



Python极简讲义

一本书入门数据分析与机器学习

张玉宏◎著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

专家力荐

Python语言是快速实现数据分析、机器学习及相关领域理论与技术的利器。本书以通俗易懂的语言和丰富的实战案例介绍了Python基础内容、数据分析和机器学习必备知识，理论结合实战，娓娓道来，是一本难得的入门好书。

知名Python讲者、16本Python系列图书作者，董付国

Python是时下非常值得学习的编程语言，也是从事数据分析和机器学习相关工作的重要基石。张玉宏博士凭借多年的Python教学经验，为大家带来了图文并茂、简单易读的Python极简讲义，相信能够带领大家轻松学习Python语言，入门数据分析与机器学习，建议大家持卷品读。

微信公众号“程序员小灰”作者、《漫画算法》作者，魏梦舒

近几年，随着数据科学领域的发展，越来越多的非计算机专业人士也开始用计算机进行辅助数据处理，Python无疑是优选语言。这本《Python极简讲义》有着非常好的阅读体验，简单明了，案例丰富，手把手引导，非常适合跨界学习。相信它能带给你不同的入门体验。

《白话强化学习与PyTorch》作者、金山办公AI技术专家，高扬博士

本书以掌握“最少必要知识”（MAKE）为写作理念，融合了Python编程、数据分析和机器学习等热门领域的入门知识，文笔流畅，语言幽默，对初学者十分友好。相信任何有志于从事数据分析和机器学习相关工作的读者，都能从此书中收获良多。

CSDN千万级流量博主、七月在线CEO，July

在数据科技（DT）时代，数据分析与机器学习基本上是每个DT职场人士的必备技能。本书以DT时代非常流行的编程语言Python为抓手，轻松幽默地讲解了数据分析和机器学习的“最少必要知识”。本书中时有体现哲学思想的内容点缀，耐人寻味，是一本很好的入门图书，不仅适合初学者阅读，对于具有一定经验的工程师也颇具参考价值。

