

# 《大数据分析：方法与应用》

## 图书基本信息

书名：《大数据分析：方法与应用》

13位ISBN编号：9787302334170

10位ISBN编号：730233417X

出版时间：2013-9-1

出版社：清华大学出版社

页数：294

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：[www.tushu000.com](http://www.tushu000.com)

# 《大数据分析：方法与应用》

## 内容概要

《应用统计学系列教材·大数据分析:方法与应用》可用做统计学、管理学、计算机科学等专业进行数据挖掘、机器学习、人工智能等相关课程的本科高年级、研究生教材或教学参考书。

# 《大数据分析：方法与应用》

## 作者简介

1990-1994年 北京师范大学数学系学习，1994年6月毕业获理学学士学位；  
1994-1997年 北京师范大学数学系学习，1997年6月毕业获理科硕士学位；  
1999-2003年 中国人民大学统计学系学习，2003年6月毕业获经济学博士学位

## 书籍目录

### 第1章大数据分析概述

#### 1.1大数据概述

##### 1.1.1什么是大数据

##### 1.1.2数据、信息与认知

##### 1.1.3数据管理与数据库

##### 1.1.4数据仓库

##### 1.1.5数据挖掘的内涵和基本特征

#### 1.2数据挖掘的产生与功能

##### 1.2.1数据挖掘的历史

##### 1.2.2数据挖掘的功能

#### 1.3数据挖掘与相关领域之间的关系

##### 1.3.1数据挖掘与机器学习

##### 1.3.2数据挖掘与数据仓库

##### 1.3.3数据挖掘与统计学

##### 1.3.4数据挖掘与智能决策

##### 1.3.5数据挖掘与云计算

#### 1.4大数据研究方法

#### 1.5讨论题目

#### 1.6推荐阅读

### 第2章数据挖掘流程

#### 2.1数据挖掘流程概述

##### 2.1.1问题识别

##### 2.1.2数据理解

##### 2.1.3数据准备

##### 2.1.4建立模型

##### 2.1.5模型评价

##### 2.1.6部署应用

#### 2.2离群点发现

##### 2.2.1基于统计的离群点检测

##### 2.2.2基于距离的离群点检测

##### 2.2.3局部离群点算法

#### 2.3不平衡数据级联算法

#### 2.4讨论题目

#### 2.5推荐阅读

### 第3章有指导的学习

#### 3.1有指导的学习概述

#### 3.2K—近邻

#### 3.3决策树

##### 3.3.1决策树的基本概念

##### 3.3.2分类回归树

##### 3.3.3决策树的剪枝

#### 3.4提升方法

#### 3.5随机森林树

##### 3.5.1随机森林树算法的定义

##### 3.5.2如何确定随机森林树算法中树的节点分裂变量

##### 3.5.3随机森林树的回归算法

#### 3.6人工神经网络

- 3.6.1 人工神经网络基本概念
- 3.6.2 感知器算法
- 3.6.3 LMS 算法
- 3.6.4 反向传播算法
- 3.6.5 神经网络相关问题讨论
- 3.7 支持向量机
- 3.7.1 最大边距分类
- 3.7.2 支持向量机问题的求解
- 3.7.3 支持向量机的核方法
- 3.8 多元自适应回归样条
- 3.9 讨论题目
- 3.10 推荐阅读
- 第4章 无指导的学习
- 4.1 关联规则
- 4.1.1 静态关联规则算法 Apriori 算法
- 4.1.2 动态关联规则算法 Carma 算法
- 4.1.3 序列规则挖掘算法
- 4.2 聚类分析
- 4.2.1 聚类分析的含义及作用
- 4.2.2 距离的定义
- 4.2.3 系统层次聚类法
- 4.2.4 K—均值算法
- 4.2.5 BIRCH 算法
- 4.2.6 基于密度的聚类算法
- 4.3 基于预测强度的聚类方法
- 4.3.1 预测强度
- 4.3.2 预测强度方法的应用
- 4.3.3 案例分析
- 4.4 聚类问题的变量选择
- 4.4.1 高斯成对罚模型聚类
- 4.4.2 各类异方差成对罚模型聚类
- 4.4.3 几种聚类变量选择的比较
- 4.5 讨论题目
- 4.6 推荐阅读
- 第5章 贝叶斯分类和因果学习
- 5.1 贝叶斯分类
- 5.2 决策论与统计决策论
- 5.2.1 决策与风险
- 5.2.2 统计决策
- 5.3 线性判别函数和二次判别函数
- 5.4 朴素贝叶斯分类
- 5.5 贝叶斯网络
- 5.5.1 基本概念
- 5.5.2 贝叶斯网络的应用
- 5.5.3 贝叶斯网络的构建
- 5.6 案例：贝叶斯网络模型在信用卡违约概率建模中的应用
- 5.7 讨论题目
- 5.8 推荐阅读
- 第6章 高维回归及变量选择

- 6.1 线性回归模型
  - 6.2 模型选择
    - 6.2.1 模型选择概述
    - 6.2.2 偏差，方差分解
    - 6.2.3 模型选择准则
    - 6.2.4 回归变量选择
  - 6.3 广义线性模型
    - 6.3.1 二点分布回归
    - 6.3.2 指数族概率分布
    - 6.3.3 广义线性模型
    - 6.3.4 模型估计
    - 6.3.5 模型检验与诊断
  - 6.4 高维回归系数压缩
    - 6.4.1 岭回归
    - 6.4.1 LASSO
    - 6.4.3 Shooting 算法
    - 6.4.4 路径算法
    - 6.4.5 其他惩罚项及Oracle性质
    - 6.4.6 软件实现
  - 6.5 总结
  - 6.6 讨论题目
  - 6.7 推荐阅读
- ## 第7章 图模型
- 7.1 图模型基本概念和性质
    - 7.1.1 图矩阵
    - 7.1.2 概率图模型概念和性质
  - 7.2 协方差选择
    - 7.2.1 用回归估计图模型
    - 7.2.2 基于最大似然框架的方法
  - 7.3 指数族图模型
    - 7.3.1 基本定义
    - 7.3.2 参数估计及假设检验
  - 7.4 谱聚类
    - 7.4.1 聚类和图划分
    - 7.4.2 谱聚类
  - 7.5 总结
  - 7.6 讨论题目
  - 7.7 推荐阅读
- ## 第8章 客户关系管理
- 8.1 协同推荐模型
    - 8.1.1 基于邻域的算法
    - 8.1.2 矩阵分解模型
  - 8.2 客户价值随机模型
    - 8.2.1 客户价值的定义
    - 8.2.2 客户价值分析模型
    - 8.2.3 客户购买状态转移矩阵
    - 8.2.4 利润矩阵
    - 8.2.5 客户价值的计算
  - 8.3 案例：银行卡消费客户价值模型

8.4推荐阅读

第9章 社会网络分析

9.1 社会网络概述

9.1.1 社会网络概念与发展

9.1.2 社会网络的基本特征

9.1.3 社群挖掘算法

9.1.4 模型的评价

9.2 案例：社会网络在学术机构合作关系上的研究

9.3 讨论题目

9.4 推荐阅读

附录A 本章R程序

第10章 自然语言模型和文本挖掘

10.1 向量空间模型

10.1.1 向量空间模型基本概念

10.1.2 特征选择准则

10.2 统计语言模型

10.2.1 n-gram模型

10.2.2 主题n-元模型

10.3 LDA模型

10.4 案例：LDA模型的热点新闻发现

10.5 推荐阅读

# 《大数据分析：方法与应用》

## 版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:[www.tushu000.com](http://www.tushu000.com)