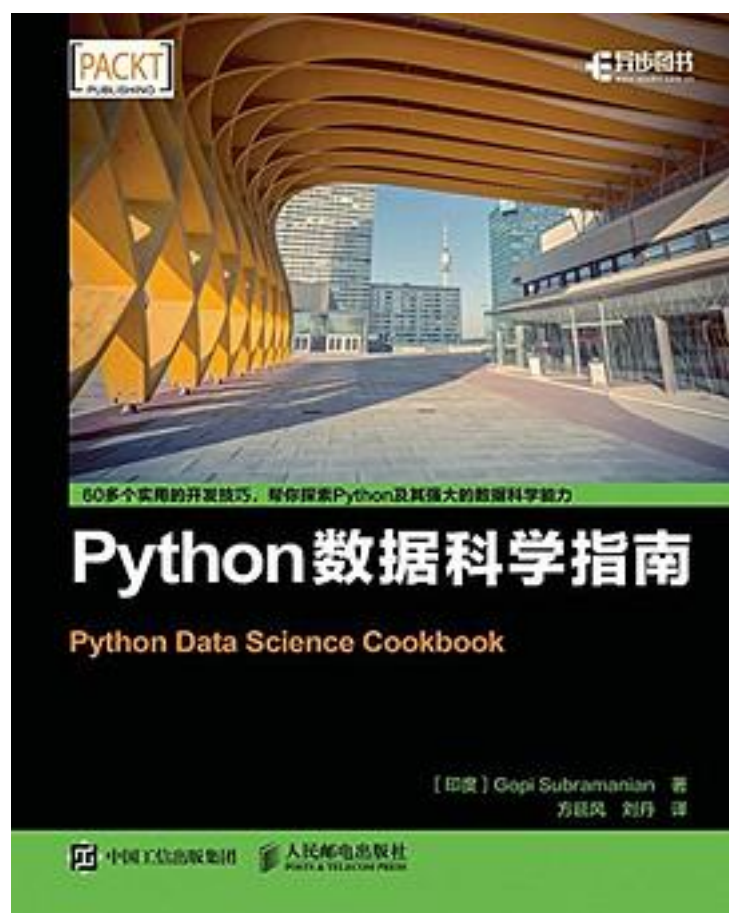


Python数据科学指南



[Python数据科学指南_下载链接1_](#)

著者:印度 Gopi Subramanian 萨伯拉曼尼安

出版者:人民邮电出版社

出版时间:2016-11-1

装帧:平装

isbn:9787115435101

本书从讲解如何在数据科学中应用Python开始，陆续介绍了Python的工作环境，如何用Python分析数据，以及数据挖掘的概念，然后又扩展到机器学习。本书还涵盖了缩减原则、集成方法、随机森林、旋转森林和超树等方面的内容，这些都是一个成功的数据科学专家所必需掌握的。

阅读本书，你将学会：

- 揭示数据科学算法的完整范畴；
- 高效地掌握和使用numpy、scipy、scikit-learn和matplotlib等Python库；
- 了解进阶回归方法的建模和变量选择；
- 进一步彻底理解集成方法的潜在含义及实施；
- 在各种各样的数值和文本数据集上解决实际问题；
- 熟悉先进的算法，如梯度提升、随机森林、旋转森林等。

本书特色：

- 内容明确且易于跟学；
- 甄选重要的任务与问题；
- 精心组织编排内容，有效解决问题；
- 清晰易懂的讲解方式；
- 书中呈现的解决方案能够直接应用到实际问题中。

作者介绍:

Gopi

Subramanian是一名数据科学家，他在数据挖掘与机器学习领域有着超过15年经验。在过去的10年中，他设计、构思、开发并领导了数据挖掘、文本挖掘、自然语言处理、信息提取和检索等多个项目，涉及不同领域和商务垂直系统。他在美国和印度的专利局共计申请了10多项专利，并以自己的名义出版了许多书籍。

目录: 第1章Python在数据科学中的应用 1

- 1.1 简介 2
- 1.2 使用字典对象 2
 - 1.2.1 准备工作 2
 - 1.2.2 操作方法 2
 - 1.2.3 工作原理 3
 - 1.2.4 更多内容 4
 - 1.2.5 参考资料 6
- 1.3 使用字典的字典 6
 - 1.3.1 准备工作 6
 - 1.3.2 操作方法 6
 - 1.3.3 工作原理 7
 - 1.3.4 参考资料 7
- 1.4 使用元组 7
 - 1.4.1 准备工作 7
 - 1.4.2 操作方法 8
 - 1.4.3 工作原理 9
 - 1.4.4 更多内容 12

