

聚类分析 (k 均值)

1. 将如下的 8 个样本点 (用 (i, j) 代表位置) 用 k 均值聚类算法分为三个簇, $x_1(2, 10), x_2(2, 5), x_3(8, 4), x_4(5, 8), x_5(7, 5), x_6(6, 4), x_7(1, 2), x_8(4, 9)$ 距离函数是欧几里得距离, 假设初始我们选择 x_1, x_4 和 x_7 分别为每个簇的中心, 用 k 均值算法给出最后的三个簇是什么?

答:

- 1) 第一步选择 x_1, x_4, x_7 为聚类中心, 可以计算出:

$$S_1(1) = \{x_1\}$$

$$S_2(1) = \{x_4, x_3, x_5, x_6, x_8\}$$

$$S_3(1) = \{x_2, x_7\}$$

计算新的聚类中心:

$$Z_1(1) = (2, 10)$$

$$Z_2(1) = (6, 6)$$

$$Z_3(1) = \left(\frac{3}{2}, \frac{7}{2}\right)$$

- 2) 选择新的聚类中心, 可以计算出:

$$S_1(2) = \{x_1, x_8\}$$

$$S_2(2) = \{x_4, x_3, x_5, x_6\}$$

$$S_3(2) = \{x_2, x_7\}$$

计算新的聚类中心:

$$Z_1(2) = (3, 19/2)$$

$$Z_2(2) = (13/2, 21/4)$$

$$Z_3(2) = \left(\frac{3}{2}, \frac{7}{2}\right)$$

- 3) 选择新的聚类中心, 可以计算出: