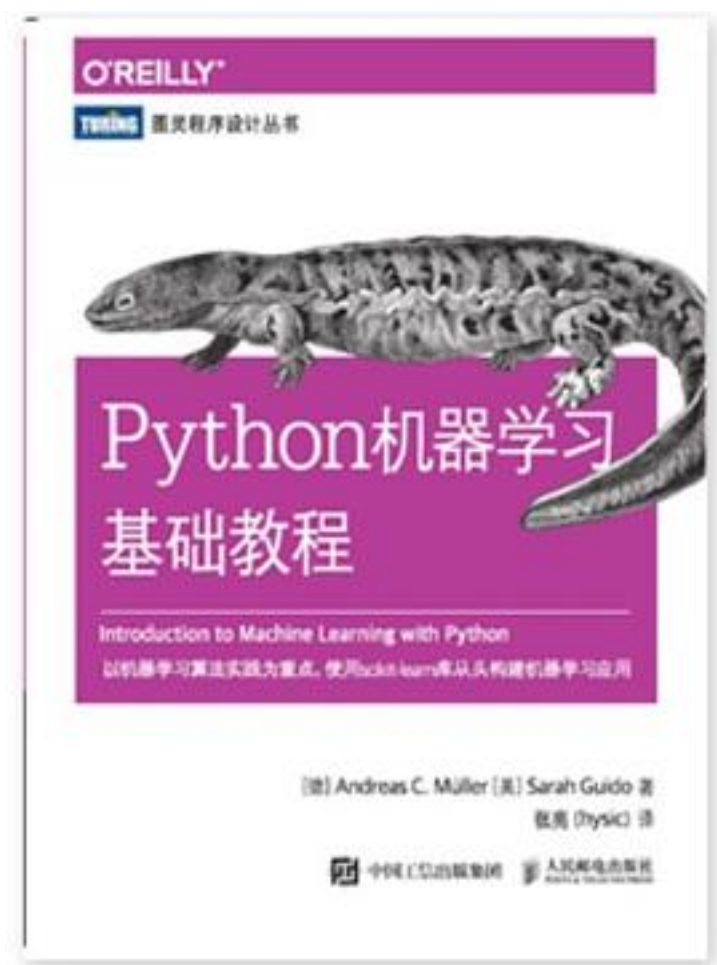


Python机器学习基础教程



[Python机器学习基础教程_下载链接1](#)

著者:[德]安德里亚斯·穆勒

出版者:人民邮电出版社

出版时间:2018-1

装帧:平装

isbn:9787115475619

本书是机器学习入门书，以Python语言介绍。主要内容包括：机器学习的基本概念及其应用；实践中最常用的机器学习算法以及这些算法的优缺点；在机器学习中待处理数

据的呈现方式的重要性，以及应重点关注数据的哪些方面；模型评估和调参的高级方法，重点讲解交叉验证和网格搜索；管道的概念；如何将前面各章的方法应用到文本数据上，还介绍了一些文本特有的处理方法。

作者介绍:

Andreas C. Muller

scikit-learn库维护者和核心贡献者。现任哥伦比亚大学数据科学研究院讲师，曾任纽约大学数据科学中心助理研究员、亚马逊公司计算机视觉应用的机器学习研究员。在波恩大学获得机器学习博士学位。

