

第五章 作业

1. 请简述 K 均值和层次聚类的异同。
2. 将如下的 8 个点聚类为 3 个簇：
 $x_1(2,10), x_2(2,5), x_3(8,4), x_4(5,8), x_5(7,5), x_6(6,4), x_7(1,2), x_8(4,9)$ 。
距离采用欧氏距离，假设初始质心分别是 x_1, x_2, x_7 ，用 K 均值算法给出：
(1) 第一次迭代后的 3 个簇的质心。
(2) 最终的 3 个簇的质心。
3. 指出在何种情况下，基于密度的聚类方法比 K 均值聚类 and 层次聚类方法更合适。通过实例说明。
4. 聚类样本足够多的情况下，K 均值算法会不会返回一个小于 K 个簇的结果？比如 $K=50$ 时，会不会产生 48 个簇的结果？分析原因。
5. 下面给出一个样本事务数据库（表 1），请对它分别实施 AGNES 算法和 DIANA 算法。

表 1 样本事务数据库

序号	属性 1	属性 2
1	1	1
2	1	2
3	2	1
4	2	2
5	3	4
6	3	5
7	4	4
8	4	5