

统计(机器)学习 Homework - 8

(1) 给定 5 个样本的集合, 样本之间的欧式距离由如下矩阵 D 表示:

$$D = [d_{ij}]_{5 \times 5} = \begin{bmatrix} 0 & 6 & 1 & 10 & 3 \\ 6 & 0 & 5 & 4 & 6 \\ 1 & 5 & 0 & 8 & 2 \\ 10 & 4 & 8 & 0 & 5 \\ 3 & 6 & 2 & 5 & 0 \end{bmatrix} \quad (0.1)$$

其中 $d_{i,j}$ 表示第 i 个样本与第 j 个样本之间的欧式距离。应用聚合层次聚类法对这 5 个样本进行聚类。

解: 在本题中, 使用最短距离 $P_{pe} = \min \{d_{ij} \mid x_i \in G_p, x_j \in G_e\}$ 作为类间距离

① 每个样本为一类, 共五类。可以发现(1)与(3)之间距离最近, 归为一类, 目前共 4 类

② 4 类之间的距离如下:

	(1,3)	(2)	(4)	(5)
(1,3)	0	5	8	2
(2)		0	4	6
(4)			0	5
(5)				0

将 (1,3) 与 (5) 归为一类, 目前共 3 类

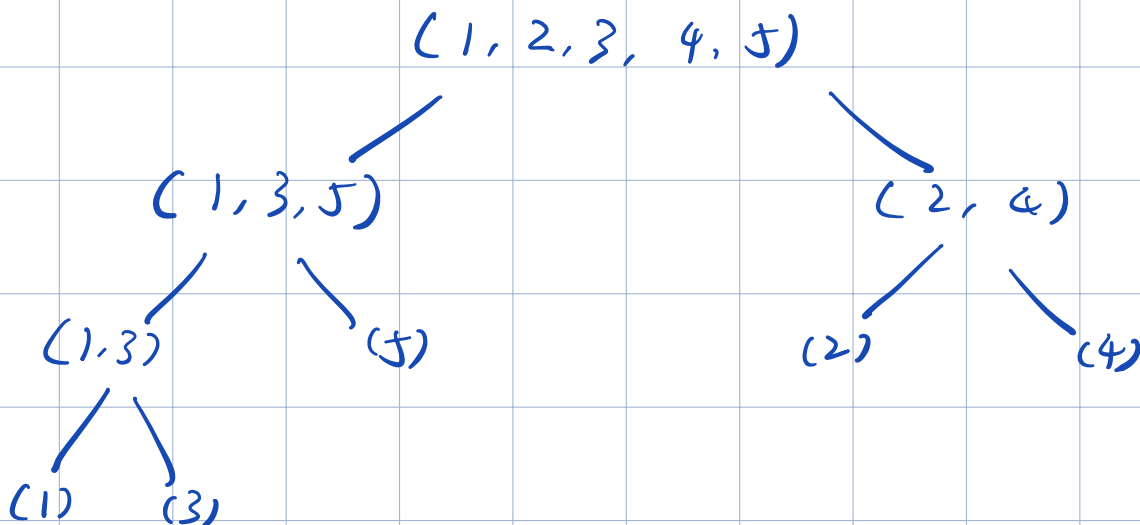
③ 3类之间的距离如下:

	(1,3,5)	(2)	(4)
(1,3,5)	0	5	5
(2)		0	4
(4)			0

将 (2) 与 (4) 合成一类, 目前共2类

④ 由于只剩下 (1,3,5) 与 (2,4) 两类, 故直接合并为一类 (1, 2, 3, 4, 5)

层次聚类结果如下:



□

(2) 给定样本集合:

$$\begin{pmatrix} 0 & 5 \\ 0 & 0 \\ 1 & 0 \\ 3 & 0 \\ 5 & 2 \end{pmatrix}$$

(0.2)

试用 k 均值算法将以上 5 个样本分到 2 个类别中 (设定 $k=2$, 迭代类别中心初始值为最后两个样本点取值)。

以上证明题请以 PDF 格式提交。

解: ① 第一轮: 中心: $(3, 0), (5, 2)$

$$\| (0, 5) - (3, 0) \|^2 = 3^2 + 5^2 = 34$$

$$\| (0, 5) - (5, 2) \|^2 = 5^2 + 3^2 = 34$$

由于距离相同, 随机分到 $(3, 0)$

$$\| (0, 0) - (3, 0) \|^2 = 3^2 + 0^2 = 9$$

$$\| (0, 0), (5, 2) \|^2 = 5^2 + 2^2 = 29$$

则 $(0, 0)$ 分到 $(3, 0)$ 的一类

$$\| (1, 0) - (3, 0) \|^2 = 2^2 + 0^2 = 4$$

$$\| (1, 0) - (5, 2) \|^2 = 4^2 + 2^2 = 20$$

则 $(1, 0)$ 分到 $(3, 0)$ 的一类

第1类样本有: $(0,5), (0,0), (1,0), (3,0)$

第2类样本有: $(5,2)$

更新中心:

$$\begin{aligned}\text{第1类中心} &: \frac{1}{4} \cdot [(0,5) + (0,0) + (1,0) + (3,0)] \\ &= (1, \frac{5}{4})\end{aligned}$$

$$\text{第2类中心: } (5,2)$$

② 第2轮: 更新中心: $(1, \frac{5}{4}), (5,2)$

$$\|(0,5) - (1, \frac{5}{4})\|^2 = 1^2 + (\frac{15}{4})^2$$

$$\|(0,5) - (5,2)\|^2 = 5^2 + 3^2$$

$(0,5)$ 分到 第1类

$$\|(0,0) - (1, \frac{5}{4})\|^2 = 1^2 + (\frac{5}{4})^2$$

$$\|(0,0) - (5,2)\|^2 = 5^2 + 2^2$$

$(0,0)$ 分到 第1类

$$\|(1,0) - (1, \frac{5}{4})\|^2 = 0^2 + (\frac{5}{4})^2$$

$$\|(1,0) - (5,2)\|^2 = 4^2 + 2^2$$

$(1,0)$ 分到 第1类

$$\|(3,0) - (1, \frac{5}{4})\|^2 = 2^2 + (\frac{5}{4})^2$$

$$\|(3,0) - (5,2)\|^2 = 2^2 + 2^2$$

$(3,0)$ 分到第1类

最后显然, $(5,2)$ 分到第2类

本轮分类结果:

第1类 $(0,5), (0,0), (1,0), (3,0)$

第2类 $(5,2)$

由于分类结果不变, 则结束迭代

□