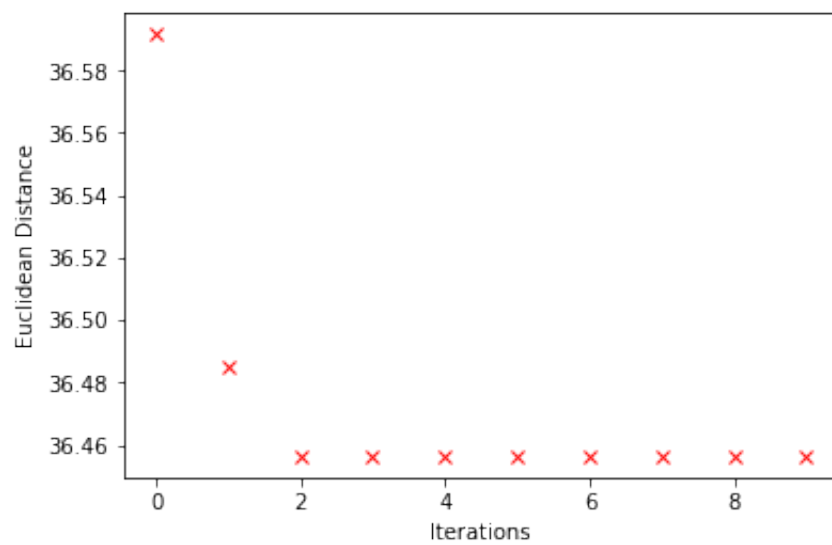
K = 200

For the case of K = 200, there are 200 clusters created. In this case all 200 data points form their own cluster. Iterations do not make a difference as the Euclidean distance stays the same.