1. Given the following 4 points with 2 attributes:

A: (2, 2), B: (2, 3), C: (3, 5), D: (4, 3).

The distance function is Euclidean distance.

Perform agglomerative hierarchical clustering using the single link (or MIN) approach and the complete link (or MAX) approach, respectively. Show the order in which the points are merged.

$$d(A,B)=1, \quad d(A,C)=\sqrt{10}, \quad d(A,D)=\sqrt{5},$$

$$d(B,C)=\sqrt{5}, \quad d(B,D)=2, \quad d(C,D)=\sqrt{5}. \text{ Thus,}$$

(1) For the single link (or MIN) approach:

A and B are merged firstly. We denote the cluster containing A and B by C1. Then, for C, D, C1:

$$d(C,C1)=\sqrt{5}, \quad d(D,C1)=2, \quad d(C,D)=\sqrt{5}.$$

Thus, D and C1 (A and B) are then merged, which is denoted by C2.

Finally, C and C2 (A, B and D) are merged.

(2) For the complete link (or MAX) approach:

A and B are merged firstly. We denote the cluster containing A and B by C1. Then, for C, D, C1:

$$d(C,C1)=\sqrt{10}, \quad d(D,C1)=\sqrt{5}, \quad d(C,D)=\sqrt{5}.$$

Thus, we have the following two orders:

(a) D and C1 (A and B) are then merged, which is denoted by C2.

Finally, C and C2 (A, B and D) are merged.

(b) C and D are then merged, which is denoted by C2.

Finally, C1 (A and B) and C2 (C and D) are merged.

2. We consider the following 6 data points:

p1: (5, 9), p2: (5, 8), p3: (3, 8), p4: (1, 2), p5: (2, 1), p6: (4, 4).

The distance function is Euclidean distance.

Find the clusters in this data set based on DBSCAN, with Eps=2 and Minpts=3.

Identify the core points, border points and noise points.

The neighborhood of each point is as follows:

N(p1)={p1, p2}, N(p2)={p1, p2, p3}, N(p3)={p2, p3},

N(p4)={p4, p5}, N(p5)={p4, p5}, N(p6)={p6}. Thus,

The core point is p2.

The border points are p1, p3.

The noise points are p4, p5, p6.

3. Given the following 6 points with 2 attributes:

A: (1, 3), B: (2, 1), C: (2, 2), D: (3, 5), E: (4, 4), F: (3, 3).

a) We need to group all 6 points into three clusters. Suppose initially we assign B, D and E as the prototype of the first, second and third cluster respectively. Use the k-Means algorithm to find the three clusters and their respective centroids after the first iteration.

b) If the initial class label of A, D and E is "C1", the initial class label of B, C and F is "C2", use the k-Means algorithm to find the two clusters and their respective centroids until convergence.

a) After the first iteration:

The first cluster is {A, B, C}, and its centroid is (5/3, 2).

The second cluster is {D}, and its centroid is (3, 5).

The third cluster is {E, F}, and its centroid is (3.5, 3.5).

b) Initially, the first cluster "C1" is {A, D, E}, and its centroid is (8/3, 4).

The second cluster "C2" is {B, C, F}, and its centroid is (7/3, 2).

After the first iteration, the first cluster "C1" is {D, E, F}, and its centroid is (4, 4).

The second cluster "C2" is {A, B, C}, and its centroid is (5/3, 2).

Then, the k-Means algorithm is convergence.