

# 《CH7聚类分析》PPT课 件

制作人：PPT创作创作  
时间：2024年X月



# 目录

- 第1章 简介
- 第2章 K均值聚类算法
- 第3章 层次聚类算法
- 第4章 密度聚类算法
- 第5章 聚类分析在金融风控中的应用
- 第6章 总结与展望

● 01

# 第1章 简介

## 聚类分析概述

聚类分析是一种将数据分组或聚类在一起的技术，通过寻找数据点之间的相似性来划分不同的类别。在实际应用中，聚类分析被广泛应用于市场细分、社交网络分析和医学图像处理等领域。其基本原理是通过计算数据点之间的距离或相似性，将它们归入同一类别或簇中。

# 什么是聚类分析？

## 数据分组

将数据点归入不同  
类别或簇

## 簇形成

形成具有相似特征  
的簇

## 相似性划分

基于数据点之间的  
相似性划分类别

# 聚类分析的应用领域

## 市场细分

根据消费者的特征  
进行市场划分

## 医学图像处理

医学影像数据的分  
类和分析

## 社交网络分析

发现社交网络中的  
群体结构

# 聚类分析的基本原理

## 距离计算

计算数据点之间的  
距离

## 簇划分

将数据点归入不同  
簇

## 相似性度量

度量数据点之间的  
相似性

# 聚类分析方法

## K均值聚类算法

根据数据点之间的距离进行簇划分

需要预先设定簇的数量

适用于数据量较大的情况

## 层次聚类算法

通过树形结构展示数据点之间的相似性

不需要预先设定簇的数量

可以反映数据点之间的层次关系

## 密度聚类算法

根据数据点的密度进行簇划分

适用于数据分布不规则的情况

对噪声和离群点有较好的鲁棒性





# 聚类评价指标

## 轮廓系数

用于评估聚类效果  
的一种指标

## DBI指数

衡量簇内的紧密度  
和簇间的分离度

## CH系数

用于评价聚类的紧  
密性和分离性

01

## 实际案例解析

展示聚类分析的具体步骤

02

## 算法应用对比

比较不同聚类算法在案例中的效果

03

## 第2章 K均值聚类算法

# K均值聚类算法原理

## K均值算法步骤详解

初始化聚类中心

分配数据点到最近的聚类中心

重新计算聚类中心

迭代直至收敛

## K均值算法的优缺点

简单易实现

对大数据集收敛速度慢

容易收敛到局部最优解



## 01 实例演示K均值算法应用

展示数据点分布

## 02 选择最佳的K值

通过轮廓系数等指标确定

03

# K均值聚类算法进阶

## 加速K均值算法的技巧

使用KD树等数据结构

## K均值算法与其他聚类算法的比较

与层次聚类、密度聚类进行对比

## K均值聚类算法 应用

K均值算法在市场分析、  
社交网络分析等领域得到  
广泛应用。然而，由于对  
初始聚类中心敏感，容易  
受到异常值影响，需谨慎  
使用。

# K均值聚类算法应用

## 在实际场景中如何应用K均值算法

客户分群

推荐系统

图像压缩

## K均值算法的局限性

对初始聚类中心敏感

受异常值影响大

需要事先确定K值





# 第三章 层次聚类算法

# 层次聚类算法原理

层次聚类算法是一种基于聚类的数据分析方法，主要分为凝聚层次聚类方法和分裂层次聚类方法。凝聚层次聚类方法从单个数据点开始，逐步合并聚类直到形成一个大的聚类。分裂层次聚类方法则是从一个大的聚类开始，逐步分裂直到得到单个数据点的聚类。这两种方法各有优缺点，需要根据具体应用场景进行选择。



## 层次聚类算法实例

通过一个实例演示层次聚类算法的具体应用，可以更直观地理解算法的实际操作步骤。在实际应用中，不同的距离度量方法会对聚类结果产生影响，选择合适的距离度量方法可以得到更准确的聚类结果。



**01 剪枝策略**  
优化聚类结果

**02 与其他聚类算法的对比**  
比较优劣

**03**

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：  
<https://d.book118.com/558141057113006050>