CSE 4309 Assignment 6

Inshaad Merchant - 1001861293

Due Date: 7 November 2024

Task 1:

The Python file is attached to this document.

Training and testing on the pendigits dataset, with k=1

classification accuracy= 97.4271

Training and testing on the pendigits dataset, with k=3

classification accuracy= 97.5033

Training and testing on the pendigits dataset, with k=5

classification accuracy= 97.6272

Task 2:

The Python file is attached to this document.

Task 3:

No, the clustering cannot be the result of the k-means algorithm. This is because the blue dots on the adjacent sides of the red dots are closer to the red's mean compared to the blue's mean, which in the first case is the red dot itself. So, the adjacent blue dots will be changed to red dots, and this will go on like this until a stable cluster is achieved.

Task 4:

Part a:

No, it will not yield the exact same results if applied on the same dataset with the same K because the EM algorithm uses a mixture of Gaussians that will

always give a different value; hence, slightly different results each time.

Part b:

Yes, the agglomerative clustering will always give the same result with the same dataset because the minimum distance with the same data sets without any ties will always give the same clusters.

Task 5:

Part a: agglomerative clustering using dmin distance

At the top level, every object is its own cluster:

$$(2), (4), (7), (11), (16), (22), (29), (37)$$

Under the top level, each level is obtained by merging the less distant clusters from the previous level: For example in this case, 2 and 4 are the closest clusters (distance = 2)

$$(2, 4), (7), (11), (16), (22), (29), (37)$$

Next, 7 is the closest element to the cluster (2,4): (distance = 3)

$$(2, 4, 7), (11), (16), (22), (29), (37)$$

Continuing this process:

$$(2, 4, 7, 11), (16), (22), (29), (37)$$

$$(2, 4, 7, 11, 16), (22), (29), (37)$$

$$(2, 4, 7, 11, 16, 22), (29), (37)$$

$$(2, 4, 7, 11, 16, 22, 29), (37)$$

$$(2, 4, 7, 11, 16, 22, 29, 37)$$

Part b: agglomerative clustering using dmax distance

At the top level, every object is its own cluster:

$$(2), (4), (7), (11), (16), (22), (29), (37)$$

Under the top level, each level is obtained by merging the clusters using the distance between them.

$$Min\_max\_dist : (2, 4) -> 2$$

Clusters: (2, 4) (7) (11) (16) (22) (29) (37)

$$Min\_max\_dist : (7, 11) -> 4$$

Clusters: (2, 4) (7, 11) (16) (22) (29) (37)

$$Min\_max\_dist : (16, 22) -> 6$$

Clusters: (2, 4) (7, 11) (16, 22) (29) (37)

$$Min\_max\_dist : (29, 37) -> 8$$

Clusters: (2, 4) (7, 11) (16, 22) (29, 37)

$$Min\_max\_dist : (2, 11) -> 9$$

Clusters: (2, 4, 7, 11) (16, 22) (29, 37)

$$Min\_max\_dist : (2, 22) -> 20$$

Clusters: (2, 4, 7, 11, 16, 22) (29, 37)

$$Min\_max\_dist : (2, 37) -> 35$$

Clusters: (2, 4, 7, 11, 16, 22, 29, 37)