

9.6 层次聚类.

1. 层次聚类 试图在不同层次对数据集进行划分, 从而形成树形的聚类结构.

$$d_{\min}(C_i, C_j) = \min_{x \in C_i, z \in C_j} \text{dist}(x, z)$$

$$d_{\max}(C_i, C_j) = \max_{x \in C_i, z \in C_j} \text{dist}(x, z)$$

$$d_{\text{avg}}(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{x \in C_i} \sum_{z \in C_j} \text{dist}(x, z).$$

输入: 数据集 $D = \{x_1, x_2, \dots, x_m\}$

聚类距离度量函数 d

聚类数 k .

过程:

for $j = 1, 2, \dots, m$ do

$C_j = \{x_j\}$

for $i = 1, 2, \dots, m$ do

for $j = 1, 2, \dots, m$ do

$M(i, j) = d(C_i, C_j)$

$M(j, i) = M(i, j)$

设置当前聚类数: $g = m$.

while $g > k$ do:

找出距离最近的两个聚类 G^* 和 C_{j^*} .

合并 G^* 和 C_{j^*} : $G^* = G^* \cup C_{j^*}$

for $j = j^* + 1, j^* + 2, \dots, g$.

将聚类 G 重编号为 $G-1$

删除 M 的第 j^* 行与 j^*+1

for $j = 1, 2, \dots, g-1$ do

$M(i^*, j) = d(G^*, C_j)$

$M(j, i^*) = M(i^*, j)$

$g = g-1$

输出 $C = \{C_1, \dots, C_k\}$