SIGAI创始人、《机器学习：原理、算法与应用》作者、百度前高级软件工程师，雷明



读者服务

微信扫码回复：38704

- 获取博文视点学院20元付费内容抵扣券
- 获取本书配套代码资源
- 获取更多技术专家分享的视频与学习资源
- 加入本书读者交流群，与更多读者互动



责任编辑：孙奇俏

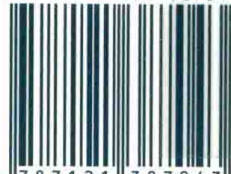
欢迎投稿

封面设计：李玲

邮箱：sunqq@phei.com.cn

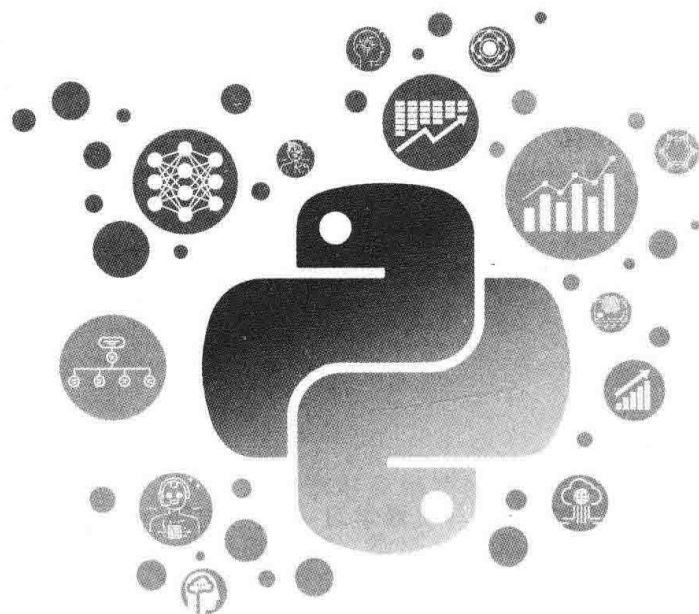
上架建议：计算机>算法

ISBN 978-7-121-38704-3



9 787121 387043 >

定价：128.00元



Python极简讲义

一本书入门数据分析与机器学习

张玉宏◎著

电子工业出版社

Publishing House of Electronics Industry

北京•BEIJING

内 容 简 介

本书以图文并茂的方式介绍了 Python 的基础内容，并深入浅出地介绍了数据分析和机器学习领域的相关入门知识。

第 1 章至第 5 章以极简方式讲解了 Python 的常用语法和使用技巧，包括数据类型与程序控制结构、自建 Python 模块与第三方模块、Python 函数和面向对象程序设计等。第 6 章至第 8 章介绍了数据分析必备技能，如 NumPy、Pandas 和 Matplotlib。第 9 章和第 10 章主要介绍了机器学习的基本概念和机器学习框架 sklearn 的基本用法。

对人工智能相关领域、数据科学相关领域的读者而言，本书是一本极简入门手册。对于从事人工智能产品研发的工程技术人员，本书亦有一定的参考价值。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目 (CIP) 数据

Python 极简讲义：一本书入门数据分析与机器学习 / 张玉宏著. —北京：电子工业出版社，2020.5
ISBN 978-7-121-38704-3

I. ①P… II. ①张… III. ①软件工具—程序设计②机器学习 IV. ①TP311.561②TP181

中国版本图书馆 CIP 数据核字 (2020) 第 039554 号

责任编辑：孙奇俏

印 刷：三河市双峰印刷装订有限公司

装 订：三河市双峰印刷装订有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱

邮编：100036

开 本：787×980 1/16

印张：36.75

字数：811.4 千字

版 次：2020 年 5 月第 1 版

印 次：2020 年 5 月第 1 次印刷

定 价：128.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819，faq@phei.com.cn。

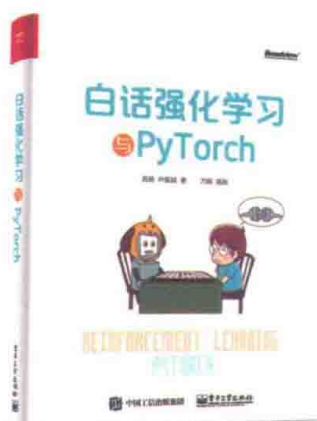
作者简介



张玉宏

大数据分析师（高级），2012年于电子科技大学获得博士学位，2009—2011年美国西北大学访问学者，2019—2020年美国IUPUI高级访问学者，YOCSEF郑州2019—2020年度副主席。现执教于河南工业大学，主要研究方向为大数据、机器学习。发表学术论文30余篇，先后撰写《深度学习之美：AI时代的数据处理与最佳实践》《品味大数据》等科技图书7本，参与编写英文学术专著2部。

好书分享



推荐语

Python 语言是快速实现数据分析、机器学习及相关领域理论与技术的利器。本书以通俗易懂的语言和丰富的实战案例介绍了 Python 基础内容、数据分析和机器学习必备知识，理论结合实战，娓娓道来，是一本难得的入门好书。

知名 Python 讲者、16 本 Python 系列图书作者，董付国

Python 是时下非常值得学习的编程语言，也是从事数据分析和机器学习相关工作的重要基石。张玉宏博士凭借多年的 Python 教学经验，为大家带来了图文并茂、简单易读的 Python 极简讲义，相信能够带领大家轻松学习 Python 语言，入门数据分析与机器学习，建议大家持卷品读。

微信公众号“程序员小灰”作者、《漫画算法》作者，魏梦舒

近几年，随着数据科学领域的发展，越来越多的非计算机专业人士也开始用计算机进行辅助数据处理，Python 无疑是优选语言。这本《Python 极简讲义》有着非常良好的阅读体验，简单明了，案例丰富，手把手引导，非常适合跨界学习。相信它能带给你不同的入门体验。

《白话强化学习与 PyTorch》作者、金山办公 AI 技术专家，高扬博士

本书以掌握“最少必要知识”（MAKE）为写作理念，融合了 Python 编程、数据分析和机器学习等热门领域的入门知识，文笔流畅，语言幽默，对初学者十分友好。相信任何有志于从事数据分析和机器学习相关工作的读者，都能从此书中收获良多。