1.4.5 参考资料	12
1.5 使用集合	13
1.5.1 准备工作	13
1.5.2 操作方法	13
1.5.3 工作原理	14
1.5.4 更多内容	15
1.6 写一个列表	16
1.6.1 准备工作	16
1.6.2 操作方法	16
1.6.3 工作原理	18
1.6.4 更多内容	19
1.7 从另一个列表创建列表——列表推导	20
1.7.1 准备工作	20
1.7.2 操作方法	20
1.7.3 工作原理	20
1.7.4 更多内容	21
1.8 使用迭代器	22
1.8.1 准备工作	22
1.8.2 操作方法	23
1.8.3 工作原理	23
1.8.4 更多内容	24
1.9 生成一个迭代器和生成器	24
1.9.1 准备工作	25
1.9.2 操作方法	25
1.9.3 工作原理	25
1.9.4 更多内容	25
1.10 使用可迭代对象	26
1.10.1 准备工作	26
1.10.2 操作方法	26
1.10.3 工作原理	27
1.10.4 参考资料	27
1.11 将函数作为变量传递	28
1.11.1 准备工作	28
1.11.2 操作方法	28
1.11.3 工作原理	28
1.12 在函数中嵌入函数	28
1.12.1 准备工作	29
1.12.2 操作方法	29
1.12.3 工作原理	29
1.13 将函数作为参数传递	29
1.13.1 准备工作	29
1.13.2 操作方法	29
1.13.3 工作原理	30
1.14 返回一个函数	30
1.14.1 准备工作	31
1.14.2 操作方法	31
1.14.3 工作原理	31
1.14.4 更多内容	32
1.15 使用装饰器改变函数行为	32
1.15.1 准备工作	32
1.15.2 操作方法	32
1.15.3 工作原理	33
1.16 使用lambda创造匿名函数	34
1.16.1 准备工作	34

1.16.2	操作方法	35
1.16.3	工作原理	35
1.17	使用映射函数	35
1.17.1	准备工作	36
1.17.2	操作方法	36
1.17.3	工作原理	36
1.17.4	更多内容	36
1.18	使用过滤器	37
1.18.1	准备工作	37
1.18.2	操作方法	37
1.18.3	工作原理	38
1.19	使用zip和izip函数	38
1.19.1	准备工作	38
1.19.2	操作方法	38
1.19.3	工作原理	38
1.19.4	更多内容	39
1.19.5	参考资料	40
1.20	从表格数据使用数组	40
1.20.1	准备工作	40
1.20.2	操作方法	41
1.20.3	工作原理	41
1.20.4	更多内容	42
1.21	对列进行预处理	43
1.21.1	准备工作	44
1.21.2	操作方法	44
1.21.3	工作原理	45
1.21.4	更多内容	45
1.22	列表排序	46
1.22.1	准备工作	46
1.22.2	操作方法	46
1.22.3	工作原理	46
1.22.4	更多内容	47
1.23	采用键排序	47
1.23.1	准备工作	48
1.23.2	操作方法	48
1.23.3	工作原理	49
1.23.4	更多内容	49
1.24	使用itertools	52
1.24.1	准备工作	52
1.24.2	操作方法	52
1.24.3	工作原理	53
第2章	Python环境	55
2.1	简介	55
2.2	使用NumPy库	55
2.2.1	准备工作	55
2.2.2	操作方法	56
2.2.3	工作原理	58
2.2.4	更多内容	64
2.2.5	参考资料	64
2.3	使用matplotlib进行绘画	64
2.3.1	准备工作	64
2.3.2	操作方法	64
2.3.3	工作原理	66
2.3.4	更多内容	72

