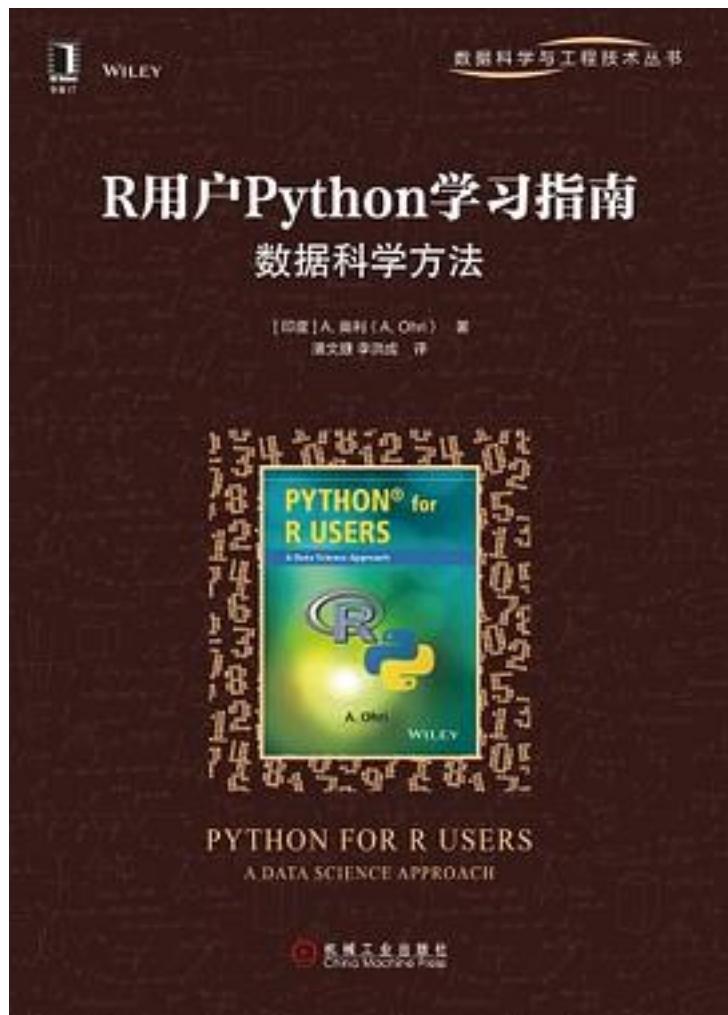


R用户Python学习指南：数据科学方法



[R用户Python学习指南：数据科学方法 下载链接1](#)

著者:[印度] A.奥利 (A.Ohri) 著

出版者:机械工业出版社

出版时间:2018-11-6

装帧:平装-胶订

isbn:9787111611950

掌握一门语言、一种工具已不足以应对当今的数据分析任务，本书旨在解决这一问题，

针对了解R但不熟悉Python（或了解Python但不熟悉R）的从业人员，详解讲解了两种语言的编程技巧和转换方法，提供大量实战案例，不讨论理论细节。

作者介绍：

[印度]A. 奥利 (A. Ohri) 著：A. 奥利 (A. Ohri)

Decisionstats.com的创始人，资深数据科学家，访问量超过100万次的知名博主。他曾为多家创业公司提供咨询服务，涉及的领域包括数据分析服务、数据分析教育和数据项目外包等。目前的研究兴趣涵盖开源分析、社交媒体分析、云计算交互等。他还著有《R for Business Analytics》和《R for Cloud Computing》。

目录: 译者序

前言

第1章 Python、R和R数据科学简介1

1.1 什么是Python1

1.2 什么是R2

1.3 什么是数据科学2

1.4 数据科学家的未来3

1.5 什么是大数据4

1.6 商务分析与数据科学5

1.7 数据科学家可用的工具6

1.8 用于数据科学的Python添加包7

1.9 Python和R之间的异同8

1.9.1 为什么R用户应该了解有关Python的更多信息8

1.9.2 为什么Python用户应该了解有关R的更多信息9

1.10 教程9

1.11 同时使用R和Python9

1.12 其他软件和Python13

1.13 将SAS与Jupyter一起使用13

1.14 如何将Python和R用于大数据分析13

1.15 什么是云计算15

1.16 如何在云上使用Python和R15

1.17 Python和R商业版本及其他替代版本16

1.17.1 数据科学家的常用Linux命令17

1.17.2 学习Git18

1.18 数据驱动的决策32

1.18.1 企业管理战略框架：非MBA和MBA的数据驱动决策指南32

1.18.2 商务分析的其他框架37

参考文献40

第2章 数据输入41

2.1 pandas中的数据输入41

2.2 网页抓取数据输入44

2.3 来自RDBMS的数据输入48

2.3.1 Windows教程50

2.3.2 安装50

2.3.3 配置ODBC52

第3章 数据检查和数据质量62

3.1 数据格式62

3.1.1 将字符串转换为日期和时间63

3.1.2 将数据框转换为NumPy数组并返回Python66

3.2 数据质量68

3.3 数据检查71

3.4 数据选择	74
3.4.1 随机选择数据	76
3.4.2 条件选择	76
3.5 R中的数据检查	79
3.5.1 来自R ggplot2添加包中的diamond数据集	83
3.5.2 修改R中的日期格式和字符串	86
3.5.3 管理R中的字符串	88
参考文献	88
第4章 探索性数据分析	89
4.1 根据分析分组	89
4.2 数值数据	89
4.3 分类数据	91
第5章 统计建模	102
5.1 回归的概念	102
5.1.1 OLS	103
5.1.2 R方	103
5.1.3 p值	104
5.1.4 异常值	104
5.1.5 多重共线性和异方差性	104
5.2 相关关系不是因果关系	105
5.2.1 关于数据科学家统计的说明	106
5.2.2 中心趋势的度量	107
5.2.3 分散度的度量	107
5.2.4 概率分布	110
5.3 R和Python中的线性回归	114
5.4 R和Python中的Logistic回归	132
5.4.1 其他概念	136
5.4.2 ROC曲线和AUC	137
5.4.3 偏差与方差	137
参考文献	138
第6章 数据可视化	139
6.1 数据可视化的概念	139
6.1.1 数据可视化的历史	139
6.1.2 Anscombe案例研究	140
6.1.3 加载添加包	141
6.1.4 获得均值和标准差	142
6.1.5 结论	144
6.1.6 数据可视化	144
6.1.7 结论	2146
6.2 Tufte关于数据可视化的工作	147
6.3 Stephen Few关于仪表盘的设计	147
6.4 基本绘图	149
6.5 高级绘图	156
6.6 交互式绘图	160
6.7 空间分析	160
6.8 R中的数据可视化	160
6.8.1 关于RStudio IDE共享R代码的注意事项	166
6.8.2 关于共享Jupyter笔记的注意事项	167
6.8.3 特别说明：关于Python的完整教程	168
参考文献	180
第7章 机器学习变得更容易	181
7.1 删除最终决策树模型中不需要的列	186
7.2 时间序列	207
7.3 关联分析	211

7.4 清洗语料库并制作词袋221

7.4.1 聚类分析224

7.4.2 Python中的聚类分析224

第8章 结论和总结233

• • • • • (收起)

[R用户Python学习指南：数据科学方法](#) [_下载链接1](#)

标签

数据科学

R

Python

评论

[R用户Python学习指南：数据科学方法](#) [_下载链接1](#)

书评

[R用户Python学习指南：数据科学方法](#) [_下载链接1](#)