

9.5 密度聚类

1. 密度聚类也称“基于密度的聚类”，此类算法假设聚类结构能够通过样本分布的紧密程度确定。通常情形下，密度聚类算法从样本密度的角度来考察样本之间的可连接性，并基于可连接样本不断扩展聚类簇以获得最终聚类结果。

2. 左-邻域：对 $o_j \in D$ ，其左-邻域包含样本 D 中与 o_j 的距离不大于 ϵ 的样本

$$N_{\epsilon}(o_j) = \{o_i \in D \mid \text{dist}(o_i, o_j) \leq \epsilon\}$$

若与核心对象有关

核心对象：若 o_j 的左-邻域至少包含 MinPts 个样本， p 。

密度直达：若 o_j 位于 o_i 的左-邻域内，且 o_j 为核心对象，则 o_j 由 o_i 密度直达。

密度可达：对 o_i 与 o_j ，若存在样本序列 p_1, p_2, \dots, p_n ，其中 $p_1 = o_i$ ， $p_n = o_j$ 且 p_{i+1} 由 p_i 密度直达，则称 o_j 由 o_i 密度可达。

密度相连：对 o_i 与 o_j ，若存在 o_k 使得 o_i 与 o_j 均由 o_k 密度直达，则称 o_i 与 o_j 密度相连。

3. “簇”定义为：由密度可达关系导出的最大的密度相连集合。形式化地说，给定邻域参数 $(\epsilon, \text{MinPts})$ ，簇 $C \subseteq D$ 是满足以下性质。

a. 可连接性： $o_i \in C, o_j \in C \Rightarrow o_i$ 与 o_j 密度相连

b. 最大性： $o_i \in C, o_j$ 由 o_i 密度可达， $\Rightarrow o_j \in C$ 。