2.4 使用scikit—learn进行机器学习 73

2.4.1 准备工作 73

2.4.2 操作方法 73

2.4.3 工作原理 75

2.4.4 更多内容 81

2.4.5 参考资料 82

第3章 数据分析——探索与争鸣 83

3.1 简介 84

3.2 用图表分析单变量数据 85

3.2.1 准备工作 85

3.2.2 操作方法 86

3.2.3 工作原理 87

3.2.4 参考资料 92

3.3 数据分组和使用点阵图 92

3.3.1 准备工作 93

3.3.2 操作方法 93

3.3.3 工作原理 95

3.3.4 参考资料 97

3.4 为多变量数据绘制散点阵图 97

3.4.1 准备工作 98

3.4.2 操作方法 98

3.4.3 工作原理 99

3.4.4 参考资料 100

3.5 使用热图 101

3.5.1 准备工作 101

3.5.2 操作方法 101

3.5.3 工作原理 102

3.5.4 更多内容 104

3.5.5 参考资料 105

3.6 实施概要统计及绘图 105

3.6.1 准备工作 105

3.6.2 操作方法 106

3.6.3 工作原理 107

3.6.4 参考资料 110

3.7 使用箱须图 110

3.7.1 准备工作 110

3.7.2 操作方法 110

3.7.3 工作原理 111

3.7.4 更多内容 112

3.8 修补数据 113

3.8.1 准备工作 113

3.8.2 操作方法 113

3.8.3 工作原理 114

3.8.4 更多内容 115

3.8.5 参考资料 116

3.9 实施随机采样 116

3.9.1 准备工作 116

3.9.2 操作方法 117

3.9.3 工作原理 117

3.9.4 更多内容 118

3.10 缩放数据 118

3.10.1 准备工作 118

3.10.2 操作方法 118

3.10.3 工作原理 119

3.10.4 更多内容	119
3.11 数据标准化	121
3.11.1 准备工作	121
3.11.2 操作方法	121
3.11.3 工作原理	122
3.11.4 更多内容	122
3.12 实施分词化	123
3.12.1 准备工作	123
3.12.2 操作方法	123
3.12.3 工作原理	124
3.12.4 更多内容	125
3.12.5 参考资料	127
3.13 删除停用词	127
3.13.1 操作方法	128
3.13.2 工作原理	129
3.13.3 更多内容	130
3.13.4 参考资料	130
3.14 词提取	130
3.14.1 准备工作	131
3.14.2 操作方法	132
3.14.3 工作原理	132
3.14.4 更多内容	133
3.14.5 参考资料	133
3.15 执行词形还原	134
3.15.1 准备工作	134
3.15.2 操作方法	134
3.15.3 工作原理	135
3.15.4 更多内容	135
3.15.5 参考资料	135
3.16 词袋模型表示文本	136
3.16.1 准备工作	136
3.16.2 操作方法	136
3.16.3 工作原理	138
3.16.4 更多内容	140
3.16.5 参考资料	141
3.17 计算词频和反文档频率	142
3.17.1 准备工作	142
3.17.2 操作方法	142
3.17.3 工作原理	144
3.17.4 更多内容	145
第4章 数据分析——深入理解	146
4.1 简介	146
4.2 抽取主成分	147
4.2.1 准备工作	148
4.2.2 操作方法	149
4.2.3 工作原理	151
4.2.4 更多内容	152
4.2.5 参考资料	154
4.3 使用核PCA	154
4.3.1 准备工作	154
4.3.2 操作方法	154
4.3.3 工作原理	156
4.3.4 更多内容	159
4.4 使用奇异值分解抽取特征	160

4.4.1	准备工作	161
4.4.2	操作方法	161
4.4.3	工作原理	162
4.4.4	更多内容	163
4.5	用随机映射给数据降维	164
4.5.1	准备工作	164
4.5.2	操作方法	165
4.5.3	工作原理	166
4.5.4	更多内容	167
4.5.5	参考资料	168
4.6	用NMF分解特征矩阵	168
4.6.1	准备工作	169
4.6.2	操作方法	170
4.6.3	工作原理	172
4.6.4	更多内容	175
4.6.5	参考资料	176
第5章	数据挖掘——海底捞针	177
5.1	简介	177
5.2	使用距离度量	178
5.2.1	准备工作	178
5.2.2	操作方法	179
5.2.3	工作原理	180
5.2.4	更多内容	183
5.2.5	参考资料	184
5.3	学习和使用核方法	184
5.3.1	准备工作	184
5.3.2	操作方法	185
5.3.3	工作原理	186
5.3.4	更多内容	187
5.3.5	参考资料	187
5.4	用k—means进行数据聚类	188
5.4.1	准备工作	188
5.4.2	操作方法	190
5.4.3	工作原理	191
5.4.4	更多内容	192
5.4.5	参考资料	193
5.5	学习向量量化	193
5.5.1	准备工作	193
5.5.2	操作方法	194
5.5.3	工作原理	197
5.5.4	更多内容	199
5.5.5	参考资料	199
5.6	在单变量数据中找出异常点	200
5.6.1	准备工作	200
5.6.2	操作方法	202
5.6.3	工作原理	203
5.6.4	更多内容	205
5.6.5	参考资料	207
5.7	使用局部异常因子方法发现异常点	207
5.7.1	准备工作	207
5.7.2	操作方法	208
5.7.3	工作原理	210
5.7.4	更多内容	216
第6章	机器学习	1217

