Long Nguyen

1001705873

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

classification accuracy=0.9743

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

classification accuracy=0.9750

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

classification accuracy=0.9763

2)

3) No, since in k-means clustering, each object is assigned and reassigned to clusters based on distances from each object to each cluster mean. This means that each reassignment should be getting closer to the center of the square, not on the boundary, since the mean would be in the center of the square. The blue cluster would be close to the center of the square, making it closer to the red dot than the red dot is to itself, meaning the red dot would be in the blue cluster near the center.

4a) The EM can produce different results if run multiple times. Example, given a dataset x1, x2, ..., xn and K, each run of the EM algorithm involves estimating w_i, μ_i, and σ_i for each N_i. Because we don't know the assignments of x_j to one of the Gaussians, we cannot compute w_i, μ_i, and σ_i easily. Therefore, estimating those values can result in different values for the initialization, resulting in whether there is bad fit or not.

4b) Yes, since the result of agglomerative clustering would be deterministic since the minimum distance between clusters is merged based on will be fixed (i.e. The distance between clusters is the same). Since the distance between clusters is the same in the same dataset, the same pairing of clusters will always be chosen in the same order.

5a) (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)

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

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

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

5b) (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)

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

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

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