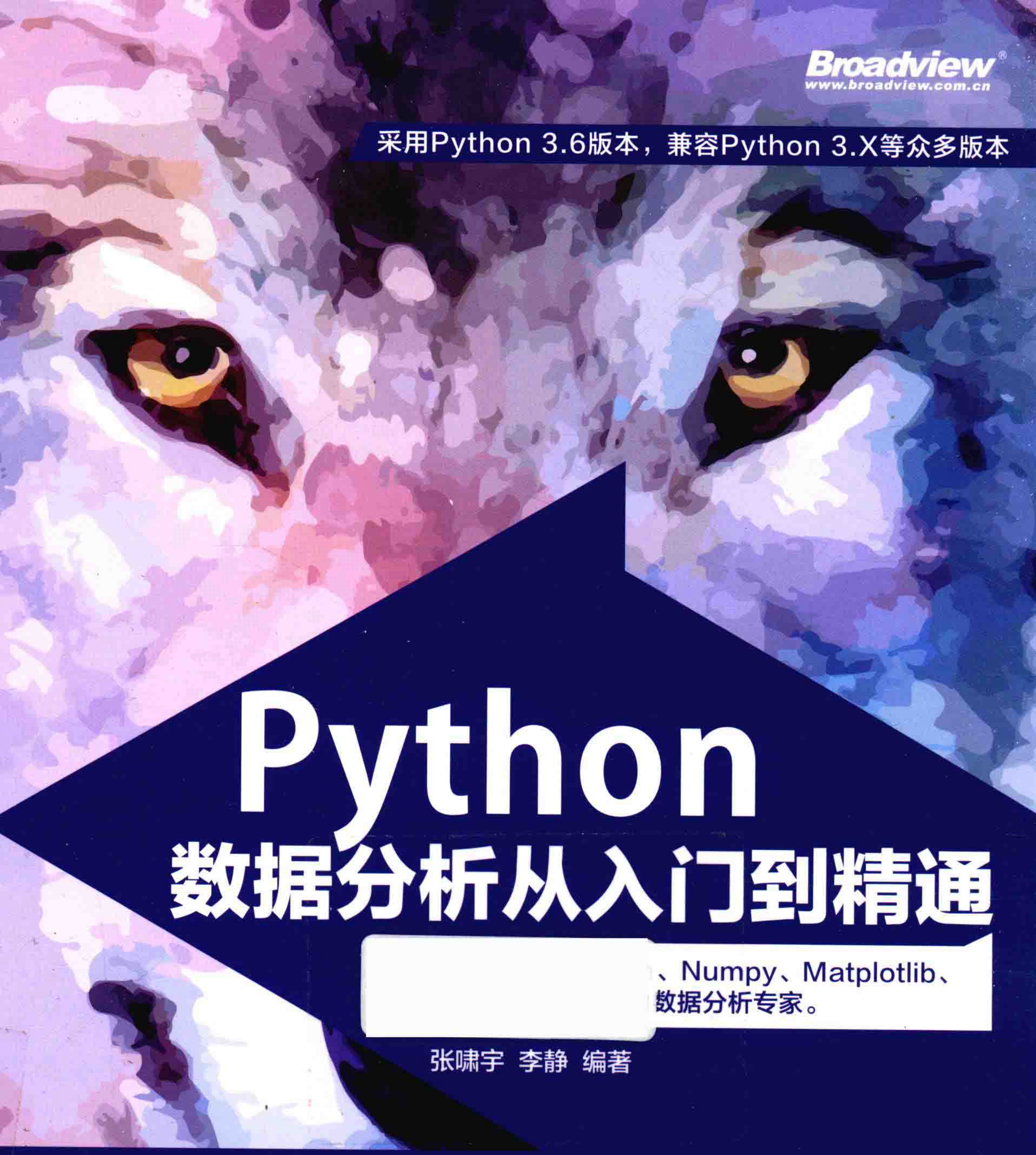
Broadview®
www.broadview.com.cn

采用Python 3.6版本，兼容Python 3.X等众多版本



Python 数据分析从入门到精通

、Numpy、Matplotlib、
数据分析专家。

张啸宇 李静 编著

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

Python

数据分析从入门到精通

张啸宇 李静 编著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

对于希望使用 Python 来完成数据分析工作的人来说，学习 IPython、Numpy、pandas、Matplotlib 这个组合是目前看来不错的方向。本书就是这样一本循序渐进的书。

