

《R语言实战——机器学习与数据分析》

图书基本信息

书名：《R语言实战——机器学习与数据分析》

13位ISBN编号：9787121286696

出版时间：2016-5

作者：左飞

页数：392

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.tushu111.com

《R语言实战——机器学习与数据分析》

内容概要

经典统计理论和机器学习方法为数据挖掘提供了必要的分析技术。《R语言实战——机器学习与数据分析》系统地介绍统计分析和机器学习领域中最重要和流行的多种技术及其基本原理，在详解有关算法的基础上，结合大量R语言实例演示了这些理论在实践中的使用方法。具体内容被分成三个部分，即R语言编程基础、基于统计的数据分析方法以及机器学习理论。统计分析与机器学习部分又具体介绍了参数估计、假设检验、极大似然估计、非参数检验方法（包括列联分析、符号检验、符号秩检验等）、方差分析、线性回归（包括岭回归和Lasso方法）、逻辑回归、支持向量机、聚类分析（包括K均值算法和EM算法）和人工神经网络等内容。同时，统计理论的介绍也为深化读者对于后续机器学习部分的理解提供了很大助益。知识结构和阅读进度的安排上既兼顾了循序渐进的学习规律，亦统筹考虑了夯实基础的必要性。《R语言实战——机器学习与数据分析》内容与实际应用结合紧密，又力求突出深入浅出、系统翔实之特色，对算法原理解释更是细致入微。

《R语言实战——机器学习与数据分析》非常适合大专院校相关专业师生自学研究之用，亦可作为数据分析和数据挖掘相关领域从业人员的参考指导用书。

书籍目录

第1章 初识R语言

1

1.1 R语言简介

1

1.2 安装与运行

3

1.3 开始使用R

5

1.4 包的使用

7

1.5 使用帮助

8

第2章 探索R数据

10

2.1 向量的创建

10

2.2 向量的运算

13

2.3 向量的筛选

15

2.4 矩阵的创建

17

2.5 矩阵的使用

20

2.5.1 矩阵的代数运算

20

2.5.2 修改矩阵的行列

22

2.5.3 对行列调用函数

23

2.6 矩阵的筛选

25

第3章 编写R程序

28

3.1 流程的控制

28

3.1.1 条件选择结构的概念

28

3.1.2 条件选择结构的语法

29

3.1.3 循环结构的基本概念

30

3.1.4 循环结构的基本语法

31

3.2 算术与逻辑

33

3.3 使用函数

34	
3.3.1	函数式语言
34	
3.3.2	默认参数值
35	
3.3.3	自定义函数
36	
3.3.4	递归的实现
38	
3.4	编写代码
40	
第4章	概率统计基础
42	
4.1	概率论的基本概念
42	
4.2	随机变量数字特征
45	
4.2.1	期望
45	
4.2.2	方差
46	
4.3	基本概率分布模型
48	
4.3.1	离散概率分布
48	
4.3.2	连续概率分布
52	
4.3.3	使用内嵌分布
55	
4.4	大数定理及其意义
59	
4.5	中央极限定理
62	
4.6	随机采样分布
65	
第5章	实用统计图形
71	
5.1	饼状图
71	
5.2	直方图
74	
5.3	核密图
78	
5.4	箱线图
81	
5.4.1	箱线图与分位数
81	
5.4.2	使用并列箱线图
84	

5.5 条形图

87

5.5.1 基本条形图及调整

87

5.5.2 堆砌与分组条形图

88

5.6 分位数与QQ图

91

第6章 数据输入/输出

99

6.1 数据的载入

99

6.1.1 基本的数据导入方法

99

6.1.2 处理其他软件的格式

103

6.1.3 读取来自网页的数据

104

6.1.4 从数据库中读取数据

106

6.2 数据的保存

108

6.3 数据预处理

109

6.3.1 常用数学函数

110

6.3.2 修改数据标签

113

6.3.3 缺失值的处理

114

第7章 高级数据结构

118

7.1 列表

118

7.1.1 列表的创建

118

7.1.2 列表元素的访问

120

7.1.3 增删列表元素

121

7.1.4 拼接列表

123

7.1.5 列表转化为向量

123

7.1.6 列表上的运算

124

7.1.7 列表的递归

125

7.2 数据框

126	
7.2.1	数据框的创建
126	
7.2.2	数据框元素的访问
128	
7.2.3	提取子数据框
129	
7.2.4	数据框行列的添加
130	
7.2.5	数据框的合并
132	
7.2.6	数据框的其他操作
134	
7.3	因子
135	
7.3.1	因子的创建
136	
7.3.2	因子中插入水平
137	
7.3.3	因子和常用函数
138	
7.4	表
140	
7.4.1	表的创建
141	
7.4.2	表中元素的访问
143	
7.4.3	表中变量的边际值
143	
第8章	统计推断
146	
8.1	参数估计
146	
8.1.1	参数估计的基本原理
146	
8.1.2	单总体参数区间估计
149	
8.1.3	双总体均值差的估计
155	
8.1.4	双总体比例差的估计
161	
8.2	假设检验
162	
8.2.1	基本概念
162	
8.2.2	两类错误
166	
8.2.3	均值检验
167	

