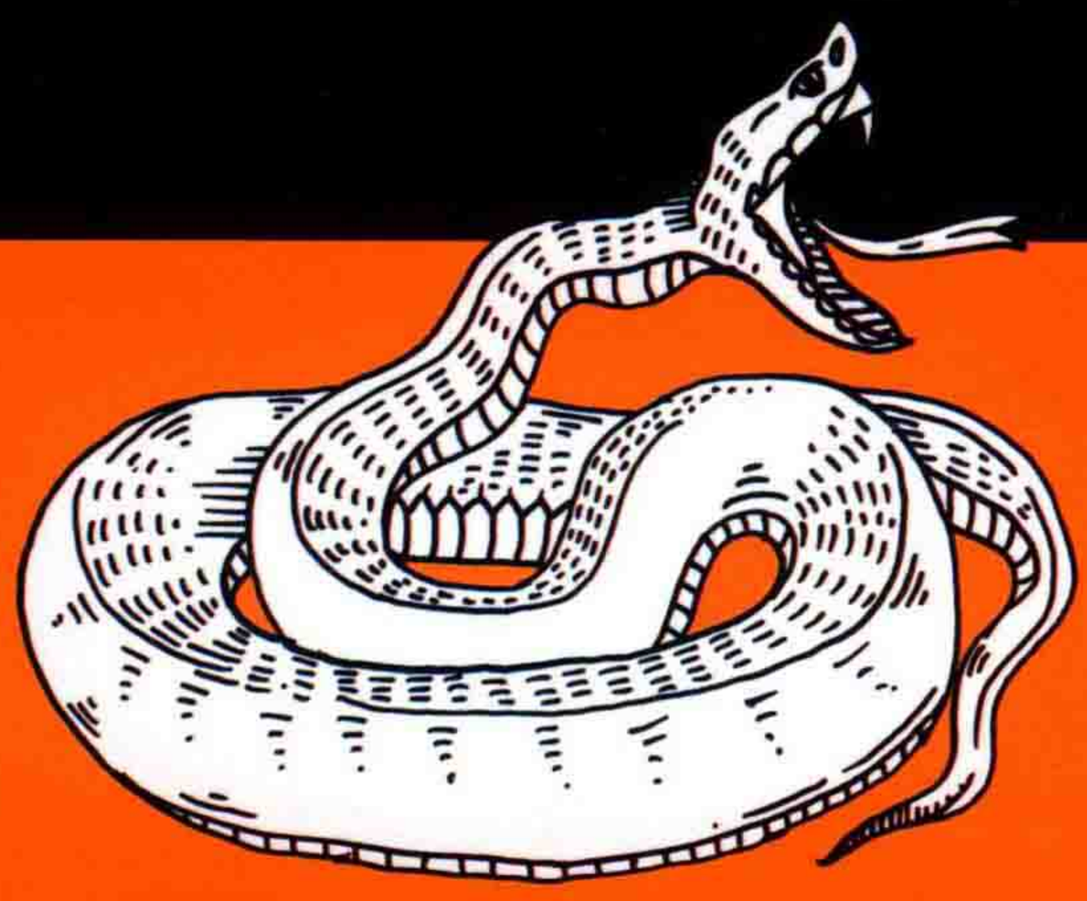
45个大型真实项目案例

95节同步微视频讲解

18小时Python 3 语法教学视频

Python 3 数据分析与机器学习实战

龙马高新教育◎编著

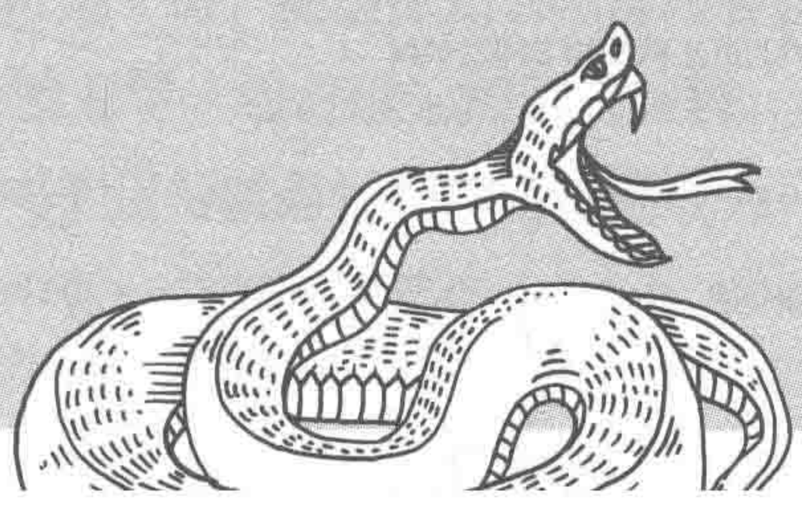


北京大学出版社
PEKING UNIVERSITY PRESS

Python 3

数据分析与机器学习实战

龙马高新教育◎编著



北京大学出版社
PEKING UNIVERSITY PRESS

内 容 提 要

机器之所以能学习，是因为大量的数据分析。

本书首先讲述数据分析的过程，然后详细介绍常用的机器学习理论、算法与案例（大型案例 29 个），最终以解决实际问题驱动成书。本书主要介绍的机器学习算法及数据分析方法包括数据预处理、分类问题、预测问题、网络爬虫、数据降维、数据压缩、关联分析、集成学习和深度学习等。