本书共 3 篇 14 章。第 1 篇是 Python 数据分析语法入门，将数据分析用到的一些语言的语法基础讲解清楚，为接下来的数据分析做铺垫。第 2 篇是 Python 数据分析工具入门，介绍了 Python 数据分析“四剑客”——IPython、Numpy、pandas、Matplotlib。第 3 篇是 Python 数据分析案例实战，包括两个案例，分别是数据挖掘和玩转大数据，为读者能真正使用 Python 进行数据分析奠定基础。

本书内容精练、重点突出、实例丰富，是广大数据分析工作者必备的参考书，同时也非常适合大、中专院校师生学习阅读，还可作为高等院校统计分析及相关专业的教材。

图书在版编目（CIP）数据

Python 数据分析从入门到精通 / 张啸宇，李静编著. —北京：电子工业出版社，2018.3

ISBN 978-7-121-33613-3

Ⅰ. ①P… Ⅱ. ①张… ②李… Ⅲ. ①软件工具—程序设计 Ⅳ. ①TP311.561

中国版本图书馆 CIP 数据核字（2018）第 020314 号

策划编辑：张月萍
责任编辑：牛　勇
特约编辑：赵树刚
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：20.75　　字数：415 千字
版　　次：2018 年 3 月第 1 版
印　　次：2018 年 6 月第 2 次印刷
印　　数：1000 册　　定价：69.00 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到 dbqq@phei.com.cn。

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

前 言

由于 Python 具有简单、易学、免费开源、可移植性、可扩展性等特点，所以它的受欢迎程度扶摇直上。再加上 Python 拥有非常丰富的库，这也使得它在数据分析领域有着越来越广泛的应用。如果你已经决定学习 Python 数据分析，但是之前没有编程经验，那么本书将会是你的正确选择。

本书的第 1 篇主要介绍学习数据分析必备的一些 Python 语法基础，包括 Python 的安装、数据类型、数据结构、模块、类、异常处理、使用 pip 安装 Python 需要的一些工具等；第 2 篇主要介绍 Python 在数据处理和科学计算方面的工具和方法，包括 IPython 交互式壳的使用、Jupyter Notebook 的使用和 Numpy 的使用，还介绍了 Python 的核心数据分析处理库 pandas，以及 Python 著名的 2D 绘图库 Matplotlib；第 3 篇通过数据挖掘和玩转大数据两个案例总结和应用前面所学的知识。

这三篇的层进正好是 Python 数据分析入门者的阶梯，读者通过学习这三部分内容，即可迈入数据分析的门槛。

本书的特点

Python 是当前非常流行的面向对象编程语言，本书将其在数据分析处理方面的特色发挥到极致。本书的主要特点如下：

- Python 被大量应用在数据挖掘和机器学习领域，其中使用极其广泛的是 IPython、Numpy、pandas、Matplotlib 等库。本书详细地介绍了这些库的组成与使用，为科学计算相关人员提供了有用的参考资料。
- 本书采取循序渐进的写作风格，对于工具的安装、使用步骤、方法技巧逐步展

开，加以图解和应用场景，即使完全不懂 Python 和数据分析的人员，也可以流畅地读完本书。

- 无论哪种语言，编程的方法、模式、数据结构、算法都是相通的。本书将科学计算、数据结构与各种工具和方法完美结合，让非 Python 读者也能融会贯通，让学习统计的人能找到更适合的统计方法和数据分析处理方法。
- 本书最后的两个实战案例适合数据分析入门者，案例的步骤详细、分析到位，能为读者入手真实项目打下良好的基础。

本书的内容安排

本书共 3 篇 14 章，主要章节规划如下：

第 1 章介绍了 Python 的发展历程、特性，帮助读者搭建最基础的数据分析环境，下载开发语言，选择开发工具，然后在此基础上开发自己的第一个 Python 程序。读者在学完本章内容后应该对 Python 有一个基础的认识，知道为什么选择它来进行数据分析。

第 2 章介绍了 Python 的基础语法，包括它的代码组织形式、如何缩进、如何注释等，以及输入/输出该如何处理，在中文环境下如何更好地使用 Python 是本章的重点，最后还通过一个实例复习了 Python 的这些语法。读者在学完本章内容后可以轻松地编写一些简单的 Python 程序。