CSDN 千万级流量博主、七月在线 CEO, July

在数据科技（DT）时代，数据分析与机器学习基本上是每个 DT 职场人士的必备技能。本书以 DT 时代非常流行的编程语言 Python 为抓手，轻松幽默地讲解了数据分析和机器学习的“最少必要知识”。本书中时有体现哲学思想的内容点缀，耐人寻味，是一本很好的入门图书，不仅适合初学者阅读，对于具有一定经验的工程师也颇具参考价值。

SIGAI 创始人、《机器学习：原理、算法与应用》作者、百度前高级软件工程师，雷明

前言

为什么写此书

我们正处于一个数据科技（Data Technology, DT）时代。在这个时代，我们的一举一动都能和数据空间留下电子印记，于是海量的社交、电商、科研大数据扑面而来。然而，太多的数据给人们带来的，可能并不是更多的洞察，反而是迷失。

因为仅就数据本身而言，它们是“一无所知”的。数据的价值，在于形成信息，变成知识，乃至升华为智慧。也就是说，这些数据如果不能进一步被“深加工”，即使数据量再“大”，也意义甚小。

于是，就派生出这么一个问题：这些数据，由谁来深加工？其实，早在2012年，《哈佛商业评论》就刊登了一篇文章并给出了答案，进行数据深加工的人就是“数据科学家”。文章还断言，数据科学家是21世纪最“性感”的职业。

但如何成为一名数据科学家呢？尤其是一名“性感”的数据科学家？作为一个有点文艺范儿的理工男，我喜欢读书，从书中获得启迪。

我曾读过埃里克·莱斯（Eric Ries）写的一本有关创业的书，书名是《精益创业》（*The Lean Startup*）。在阅读这本书时，有一个概念深入我心，那就是“最小可行产品”（Minimum Viable Product, MVP）。围绕这个概念，创业者在创业初期不可贪多求全，而应该先做出一个最小的可用产品，拿到市场上去检验，然后根据反馈反复迭代，打磨升级，最终做出比较完善、比较成功的产品。

后来，我看到有人提出了类似的概念——最少必要知识（Minimal Actionable Knowledge and Experience, MAKE）。这个概念点醒了我，我觉得自己大概找到了“如何成为一名数据科学家”这个问题的答案：You can MAKE it!

先来说说什么是 MAKE。它指的是入门某个新领域切实可行的最小知识集合。MAKE 说起来好像比较高级，但实际上，它背后有一个支撑它的朴素原则——Pareto 原则（亦称 80-20 原则），即 80% 的工作问题可以通过掌握 20% 的知识来解决。

同样，想成为一名“性感”的数据科学家，一条路自然是按部就班地学习所有技能——十年磨一剑，但这样做的风险在于，当你“携剑下山”时，别人可能已经用上了飞机、大炮。这样的对垒，你胜算几何？

其实还有另一条备选之路，那就是走一走 MAKE 之道。在学习某项技能（如 Python、数据分析、机器学习）时，我们要想办法在最短的时间内，摸索清楚这项技能的“最少必要知识”。一方面，它已然可以帮我们解决工作中的大部分问题；另一方面，入门之后，技能的提升通道可以在实践中寻得，缺啥补啥。有明确的任务导向，学习就会有如神助，这也是当前时代的快节奏学习法。

说到这里，本书的写作初衷就呼之欲出了。是的，这本书的定位就是，为初学者提供关于数据科学的“最少必要知识”，从而让你获得那份最“性感”的工作。这些知识包含了成为数据科学家所需要掌握的基础内容——Python 编程、数据分析、机器学习。

本书内容

本书主要介绍 Python 的基础知识、数据分析的必备技能，以及机器学习相关内容。全书共分 10 章，每章的内容简介如下。

第 1 章 初识 Python 与 Jupyter

Python 是最具人气的编程语言之一，Jupyter 是人气与口碑俱佳的 Python 开发平台。本章将介绍 Python 和 Jupyter 的基本内容，包括 Python 的安装与运行，以及文学化编程利器 Jupyter 的使用方法。

第 2 章 数据类型与程序控制结构

本章将介绍 Python 的基础语法及常见的数据类型，包括数值型、布尔类型、字符串型、列表、元组、字典、集合等。此外，本章还将介绍三种程序控制结构（顺序结构、选择结构和循环结构）和高效的推导式。

