

第五章作业

1. 聚类算法可分为哪几类？影响聚类算法性能优劣的因素有哪些？

2. 请给出最小张树聚类算法的具体过程，并分析其优缺点。

3. 有以下四个样本：

$$X_1 = (3,5)^T, X_2 = (5,1)^T, X_3 = (1,0)^T, X_4 = (1,4)^T$$

初始划分为两类， $\omega_1: \{X_1, X_2\}$ 和 $\omega_2: \{X_3, X_4\}$ ，则：

(1) 若将 X_2 移到 ω_2 类中，试计算出转移前后的总的类内离散度矩阵 S_w ，请给出具体计算过程。

(2) 若使用 S_w 的行列式作为聚类准则，(1)中的转移是否合适？请给出具体过程。

(3) 若使用 C-均值算法的准则函数，(1)中的转移是否合适？请给出具体过程。

4. 有以下五个样本：

$$X_1 = (0,1,2,1,2,4)^T,$$

$$X_2 = (3,2,3,1,2,1)^T,$$

$$X_3 = (1,0,0,0,1,1)^T,$$

$$X_4 = (2,1,0,2,1,2)^T,$$

$$X_5 = (0,0,1,0,1,0)^T,$$

请按照最小距离准则用层次聚类算法进行聚类分析(距离 D 直接用根号表示)，并给出层次聚类示意图。

5. 已知一个样本集合：

$$\{(0,0)^T, (1,1)^T, (2,1)^T, (4,4)^T, (5,3)^T, (5,4)^T, (6,3)^T, (6,5)^T\}$$

试用近邻函数法进行聚类分析，请给出具体过程和最终的聚类结果。

6. 查阅基于密度的聚类算法 DBSCAN，了解其相关概念定义，并给出算法的伪代码表达。