

1. We consider the following 6 data points:

$p_1: (5, 9), p_2: (5, 8), p_3: (3, 8), p_4: (1, 2), p_5: (2, 1), p_6: (4, 4)$.

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:

$N(p_1)=\{p_1, p_2\}, N(p_2)=\{p_1, p_2, p_3\}, N(p_3)=\{p_2, p_3\},$

$N(p_4)=\{p_4, p_5\}, N(p_5)=\{p_4, p_5\}, N(p_6)=\{p_6\}$. Thus,

The core point is p_2 .

The border points are p_1, p_3 .

The noise points are p_4, p_5, p_6 .

2. 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 C_1 .

Then, for C, D, C_1 :

$$d(C, C_1) = \sqrt{5}, \quad d(D, C_1) = 2, \quad d(C, D) = \sqrt{5}.$$

Thus, D and C_1 (A and B) are then merged, which is denoted by C_2 .

Finally, C and C_2 (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 C_1 .

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.