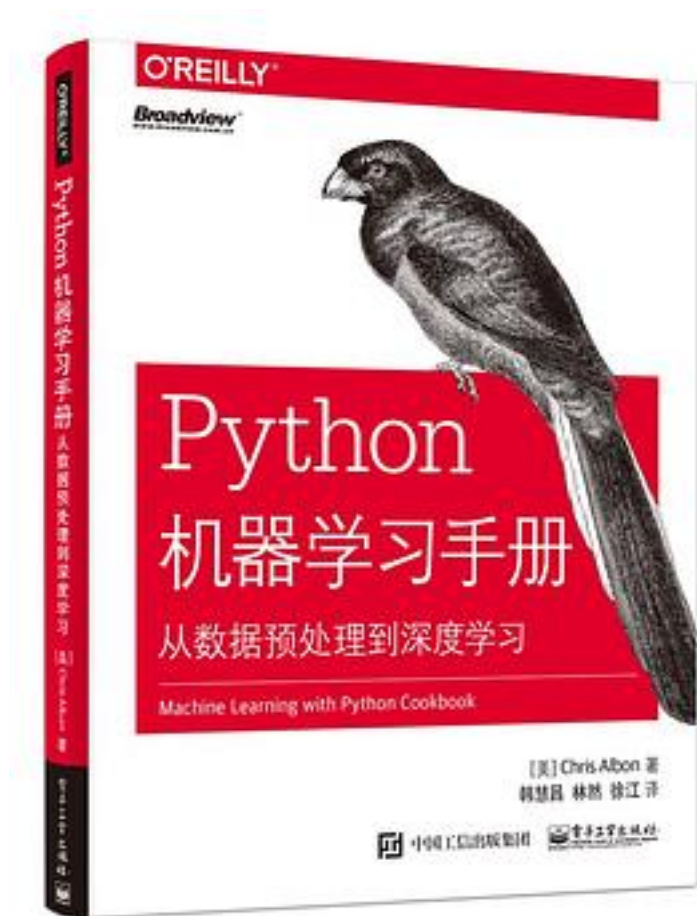


Python机器学习手册：从数据预处理到深度学习



[Python机器学习手册：从数据预处理到深度学习_下载链接1](#)

著者:【美】克里斯·阿尔本 (Chris Albon)

出版者:电子工业出版社

出版时间:2019-7

装帧:平装

isbn:9787121369629

《Python机器学习手册：从数据预处理到深度学习》采用基于任务的方式来介绍如何在机器学习中使用Python。书中有近200个独立的解决方案，针对的都是数据科学家或机器学习工程师在构建模型时可能遇到的常见任务，涵

盖从简
单的矩阵和向量运算到特征工程以及神经网络的构建。所有方案都提供了相关代码，读者可以复制并粘贴这些代码，用在自己的程序中。

《Python机器学习手册：从数据预处理到深度学习》不是机器学习的入门书，适合熟悉机器学习理论和概念的读者阅读。你可以将本书作为案头参考书，在机器学习的日常开发中遇到问题时，随时借鉴书中代码，快速解决问题。

作者介绍:

目录: 第1章 向量、矩阵和数组	1
1.0 简介	1
1.1 创建一个向量	1
1.2 创建一个矩阵	2
1.3 创建一个稀疏矩阵	3
1.4 选择元素	5
1.5 展示一个矩阵的属性	6
1.6 对多个元素同时应用某个操作	7
1.7 找到最大值和最小值	8
1.8 计算平均值、方差和标准差	9
1.9 矩阵变形	10
1.10 转置向量或矩阵	11
1.11 展开一个矩阵	12
1.12 计算矩阵的秩	13
1.13 计算行列式	14
1.14 获取矩阵的对角线元素	14
1.15 计算矩阵的迹	15
1.16 计算特征值和特征向量	16
1.17 计算点积	17
1.18 矩阵的相加或相减	18
1.19 矩阵的乘法	19
1.20 计算矩阵的逆	20
1.21 生成随机数	21
第2章 加载数据	23
2.0 简介	23
2.1 加载样本数据集	23
2.2 创建仿真数据集	25
2.3 加载CSV 文件	28
2.4 加载Excel 文件	29
2.5 加载JSON 文件	29
2.6 查询SQL 数据库	31
第3章 数据整理	33
3.0 简介	33
3.1 创建一个数据帧	34
3.2 描述数据	35
3.3 浏览数据帧	37
3.4 根据条件语句来选择行	39
3.5 替换值	40
3.6 重命名列	41
3.7 计算最小值、最大值、总和、平均值与计数值	43
3.8 查找唯一值	44