第 3 章介绍了 Python 的数据类型与流程控制语句。如果读者已有编程基础，那么阅读本章内容不会有任何压力。如果没有编程基础，那么学习一门语言的流程控制最关键的就是这些知识。读者在阅读完本章后就能轻松阅读更大的 Python 程序。

第 4 章介绍了可复用的函数与模块。这些内容较为复杂，但却是进行数据分析的关键。每个数据处理过程我们都会用到函数或模块，而我们后期用到的数据分析库也可以说是一个大函数。所以学习完本章内容，读者应该能够看明白一个完整的 Python 库。

第 5 章介绍了数据结构与算法，这是数据分析的基础，也是人工智能的基础。利用算法我们可以找到解决方案，也可以找到最优路径，还可以更高效地完成数据分析任务。读者如果没有看懂本章内容，一定要反复阅读，直到学会为止。

第 6 章介绍了面向对象的 Python。面向对象已经成为每门语言都具备的特性，类、对象、继承这些概念都是面向对象的基础。如果读者没有编程经验，则阅读本章可能会

有一定的难度，但是了解了对象的概念，就能学会如何编写更高效的代码、如何让代码和代码之间联动起来。

第 7 章介绍了异常处理与程序调试。机器毕竟不是人，如果出现错误，则可能会导致死机，或者数据出错。为了防止这些错误的发生，或者防止程序的使用方能得到反馈，我们必须学会 Python 的异常处理功能。并且当程序发生错误时，我们还要通过程序调试找到错误所在。

第 8 章介绍了 pip 软件包管理。既然在做数据分析时我们要用到很多数据分析库，那么如何下载、安装或管理这些库就成了数据分析的第一步。pip 就是这样一个工具，它能下载、安装、更新、显示、搜索我们需要的数据分析库。

第 9 章介绍了 IPython 科学计算库，它是使用 Python 进行数据分析、处理、呈现的重要选择之一。本章主要介绍了 Python 科学计算库的安装方法、IPython 壳的一些特性和基本功能、Jupyter Notebook 的安装和使用方法。IPython 壳的使用是本章的重点，也是数据分析处理的基础工具，希望读者能够消化本章内容，为真正做好数据项目打下基础。

第 10 章介绍了 Numpy 科学计算库，主要介绍了它的数组对象及数组对象的一些基本属性和生成数组的基本方法，还包括数组的索引和分片等基本操作，这部分内容是 Numpy 数据处理的核心。本章介绍的代数运算函数、线性代数、统计函数等内容会让读者觉得有些困惑，但这已经进入了数据分析的关键时刻，所以仍建议读者对本章的内容融会贯通。

第 11 章介绍了 pandas 数据分析处理库，主要包括它的序列、数据框的基本操作，还包括 pandas 里处理默认值、读取常见格式的文本数据，以及数据的组合和分组操作。最后介绍了 pandas 的时间序列和一个处理实际数据集的案例，读者掌握了这些内容，就可以更好地处理数据。

第 12 章介绍了 Matplotlib 的 Pyplot 和 Artist 模块，以及 pandas 的绘图功能。对于读者来说，Pyplot 模块是需要掌握的，Artist 模块是需要了解的，pandas 的绘图功能在实际数据分析中要能熟练应用。

第 13 章是数据挖掘的案例。首先介绍了著名的贝叶斯理论，然后实现了贝叶斯分类器，最后实现了协同过滤算法，这些都是数据挖掘、分析领域的基础算法。建议读者尝试自己编写代码，熟练掌握贝叶斯分类器和协同过滤算法的使用。

第 14 章是玩转大数据的案例。鉴于本书主要针对数据分析入门者，所以本章也逐

步实现了数据的分析过程，从了解数据到分析数据，最后到代码实现，相信读者学完本章内容后，就能真正动手分析大数据了。

本书由浅入深、从理论到实践，尤其适合初学者逐步学习和完善自己的知识结构。

请访问 www.broadview.com.cn/33613 下载本书配套代码。

适合阅读本书的读者

- 希望从事数据分析相关工作的人员。
- 数据分析工作人员。
- 大数据从业人员。
- Python 爱好者。
- 人工智能从业人员。
- 统计行业的人员。
- 大、中专院校统计相关专业的学生。

编　者

目 录

第 1 篇 Python 数据分析语法入门