8.3 极大似然估计	172
8.3.1 极大似然法的基本原理	172
8.3.2 求极大似然估计的方法	174
8.3.3 极大似然估计应用举例	176
第9章 非参数检验方法	181
9.1 列联分析	181
9.1.1 类别数据与列联表	181
9.1.2 皮尔逊 (Pearson) 的卡方检验	182
9.1.3 列联分析应用条件	186
9.1.4 费希尔 (Fisher) 的确切检验	188
9.2 符号检验	190
9.3 威尔科克森 (Wilcoxon) 符号秩检验	195
9.4 威尔科克森 (Wilcoxon) 的秩和检验	199
9.5 克鲁斯卡尔-沃利斯 (Kruskal-Wallis) 检验	204
第10章 一元线性回归	208
10.1 回归分析的性质	208
10.2 回归的基本概念	210
10.2.1 总体的回归函数	210
10.2.2 随机干扰的意义	211
10.2.3 样本的回归函数	213
10.3 回归模型的估计	214
10.3.1 普通最小二乘法原理	214
10.3.2 一元线性回归的应用	216
10.3.3 经典模型的基本假定	218
10.3.4 总体方差的无偏估计	

222	
10.3.5	估计参数的概率分布
225	
10.4	正态条件下的模型检验
227	
10.4.1	拟合优度的检验
227	
10.4.2	整体性假定检验
231	
10.4.3	单个参数的检验
233	
10.5	一元线性回归模型预测
234	
10.5.1	点预测
234	
10.5.2	区间预测
235	
第11章	线性回归进阶
239	
11.1	多元线性回归模型
239	
11.2	多元回归模型估计
241	
11.2.1	最小二乘估计量
241	
11.2.2	多元回归的实例
242	
11.2.3	总体参数估计量
245	
11.3	多元回归模型检验
247	
11.3.1	线性回归的显著性
247	
11.3.2	回归系数的显著性
249	
11.4	多元线性回归模型预测
250	
11.5	其他回归模型函数形式
253	
11.5.1	双对数模型以及生产函数
253	
11.5.2	倒数模型与菲利普斯曲线
255	
11.5.3	多项式回归模型及其分析
258	
11.6	回归模型的评估与选择
260	
11.6.1	嵌套模型选择
261	

11.6.2 赤池信息准则	262
11.6.3 逐步回归方法	265
11.7 现代回归方法的新进展	269
11.7.1 多重共线性	269
11.7.2 岭回归	270
11.7.3 从岭回归到Lasso	271
第12章 方差分析方法	275
12.1 方差分析的基本概念	275
12.2 单因素方差分析方法	278
12.2.1 基本原理	278
12.2.2 分析步骤	279
12.2.3 强度测量	280
12.3 双因素方差分析方法	281
12.3.1 无交互作用的分析	281
12.3.2 有交互作用的分析	286
12.4 多重比较	289
12.4.1 多重t检验	290
12.4.2 Dunnett检验	291
12.4.3 Tukey的HSD检验	294
12.4.4 Newman-Keuls检验	298
12.5 方差齐性的检验方法	301
12.5.1 Bartlett检验法	301
12.5.2 Levene检验法	303
第13章 聚类分析	307
13.1 聚类的概念	

307	
13.2 K均值算法	
308	
13.2.1 距离度量	
309	
13.2.2 算法描述	
310	
13.2.3 应用实例	
312	
13.3 最大期望算法	
314	
13.3.1 算法原理	
314	
13.3.2 收敛探讨	
319	
13.4 高斯混合模型	
320	
13.4.1 模型推导	
320	
13.4.2 应用实例	
323	
第14章 支持向量机	
326	
14.1 从逻辑回归到线性分类	
326	
14.2 线性可分的支持向量机	
330	
14.2.1 函数距离与几何距离	
330	
14.2.2 最大间隔分类器	
332	
14.2.3 拉格朗日乘数法	
334	
14.2.4 对偶问题的求解	
339	
14.3 松弛因子与软间隔模型	
343	
14.4 非线性支持向量机方法	
345	
14.4.1 从更高维度上分类	
345	
14.4.2 非线性核函数方法	
347	
14.4.3 默瑟定理与核函数	
350	
14.5 对数据进行分类的实践	
350	
14.5.1 基本建模函数	
351	

14.5.2 分析建模结果	355
第15章 人工神经网络	358
15.1 从感知机开始	358
15.1.1 感知机模型	358
15.1.2 感知机学习	360
15.1.3 多层感知机	362
15.2 基本神经网络	365
15.2.1 神经网络结构	365
15.2.2 符号标记说明	366
15.2.3 后向传播算法	368
15.3 神经网络实践	370
15.3.1 核心函数介绍	370
15.3.2 应用分析实践	372
参考文献	375

《R语言实战——机器学习与数据分析》

精彩短评

1、本书出版除了增添电子工业出版社的业务绩效外毫无意义。本书主要内容中前12章和方匡南同系列的那本书中的内容完全重复，11章“现代回归方法的新进展”本应是区别方著的亮点，作者仅用三页就介绍完了。13、14、15这三章与方著不同，但是写得很粗略，关于分类和判别可参考何晓群的《现代统计分析方法与应用》，关于神经网络部分可参考吴喜之的新书。本书非常不值看。

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.tushu111.com