3.9 处理缺失值	45
3.10 删除一列	47
3.11 删除一行	48
3.12 删除重复行	49
3.13 根据值对行分组	51
3.14 按时间段对行分组	52
3.15 遍历一个列的数据	54
3.16 对一列的所有元素应用某个函数	55
3.17 对所有分组应用一个函数	56
3.18 连接多个数据帧	57
3.19 合并两个数据帧	59
第4章 处理数值型数据	63
4.0 简介	63
4.1 特征的缩放	63
4.2 特征的标准化	65
4.3 归一化观察值	66
4.4 生成多项式和交互特征	69
4.5 转换特征	70
4.6 识别异常值	71
4.7 处理异常值	73
4.8 将特征离散化	75
4.9 使用聚类的方式将观察值分组	77
4.10 删除带有缺失值的观察值	79
4.11 填充缺失值	81
第5章 处理分类数据	83
5.0 简介	83
5.1 对nominal 型分类特征编码	84
5.2 对ordinal 分类特征编码	86
5.3 对特征字典编码	88
5.4 填充缺失的分类值	91
5.5 处理不平衡分类	93
第6章 处理文本	97
6.0 简介	97
6.1 清洗文本	97
6.2 解析并清洗HTML	99
6.3 移除标点	100
6.4 文本分词	101
6.5 删除停止词 (stop word)	102
6.6 提取词干	103
6.7 标注词性	104
6.8 将文本编码成词袋 (Bag of Words)	107
6.9 按单词的重要性加权	109
第7章 处理日期和时间	113
7.0 简介	113
7.1 把字符串转换成日期	113
7.2 处理时区	115
7.3 选择日期和时间	116
7.4 将日期数据切分成多个特征	117
7.5 计算两个日期之间的时间差	118
7.6 对一周内的各天进行编码	119
7.7 创建一个滞后的特征	120
7.8 使用滚动时间窗口	121
7.9 处理时间序列中的缺失值	123
第8章 图像处理	127

8.0 简介	127
8.1 加载图像	128
8.2 保存图像	130
8.3 调整图像大小	131
8.4 裁剪图像	132
8.5 平滑处理图像	133
8.6 图像锐化	136
8.7 提升对比度	138
8.8 颜色分离	140
8.9 图像二值化	142
8.10 移除背景	144
8.11 边缘检测	148
8.12 角点检测	150
8.13 为机器学习创建特征	153
8.14 将颜色平均值编码成特征	156
8.15 将色彩直方图编码成特征	157
第9章 利用特征提取进行特征降维	161
9.0 简介	161
9.1 使用主成分进行特征降维	161
9.2 对线性不可分数据进行特征降维	164
9.3 通过最大化类间可分性进行特征降维	166
9.4 使用矩阵分解法进行特征降维	169
9.5 对稀疏数据进行特征降维	170
第10章 使用特征选择进行降维	173
10.0 简介	173
10.1 数值型特征方差的阈值化	173
10.2 二值特征的方差阈值化	175
10.3 处理高度相关性的特征	176
10.4 删除与分类任务不相关的特征	178
10.5 递归式特征消除	180
第11章 模型评估	183
11.0 简介	183
11.1 交叉验证模型	183
11.2 创建一个基准回归模型	187
11.3 创建一个基准分类模型	188
11.4 评估二元分类器	190
11.5 评估二元分类器的阈值	193
11.6 评估多元分类器	197
11.7 分类器性能的可视化	198
11.8 评估回归模型	201
11.9 评估聚类模型	203
11.10 创建自定义评估指标	204
11.11 可视化训练集规模的影响	206
11.12 生成对评估指标的报告	208
11.13 可视化超参数值的效果	209
第12章 模型选择	213
12.0 简介	213
12.1 使用穷举搜索选择最佳模型	213
12.2 使用随机搜索选择最佳模型	216
12.3 从多种学习算法中选择最佳模型	218
12.4 将数据预处理加入模型选择过程	220
12.5 用并行化加速模型选择	221
12.6 使用针对特定算法的方法加速模型选择	223
12.7 模型选择后的性能评估	224