Sarah Guido

Mashable公司数据科学家，曾担任Bitly公司首席数据科学家。

目录: 前言 ix

第1章 引言 1

1.1 为何选择机器学习 1

1.1.1 机器学习能够解决的问题 2

1.1.2 熟悉任务和数据 4

1.2 为何选择Python 4

1.3 scikit-learn 4

1.4 必要的库和工具 5

1.4.1 Jupyter Notebook 6

1.4.2 NumPy 6

1.4.3 SciPy 6

1.4.4 matplotlib 7

1.4.5 pandas 8

1.4.6 mglearn 9

1.5 Python 2与Python 3的对比 9

1.6 本书用到的版本 10

1.7 第一个应用：鸢尾花分类 11

1.7.1 初识数据 12

1.7.2 衡量模型是否成功：训练数据与测试数据 14

1.7.3 要事第一：观察数据 15

1.7.4 构建第一个模型：k近邻算法 16

1.7.5 做出预测 17

1.7.6 评估模型 18

1.8 小结与展望 19

第2章 监督学习 21

2.1 分类与回归 21

2.2 泛化、过拟合与欠拟合 22

2.3 监督学习算法 24

2.3.1 一些样本数据集 25

2.3.2 k近邻 28

2.3.3 线性模型 35

2.3.4 朴素贝叶斯分类器 53

2.3.5 决策树 54

2.3.6 决策树集成 64

2.3.7 核支持向量机 71

2.3.8 神经网络（深度学习） 80

2.4 分类器的不确定度估计	91
2.4.1 决策函数	91
2.4.2 预测概率	94
2.4.3 多分类问题的不确定度	96
2.5 小结与展望	98
第3章 无监督学习与预处理	100
3.1 无监督学习的类型	100
3.2 无监督学习的挑战	101
3.3 预处理与缩放	101
3.3.1 不同类型的预处理	102
3.3.2 应用数据变换	102
3.3.3 对训练数据和测试数据进行相同的缩放	104
3.3.4 预处理对监督学习的作用	106
3.4 降维、特征提取与流形学习	107
3.4.1 主成分分析	107
3.4.2 非负矩阵分解	120
3.4.3 用t-SNE进行流形学习	126
3.5 聚类	130
3.5.1 k 均值聚类	130
3.5.2 凝聚聚类	140
3.5.3 DBSCAN	143
3.5.4 聚类算法的对比与评估	147
3.5.5 聚类方法小结	159
3.6 小结与展望	159
第4章 数据表示与特征工程	161
4.1 分类变量	161
4.1.1 One-Hot编码（虚拟变量）	162
4.1.2 数字可以编码分类变量	166
4.2 分箱、离散化、线性模型与树	168
4.3 交互特征与多项式特征	171
4.4 单变量非线性变换	178
4.5 自动化特征选择	181
4.5.1 单变量统计	181
4.5.2 基于模型的特征选择	183
4.5.3 迭代特征选择	184
4.6 利用专家知识	185
4.7 小结与展望	192
第5章 模型评估与改进	193
5.1 交叉验证	194
5.1.1 scikit-learn中的交叉验证	194
5.1.2 交叉验证的优点	195
5.1.3 分层k 折交叉验证和其他策略	196
5.2 网格搜索	200
5.2.1 简单网格搜索	201
5.2.2 参数过拟合的风险与验证集	202
5.2.3 带交叉验证的网格搜索	203
5.3 评估指标与评分	213
5.3.1 牢记最终目标	213
5.3.2 二分类指标	214
5.3.3 多分类指标	230
5.3.4 回归指标	232
5.3.5 在模型选择中使用评估指标	232
5.4 小结与展望	234
第6章 算法链与管道	236

- 6.1 用预处理进行参数选择 237
- 6.2 构建管道 238
- 6.3 在网格搜索中使用管道 239
- 6.4 通用的管道接口 242
 - 6.4.1 用make_pipeline方便地创建管道 243
 - 6.4.2 访问步骤属性 244
 - 6.4.3 访问网格搜索管道中的属性 244
- 6.5 网格搜索预处理步骤与模型参数 246
- 6.6 网格搜索选择使用哪个模型 248
- 6.7 小结与展望 249
- 第7章 处理文本数据 250
 - 7.1 用字符串表示的数据类型 250
 - 7.2 示例应用：电影评论的情感分析 252
 - 7.3 将文本数据表示为词袋 254
 - 7.3.1 将词袋应用于玩具数据集 255
 - 7.3.2 将词袋应用于电影评论 256
 - 7.4 停用词 259
 - 7.5 用tf-idf缩放数据 260
 - 7.6 研究模型系数 263
 - 7.7 多个单词的词袋（n元分词） 263
 - 7.8 高级分词、词干提取与词形还原 267
 - 7.9 主题建模与文档聚类 270
 - 7.10 小结与展望 277
- 第8章 全书总结 278
 - 8.1 处理机器学习问题 278
 - 8.2 从原型到生产 279
 - 8.3 测试生产系统 280
 - 8.4 构建你自己的估计器 280
 - 8.5 下一步怎么走 281
 - 8.5.1 理论 281
 - 8.5.2 其他机器学习框架和包 281
 - 8.5.3 排序、推荐系统与其他学习类型 282
 - 8.5.4 概率建模、推断与概率编程 282
 - 8.5.5 神经网络 283
 - 8.5.6 推广到更大的数据集 283
 - 8.5.7 磨练你的技术 284
 - 8.6 总结 284
 - 关于作者 285
 - 关于封面 285
 - • • • • • ([收起](#))