第 3 章 自建 Python 模块与第三方模块

本章将介绍 Python 的自定义模块及常用的第三方模块，包括 `collection`、`datetime`、`json`、`random` 等模块。

第 4 章 Python 函数

本章将讨论 Python 的函数定义、函数参数（关键字参数、可变参数、默认参数等）的“花式”传递、函数的递归调用，以及函数式编程。

第 5 章 Python 高级特性

本章将介绍 Python 中的一些高阶应用，这些高阶应用能让我们更高效地写出更专业的 Python 代码。本章内容涉及面向对象程序设计思想、生成器与迭代器、文件操作、异常处理及错误调试等。

第 6 章 NumPy 向量计算

本章将讨论 NumPy 数组的构建、方法和属性，介绍 NumPy 的广播机制、布尔索引、数组的堆叠，以及爱因斯坦求和约定等。

第 7 章 Pandas 数据分析

Pandas 是数据分析的利器，本章将主要介绍 Pandas 的两种常用数据处理结构：Series 和 DataFrame。同时介绍基于 Pandas 的文件读取与分析，涉及数据的清洗、条件过滤、聚合与分组等。

第 8 章 Matplotlib 与 Seaborn 可视化分析

Matplotlib 和 Seaborn 是非常好用的数据可视化包，本章将主要介绍 Matplotlib 和 Seaborn 的基本用法，并基于此绘制可视化图形，包括散点图、条形图、直方图、饼图等。同时，本章将以谷歌流感趋势数据为例，结合 Pandas 进行可视化分析。

第 9 章 机器学习初步

本章将主要介绍有关机器学习的初步知识，包括机器学习的定义，机器学习的几个主要流派，并讨论机器学习模型的性能评估指标，包括混淆矩阵、查准率、查全率、P-R 曲线、ROC 曲线等。

第 10 章 sklearn 与经典机器学习算法

本章将主要讲解知名机器学习框架 sklearn 的用法，并介绍几种经典机器学习算法的原理和实战，这些算法包括线性回归、 k -近邻算法、Logistics 回归、神经网络学习算法、 k 均值聚类算法等。

阅读准备

要想运行本书中的示例代码，需要提前安装如下系统及软件。

- 操作系统：Windows、macOS、Linux 均可。
- Python 环境：建议使用 Anaconda 安装，确保版本为 Python 3.x 即可。
- NumPy：建议使用 Anaconda 安装 NumPy 1.18 及以上版本。
- Pandas：建议使用 Anaconda 安装 Pandas 1.0.1 及以上版本。
- sklearn：建议使用 Anaconda 安装 sklearn 0.22.1 及以上版本。

联系作者

数据科学是一个前沿且广袤的研究领域，很少有人能对其每个研究方向都有深刻的认知。我自认才疏学浅，同时限于时间与篇幅，书中难免出现理解偏差和错缪之处。若读者朋友们在阅读本书的过程中发现问题，希望能及时与我联系，我将在第一时间修正并对此不胜感激。

邮件地址：zhangyuhong001@gmail.com。

致谢

这本《Python 极简讲义：一本书入门数据分析与机器学习》，从构思大纲、查阅资料、撰写内容、绘制图片，到出版成书，历时一年有余。图书得以面市，自然得益于多方面的帮助和支持。在信息获取上，我学习并吸纳了很多精华知识，书中也尽可能地给出了文献出处，如有疏漏，望来信告知。在这里，我对这些高价值资料的提供者、生产者，表示深深的敬意和感谢。

很多人在这本书的出版过程中扮演了重要角色——电子工业出版社博文视点的孙奇俏老师在选题策划和文字编辑上，国网河北电科院计量中心的赵佩、河南工业大学的苏灏等在文稿校对上，都付出了辛勤的劳动，在此对他们一并表示感谢。

