

# Statistical Learning HW8

丁泽衡

18307110088

(1) 解: 距离矩阵题目中已给出; 类间距离用单链接 (最近距离) 计算;

具体实现如下:

~~step 1~~ iteration 1:

step 1: 各样本自成一类, 共5类 ( $i=1,2,3,4,5$ )

step 2: 合并样本3与5为一类 (距离最小, 为1)

step 3: 新类记为“3号”, 新的距离矩阵 (单链接) 为:

$$D_1 = \begin{matrix} & \textcircled{1} & \textcircled{2} & \textcircled{3} & \textcircled{4} & \textcircled{5} \\ \textcircled{1} & 0 & 6 & 2 & 9 & \\ \textcircled{2} & 6 & 0 & 5 & 4 & \\ \textcircled{3} & 2 & 5 & 0 & 5 & \\ \textcircled{4} & 9 & 4 & 5 & 0 & \\ \textcircled{5} & & & & & \end{matrix}$$

即新类与1距离: 2

与2: 5

与4: 5

after iteration 1, 共4类: 样本1, 样本2, {样本3 & 样本5}, 样本4  
 $\textcircled{1} \quad \textcircled{2} \quad \textcircled{3} \quad \textcircled{4}$

iteration 2:

由  $D_1$ , 合并 {3,5} 与 1, 因为此时类1与类 {3,5} 距离最小 (为2)

∴ 新类记为“1号”

$$D_2 = \begin{matrix} & \textcircled{1} & \textcircled{2} & \textcircled{4} \\ \textcircled{1} & 0 & 5 & 5 & \\ \textcircled{2} & 5 & 0 & 4 & \\ \textcircled{4} & 5 & 4 & 0 & \end{matrix}$$

即新类 {1,3,5} 与2距离: 5

与4距离: 5

after iteration 2, 共3类: {样本1, 3, 5}, 样本2, 样本4  
 $\textcircled{1} \quad \textcircled{2} \quad \textcircled{4}$

iteration 3:

由  $D_2$ , 合并 2与4为新类 {2,4} (此时这两类距离最小, 为4)

∴ 记为“2号”

$$D_3 = \begin{matrix} & \textcircled{1} & \textcircled{2} \\ \textcircled{1} & 0 & 5 & \\ \textcircled{2} & 5 & 0 & \end{matrix}$$

即新类 {2,4} 与类 {1,3,5} 距离为: 5

after iteration 3, 共2类: {样本1, 3, 5} 与 {样本2, 4}  
 $\textcircled{1} \quad \textcircled{2}$

iteration 4:

由  $D_3$ , 合并 {2,4} 与 {1,3,5} 为新类 {1,2,3,4,5}

且此时仅有这一个类, ∴ 终止计算

⇒ after iteration 4, 全聚为一个类: {样本1, 2, 3, 4, 5}

①

#

iteration

(2) 解: ① 初始质心坐标为  $m_1 = x_4 = (5, 0)$ ;  $m_2 = x_5 = (5, 2)$  ( $x_5$  为类  $G_1, G_2$ )

② 计算各样本到质心距离: (用欧氏距离)

$$d_{11} = \sqrt{(0-5)^2 + (2-0)^2} = \sqrt{29} \quad d_{12} = \sqrt{5^2 + 0^2} = 5 \quad d_{12} < d_{11} \quad \therefore x_1 \text{ 分入类 } G_2 \text{ 中}$$

$$d_{21} = \sqrt{5^2 + 0^2} = 5 \quad d_{22} = \sqrt{5^2 + 2^2} = \sqrt{29} \quad d_{21} < d_{22} \quad \therefore x_2 \text{ 分入类 } G_1 \text{ 中}$$

$$\cancel{d_{31} = \sqrt{4^2 + 2^2} = \sqrt{20}} \quad d_{32} = \sqrt{4^2 + 2^2} = \sqrt{20} \quad d_{31} < d_{32} \quad \therefore x_3 \text{ 分入类 } G_1 \text{ 中}$$

$$d_{31} = \sqrt{4^2 + 0^2} = 4$$

$$d_{41} = 0 \quad d_{42} = \sqrt{2^2} = 2 \quad \Rightarrow x_4 \text{ 分入类 } G_1$$

$$d_{51} = \sqrt{2^2} = 2 \quad d_{52} = 0 \quad \Rightarrow x_5 \text{ 分入类 } G_2$$

第一轮迭代后,  $x_1, x_5 \in G_2$   $x_2, x_3, x_4 \in G_1$

③ 更新类中心:

$$m_2 = \frac{x_1 + x_5}{2} = \left(\frac{5}{2}, 2\right) = (2.5, 2) \quad m_1 = \frac{x_2 + x_3 + x_4}{3} = (2, 0)$$

④ 重复迭代:

$$d_{11} = \sqrt{2^2 + 2^2} = \sqrt{8} \quad d_{12} = \sqrt{2.5^2} = 2.5 \quad d_{11} > d_{12} \quad \therefore x_1 \in G_2$$

$$d_{21} = \sqrt{2^2} = 2 \quad d_{22} = \sqrt{2.5^2 + 2^2} = \frac{\sqrt{41}}{2} \quad d_{21} < d_{22} \quad \therefore x_2 \in G_1$$

$$d_{31} = \sqrt{1^2} = 1 \quad d_{32} = \sqrt{1.5^2 + 2^2} = 2.5 \quad d_{31} < d_{32} \quad \therefore x_3 \in G_1$$

$$d_{41} = \sqrt{3^2} = 3 \quad d_{42} = \sqrt{2.5^2 + 2^2} = \frac{\sqrt{41}}{2} \quad d_{41} < d_{42} \quad \therefore x_4 \in G_1$$

$$d_{51} = \sqrt{3^2 + 2^2} = \sqrt{13} \quad d_{52} = \sqrt{2.5^2} = 2.5 = \frac{5}{2} \quad d_{51} > d_{52} \quad \therefore x_5 \in G_2$$

可见, 更新类中心后迭代仍未改变结果,  $\therefore$  已收敛, 聚类结束

综上, 聚类结果为: 点  $x_2, x_3, x_4$  为一类, 类中心为  $(2, 0)$

点  $x_1, x_5$  为另一类, 类中心为  $(2.5, 2)$

#