[Python机器学习基础教程_下载链接1_](#)

标签

机器学习

Python

MachineLearning

python

数据分析

人工智能

计算机科学

计算机

评论

强推，需要一定的python和机器学习的基础，里面没有理论只有实际运用，内容哪里过时了？基于python3.5的机器学习书，这本书已经走在很前面了。

不小心买了这本，才发现有些过时了，不过能用就是了，可惜没讲理论，只是单纯介绍算法的优点，看完才明白难怪别人说低层和中层的机器学习工程师只是在调参而已，争取看完去看一下算法理论好了。

只能说，要是能早点看到这本书就好了

scikit-learn库核心贡献者所写，介绍了很多经典机器学习算法。

过于简单粗暴了.....

囫圇吞枣看完了这个书 由于有的代码敲不了老报错 代码实现部分准备看其他的教程了
写的结果还是很棒棒的

20191028—01~1—6 20191031—02~6—20 (引言) +250— (处理文本数据)
20191106—03~254— 结束了七八两章，顺便学习与之相关的概念
总感觉内容干巴巴的，不是特别好看

结构非常好，条理很清晰。python机器学习的书里面目前最喜欢这本。

机器学习的入门书，通过这本书可以对机器学习建立感性认识。如果只是学习机器学习的理论推导，无法深刻理解各种数据处理方法带来的效果，通过Scikit-Learn和这本书可以快速建立起这种理解。

相当好的入门书，sklearn实在是太强大了

《Python for Data Analysis》的续作，学完数据分析之后开始了机器学习~非常愉快的学习之旅，让你快速上手Python机器学习（调包，调参，交叉验证，网格搜索），我花了三周时间读完，从此以后不再担忧不会用Python机器学习了。下一步，加强实践。

sklearn嘛

偏向实践方向的sklearn介绍的书籍。

作者写的书就是不一样。本书理论不多，偏重于实操，基于sklearn讲解常用的机器学习模型。技术书籍真的要读几遍，二刷完全不觉得我曾经看过，本书的理论虽然不深，但可谓是深入浅出。

很经典

翻译真的很渣 具体写算法的地方有点啰嗦 没有数学推导 最好再加上100页的数学文字那么多很不好理解 但是这本书的重点应该是在跑模型前后的工作

翻译的太差了, 稀里糊涂 说什么都不知道.中午都语法不通.还不如看 官方 文档

作为基础入门书，讲的挺好的，很多算法都讲到了，也讲了很多实用的技巧。

自学一个月成功变成调包侠

基本上告诉你用了库可以解决很多很多问题。理论一笔带过，当天书似的通读一遍。看完之后决定死磕理论。

[Python机器学习基础教程_下载链接1](#)

书评

在基于TensorFlow的深度学习框架大红大紫之前，其实在2010年前后流行过很多的经典机器学习框架。比如KNN，比如支撑向量机，比如随机森林。相对于深度学习的理论，这些经典的机器学习算法构建在更为精密的数学推导上。运筹学，最优化理论，数学分析，数理统计和随机过程构成了这...

扫码关注公众号

「图灵的猫」，点击“学习资料”菜单，可以获得海量python、机器学习、深度学习书籍、课程资源，以及书中对应习题答案和代码。后台回复SSR更有机场节点相送~
入门避坑指南

自学三年，基本无人带路，转专业的我自然是难上加难，踩过无数坑，走过很多弯路。
这里我...

扫码关注公众号

「图灵的猫」，点击“学习资料”菜单，可以获得海量python、机器学习、深度学习书籍、课程资源，以及书中对应习题答案和代码。后台回复SSR更有机场节点相送~

入门避坑指南

自学三年，基本无人带路，转专业的我自然是难上加难，踩过无数坑，走过很多弯路。
这里我...

[Python机器学习基础教程_下载链接1](#)