张玉宏

2020 年 3 月于美国

读者服务

微信扫码回复：38704



- 获取博文视点学院 20 元付费内容抵扣券
- 获取本书配套代码资源
- 获取更多技术专家分享的视频与学习资源
- 加入本书读者交流群，与更多读者互动

目录

第 1 章 初识 Python 与 Jupyter	1
1.1 Python 概要	2
1.1.1 为什么要学习 Python	2
1.1.2 Python 中常用的库	2
1.2 Python 的版本之争	4
1.3 安装 Anaconda	5
1.3.1 Linux 环境下的 Anaconda 安装	5
1.3.2 conda 命令的使用	6
1.3.3 Windows 环境下的 Anaconda 安装	7
1.4 运行 Python	11
1.4.1 验证 Python	11
1.4.2 Python 版本的 Hello World	12
1.4.3 Python 的脚本文件	13
1.4.4 代码缩进	15
1.4.5 代码注释	17
1.5 Python 中的内置函数	17
1.6 文学化编程——Jupyter	20
1.6.1 Jupyter 的由来	20
1.6.2 Jupyter 的安装	21

1.6.3	Jupyter 的使用	23
1.6.4	Markdown 编辑器	26
1.7	Jupyter 中的魔法函数	31
1.7.1	%lsmagic 函数	31
1.7.2	%matplotlib inline 函数	32
1.7.3	%timeit 函数	32
1.7.4	%%writefile 函数	33
1.7.5	其他常用的魔法函数	34
1.7.6	在 Jupyter 中执行 shell 命令	35
1.8	本章小结	35
1.9	思考与提高	36
第 2 章	数据类型与程序控制结构	40
2.1	为什么需要不同的数据类型	41
2.2	Python 中的基本数据类型	42
2.2.1	数值型 (Number)	42
2.2.2	布尔类型 (Boolean)	45
2.2.3	字符串型 (String)	45
2.2.4	列表 (List)	49
2.2.5	元组 (Tuple)	59
2.2.6	字典 (Dictionary)	62
2.2.7	集合 (Set)	65
2.3	程序控制结构	67
2.3.1	回顾那段难忘的历史	67
2.3.2	顺序结构	69
2.3.3	选择结构	70
2.3.4	循环结构	74
2.4	高效的推导式	80
2.4.1	列表推导式	80
2.4.2	字典推导式	83
2.4.3	集合推导式	83

2.5	本章小结	84
2.6	思考与提高	84
第 3 章	自建 Python 模块与第三方模块.....	90
3.1	导入 Python 标准库	91
3.2	编写自己的模块	93
3.3	模块的搜索路径	97
3.4	创建模块包	100
3.5	常用的内建模块	103
3.5.1	collection 模块.....	103
3.5.2	datetime 模块.....	110
3.5.3	json 模块.....	115
3.5.4	random 模块	118
3.6	本章小结	121
3.7	思考与提高	122
第 4 章	Python 函数	124
4.1	Python 中的函数	125
4.1.1	函数的定义	125
4.1.2	函数返回多个值	127
4.1.3	函数文档的构建	128
4.2	函数参数的“花式”传递.....	132
4.2.1	关键字参数	132
4.2.2	可变参数	133
4.2.3	默认参数	136
4.2.4	参数序列的打包与解包.....	138
4.2.5	传值还是传引用	142
4.3	函数的递归	146
4.3.1	感性认识递归	146
4.3.2	思维与递归思维	148

4.3.3 递归调用的函数	149
4.4 函数式编程的高阶函数.....	151
4.4.1 lambda 表达式.....	152
4.4.2 filter()函数.....	153
4.4.3 map()函数.....	155
4.4.4 reduce()函数.....	157
4.4.5 sorted()函数.....	158
4.5 本章小结	159
4.6 思考与提高	160
 第 5 章 Python 高级特性	165
5.1 面向对象程序设计	166
5.1.1 面向过程与面向对象之辩.....	166
5.1.2 类的定义与使用	169
5.1.3 类的继承	173
5.2 生成器与迭代器	176
5.2.1 生成器	176
5.2.2 迭代器	183
5.3 文件操作	187
5.3.1 打开文件	187
5.3.2 读取一行与读取全部行.....	191
5.3.3 写入文件	193
5.4 异常处理	193
5.4.1 感性认识程序中的异常.....	194
5.4.2 异常处理的三步走	195
5.5 错误调试	197
5.5.1 利用 print()输出观察变量	197
5.5.2 assert 断言	198
5.6 本章小结	201
5.7 思考与提高	202