

1 绪论

2 聚类分析

2.1 距离聚类的概念

2.2 相似性测度和聚类准则

2.3 基于距离阈值的聚类算法

2.4 层次聚类法

2.5 动态聚类法

聚类过程中，聚类中心位置或个数发生变化。
两种常用的算法：

- K - 均值算法（或 C - 均值算法）
- 迭代自组织的数据分析算法

2.5.1 K - 均值算法

基于使聚类准则函数最小化。

准则函数 聚类集中每一样本点到该类中心的距离平方和。

对于第 j 个聚集类，准则函数定义为：

$$J_j = \sum_{i=1}^{N_j} \|X_i - Z_j\|^2, X_i \in S_j$$

算法描述

算法讨论

聚类准则函数 J_k 与 K 的关系曲线

2.5.2 迭代自组织的数据分析算法

迭代自组织的数据分析算法也常称为 ISODATA 算法 (Iterate Self-Organizing Data Analysis Techniques Algorithm, ISODATA).

算法特点

- 加入了试探性步骤，组成人机交互的结构；
- 可以通过类的自动合并与分裂得到较合理的类别数。

基本思路

1. 选择初始值
2. 按最邻近规则进行分类
3. 聚类后的处理：计算各类中的距离函数等指标
4. 判断结果是否符合要求，符合则结束，否则回到 2

算法描述 P31

常用指标 各指标综合考虑

1. 聚类中心之间的距离
2. 诸聚类域中样本数目
3. 诸聚类域中样本的标准差向量

3 判别函数及几何分类法

3.1 判别函数

统计模式识别

{聚类分析法（第二章）判别函数法 { 线性判别函数法 非线性判别函数法 } => xx 统计决策方法

3.1.1 判别函数

定义 直接用来对模式进行分类的准则函数。