

第二章作业

● 习题作业：

1. 已知 8 个样本点 $x_1 = (0, 0)^T$, $x_2 = (2, 0)^T$, $x_3 = (0, 2)^T$, $x_4 = (2, 2)^T$, $x_5 = (6, 6)^T$, $x_6 = (8, 6)^T$, $x_7 = (6, 8)^T$, $x_8 = (8, 8)^T$, 利用 K-均值算法将上述样本聚为 2 类, (1) 要求选用样本点 x_1 和 x_5 分别作为两类中心的初始位置。(2) 要求选用样本点 x_2 和 x_4 分别作为两类中心的初始位置。尝试分析初始点的选取对聚类效果的影响?

2. 已知 5 个样本, 每个样本 5 个特征, 数据如下:

$$x_1 = (0, 3, 1, 2, 0)^T, \quad x_2 = (1, 3, 0, 1, 0)^T, \quad x_3 = (3, 3, 0, 0, 1)^T, \quad x_4 = (1, 1, 0, 2, 0)^T, \\ x_5 = (3, 2, 1, 2, 1)^T, \text{ 进行分级聚类, 相似性度量采用最小距离准则, 最终分为 3 类。并画出聚类分级树。}$$

3. 设有 5 个四维模式, 按最小距离准则和Tanimoto测度进行系统分级聚类分析。

$$\begin{aligned} x_1 &= \{1 \ 0 \ 1 \ 0\} \\ x_2 &= \{0 \ 1 \ 0 \ 1\} \\ x_3 &= \{0 \ 1 \ 0 \ 0\} \\ x_4 &= \{0 \ 0 \ 0 \ 0\} \\ x_5 &= \{1 \ 0 \ 0 \ 0\} \end{aligned}$$

4. 给定一组数据

$$x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad x_4 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad x_5 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \\ x_6 = \begin{bmatrix} 4 \\ -5 \end{bmatrix}, \quad x_7 = \begin{bmatrix} 3 \\ -5 \end{bmatrix}, \quad x_8 = \begin{bmatrix} 4 \\ -4 \end{bmatrix}, \quad x_9 = \begin{bmatrix} 3 \\ -4 \end{bmatrix}, \quad x_{10} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

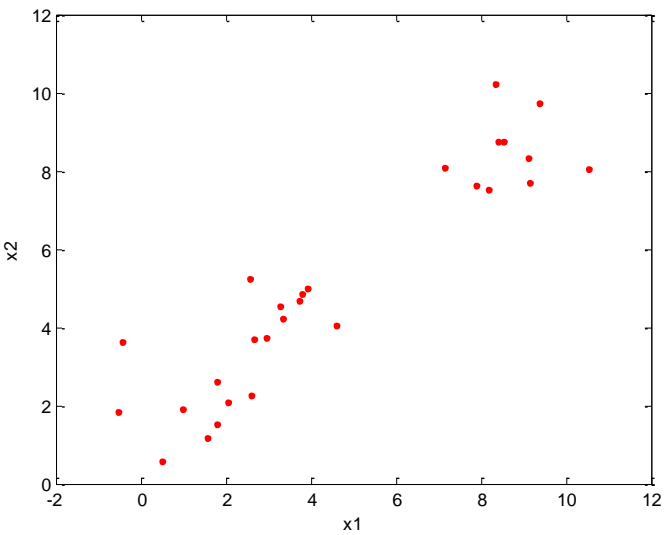
采用欧氏距离, 设置距离阈值为 3, 分别求取各数据点的局部密度值 ρ_i 及各点与更高密度值数据的距离 δ_i , 利用基于密度峰值的聚类方法确定这些数据的聚类中心数并实现聚类。

● 计算机作业：

编写 K-均值聚类算法程序，采用下面的数据进行聚类分析。

数据为：

x1	x2
-0.5200	1.8539
2.5849	2.2481
0.9919	1.9234
2.9443	3.7382
-0.4240	3.6220
1.7762	2.6264
2.0581	2.0918
1.5754	1.1924
1.7971	1.5387
0.4869	0.5940
7.8736	7.6255
8.1850	7.5291
9.3666	9.7513
8.4139	8.7532
10.5374	8.0650
9.1401	7.7072
7.1372	8.0828
8.5458	8.7662
8.3479	10.2368
9.1033	8.3269
3.7794	4.8633
3.7210	4.6794
3.2663	4.5548
3.9355	5.0016
2.5560	5.2594
4.6123	4.0442



2.6765 3.6859

3.3384 4.2267