6.1 简介	217
6.2 为建模准备数据	218
6.2.1 准备工作	218
6.2.2 操作方法	218
6.2.3 工作原理	221
6.2.4 更多内容	222
6.3 查找最近邻	223
6.3.1 准备工作	224
6.3.2 操作方法	226
6.3.3 工作原理	227
6.3.4 更多内容	229
6.3.5 参考资料	230
6.4 用朴素贝叶斯分类文档	230
6.4.1 准备工作	232
6.4.2 操作方法	232
6.4.3 工作原理	238
6.4.4 更多内容	242
6.4.5 参考资料	242
6.5 构建决策树解决多类问题	243
6.5.1 准备工作	244
6.5.2 操作方法	247
6.5.3 工作原理	249
6.5.4 更多内容	251
6.5.5 参考资料	252
第7章 机器学习	2253
7.1 简介	253
7.2 回归方法预测实数值	254
7.2.1 准备工作	255
7.2.2 操作方法	256
7.2.3 工作原理	259
7.2.4 更多内容	263
7.2.5 参考资料	267
7.3 学习L2缩减回归——岭回归	267
7.3.1 准备工作	268
7.3.2 操作方法	268
7.3.3 工作原理	271
7.3.4 更多内容	273
7.3.5 参考资料	276
7.4 学习L1缩减回归——LASSO	276
7.4.1 准备工作	277
7.4.2 操作方法	277
7.4.3 工作原理	280
7.4.4 更多内容	283
7.4.5 参考资料	283
7.5 L1和L2缩减交叉验证迭代	283
7.5.1 准备工作	284
7.5.2 操作方法	284
7.5.3 工作原理	288
7.5.4 更多内容	294
7.5.5 参考资料	295
第8章 集成方法	296
8.1 简介	296
8.2 理解集成——挂袋法	297
8.2.1 准备工作	298

8.2.2 操作方法	298
8.2.3 工作原理	300
8.2.4 更多内容	304
8.2.5 参考资料	305
8.3 理解集成——提升法	305
8.3.1 准备工作	307
8.3.2 操作方法	307
8.3.3 工作原理	312
8.3.4 更多内容	319
8.3.5 参考资料	319
8.4 理解集成——梯度提升	320
8.4.1 准备工作	321
8.4.2 操作方法	321
8.4.3 工作原理	325
8.4.4 更多内容	330
8.4.5 参考资料	330
第9章 生长树	331
9.1 简介	331
9.2 从生长树到生长森林——随机森林	332
9.2.1 准备工作	333
9.2.2 操作方法	333
9.2.3 工作原理	336
9.2.4 更多内容	340
9.2.5 参考资料	342
9.3 生成超随机树	342
9.3.1 准备工作	343
9.3.2 操作方法	343
9.3.3 工作原理	345
9.3.4 更多内容	349
9.3.5 参考资料	349
9.4 生成旋转森林	349
9.4.1 准备工作	350
9.4.2 操作方法	350
9.4.3 工作原理	353
9.4.4 更多内容	358
9.4.5 参考资料	358
第10章 大规模机器学习——在线学习	359
10.1 简介	359
10.2 用感知器作为在线学习算法	360
10.2.1 准备工作	361
10.2.2 操作方法	362
10.2.3 工作原理	363
10.2.4 更多内容	366
10.2.5 参考资料	367
10.3 用随机梯度下降解决回归问题	367
10.3.1 准备工作	369
10.3.2 操作方法	369
10.3.3 工作原理	370
10.3.4 更多内容	373
10.3.5 参考资料	375
10.4 用随机梯度下降解决分类问题	375
10.4.1 准备工作	376
10.4.2 操作方法	376
10.4.3 工作原理	377

10.4.4 更多内容 379
10.4.5 参考资料 380
• • • • • ([收起](#))

[Python数据科学指南_下载链接1](#)

标签

Python

数据科学

数据分析

计算机

编程

数据

工作-数据分析/数据科学/商业智能

python

评论

关于代码入门和培养DS的代码习惯是挺不错的，但是涉及理论公式比较少，在调用sklearn库的时候，一些详细参数还是缺乏全面的解释，建议配合相关ML理论书籍一起看，可能效果会更好。此书只限DS入门代码，不是深入的那种，不过挺全的。我做的笔记也放在博客里了，有兴趣的朋友可以看看，相互交流：https://blog.csdn.net/qq_33704653/article/category/7586265

Python进阶全靠这本书，好适合我这种半吊子老白，推荐！

[Python数据科学指南_下载链接1](#)

书评

[Python数据科学指南_下载链接1](#)