第13章 线性回归	227
13.0 简介	227
13.1 拟合一条直线	227
13.2 处理特征之间的影响	229
13.3 拟合非线性关系	231
13.4 通过正则化减少方差	233
13.5 使用套索回归减少特征	235
第14章 树和森林	237
14.0 简介	237
14.1 训练决策树分类器	237
14.2 训练决策树回归模型	239
14.3 可视化决策树模型	240
14.4 训练随机森林分类器	243
14.5 训练随机森林回归模型	244
14.6 识别随机森林中的重要特征	245
14.7 选择随机森林中的重要特征	248
14.8 处理不均衡的分类	249
14.9 控制决策树的规模	250
14.10 通过boosting 提高性能	252
14.11 使用袋外误差 (Out-of-Bag Error) 评估随机森林模型	253
第15章 KNN	255
15.0 简介	255
15.1 找到一个观察值的最近邻	255
15.2 创建一个KNN 分类器	258
15.3 确定最佳的邻域点集的大小	260
15.4 创建一个基于半径的最近邻分类器	261
第16章 逻辑回归	263
16.0 简介	263
16.1 训练二元分类器	263
16.2 训练多元分类器	265
16.3 通过正则化来减小方差	266
16.4 在超大数据集上训练分类器	267
16.5 处理不均衡的分类	269
第17章 支持向量机	271
17.0 简介	271
17.1 训练一个线性分类器	271
17.2 使用核函数处理线性不可分的数据	274
17.3 计算预测分类的概率	278
17.4 识别支持向量	279
17.5 处理不均衡的分类	281
第18章 朴素贝叶斯	283
18.0 简介	283
18.1 为连续的数据训练分类器	284
18.2 为离散数据和计数数据训练分类器	286
18.3 为具有二元特征的数据训练朴素贝叶斯分类器	287
18.4 校准预测概率	288
第19章 聚类	291
19.0 简介	291
19.1 使用K-Means 聚类算法	291
19.2 加速K-Means 聚类	294
19.3 使用Meanshift 聚类算法	295
19.4 使用DBSCAN 聚类算法	296
19.5 使用层次合并聚类算法	298
第20章 神经网络	301

20.0 简介	301
20.1 为神经网络预处理数据	302
20.2 设计一个神经网络	304
20.3 训练一个二元分类器	307
20.4 训练一个多元分类器	309
20.5 训练一个回归模型	311
20.6 做预测	313
20.7 可视化训练历史	315
20.8 通过权重调节减少过拟合	318
20.9 通过提前结束减少过拟合	320
20.10 通过Dropout 减少过拟合	322
20.11 保存模型训练过程	324
20.12 使用k 折交叉验证评估神经网络	326
20.13 调校神经网络	328
20.14 可视化神经网络	331
20.15 图像分类	333
20.16 通过图像增强来改善卷积神经网络的性能	337
20.17 文本分类	339
第21 章 保存和加载训练后的模型	343
21.0 简介	343
21.1 保存和加载scikit-learn 模型	343
21.2 保存和加载Keras 模型	345
• • • • • (收起)	

[Python机器学习手册：从数据预处理到深度学习_下载链接1](#)

标签

机器学习

Python

sklearn

计算机

keras

数据分析

美国

评论

利用sklearn进行机器学习，利用keras进行深度学习，本书门槛在机器学习一众书籍中是难得得低，虽然作者反复强调这是一本技术手册，但是专注掉包教学未也是极好的。

在上图借到了开刷！ //暂时搁一下//和超图一起啃吧//四天里看完，写得好翻得也好~

还可以

工具书

Hmm。。内容没有想象的好，入门玩家看看还行

[Python机器学习手册：从数据预处理到深度学习 下载链接1](#)

书评

[Python机器学习手册：从数据预处理到深度学习 下载链接1](#)