全书共 17 章分为三大部分：第 0~3 章介绍 Python 的基础知识、安装和基本语法；第 4~7 章介绍 Python 的编程、机器学习基础和 Python 中常用的第三方库函数及数据预处理的基本方法；第 8~16 章介绍常用的机器学习分析算法及深度学习。每章都结合多个经典案例图文并茂地介绍机器学习的原理和实现方法。

本书通俗易懂，且赠送全程同步教学录像和 Python 3 编程基础录像，属学习 Python 及机器学习理论和数据分析的入门与提高课程，对于不熟悉 Python 又想学习机器学习相关算法的初学者来说非常适合。

图书在版编目（CIP）数据

Python 3 数据分析与机器学习实战 / 龙马高新教育编著 . — 北京：北京大学出版社，2018.8
ISBN 978-7-301-29566-3

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

中国版本图书馆 CIP 数据核字 (2018) 第 116684 号

书　　名	Python 3 数据分析与机器学习实战 PYTHON 3 SHUJU FENXI YU JIQI XUEXI SHIZHAN
著作责任者	龙马高新教育 编著
责任编辑	尹 毅
标准书号	ISBN 978-7-301-29566-3
出版发行	北京大学出版社
地　　址	北京市海淀区成府路 205 号　100871
网　　址	http://www.pup.cn　　新浪微博：@ 北京大学出版社
电子信箱	pup7@ pup.cn
电　　话	邮购部 62752015　发行部 62750672　编辑部 62570390
印 刷 者	北京大学印刷厂
经 销 者	新华书店
	787 毫米 ×1092 毫米　16 开本　19.75 印张　436 千字
	2018 年 8 月第 1 版　2018 年 8 月第 1 次印刷
印　　数	1—4000 册
定　　价	69.00 元

前　言

本书专为 Python 初学者和爱好者打造，旨在使读者学会、掌握 Python 相关知识和技能，并能进行项目开发。当您认真、系统地学习本书之后，就可以骄傲地说“我是一位真正的 Python 程序员了！”，哪怕目前您还是初学者。

机器学习 (Machine Learning, ML) 是一门多领域交叉的学科，是人工智能的核心，其应用遍及人工智能的各个领域，专门研究计算机是如何模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构，使之不断改善自身的性能。在机器学习过程中，需要使用大量数据，而数据分析是指用适当的方法对收集来的大量数据进行分析，提取有用信息和形成结论而对数据加以详细研究和概括总结的过程。

◈ 为什么要写这样一本书

古人云：“不闻不若闻之，闻之不若见之，见之不若知之，知之不若行之。”实践对于学习知识的重要性由此可见一斑，而理论知识与实践经验的脱节，恰恰是很多 Python 图书的写照。从项目开发经验入手，结合理论知识的讲解，便成为本书的立足点，也转化成了作者对本书的要求。本书的目标就是让初学者快速成为一个 Python 程序员，并拥有项目开发技能，在未来的职场中有一个较高的起点。

◈ 读者对象

- 没有任何 Python 基础的初学者
- 有一定的 Python 基础，想精通 Python 数据分析与机器学习的人员
- 有一定的 Python 基础，没有项目经验的人员
- 正在进行毕业设计的学生
- 大专院校及培训学校的教师和学生

◈ Python 最佳学习途径

◈ 本书特色

■ **零基础也能入门**

无论您是否从事计算机相关行业，是否接触过 Python，是否使用 Python 开发过项目，都能从本书开启学习之旅。

■ **专业的项目指导**

本书结合实际工作中的范例逐一讲解 Python 的各种知识和技术，使您在实战中掌握知识，轻松拥有项目经验。

◈ 配套资源

■ **源代码及视频下载**

扫描下方二维码或在浏览器中输入下载链接“http://v.51pcbook.cn/download/29566.html”，即可下载本书配套素材与视频资源。

① **提示：** 如果下载链接失效，请加入“办公之家”QQ 群（218192911），联系管理员获取最新下载链接。

① **注意：** 如加入 QQ 群时，系统提示此群已满，请根据验证信息加入新群。

■ **扫描二维码观看同步视频**

使用微信、QQ 或浏览器中的“扫一扫”功能，扫描每节中对应的二维码，即可观看相应的同步教学视频。

■ **手机版打包视频**

读者可以扫描下方二维码下载龙马高新教育手机 APP，将其直接安装到手机中，随时随地问同学、问专家，尽享海量资源。同时，我们也会不定期地向读者推送学习中常见的难点、使用技巧、行业应用等精彩内容，让学习更加简单高效。

◈ 作者团队

本书由龙马高新教育策划，由史卫亚任主编，于俊伟任副主编。其中第 0~9 章、11、14、16 章由河南工业大学史卫亚老师编著，第 10、12、13、15 章由河南工业大学于俊伟

老师编著。在编写过程中，编者竭尽所能地为读者呈现最好、最全的实用功能，但仍难免有疏漏和不妥之处，敬请广大读者指正。读者若在学习过程中产生疑问，或有任何建议，可以通过以下方式联系我们。

投稿信箱：pup7@pup.cn

读者信箱：2751801073@qq.com

读者交流 QQ 群：218192911（办公之家）

注意：如加入 QQ 群时，系统提示此群已满，请根据验证信息加入新群。

目　录

第0章 本书的技术体系

Python 语言是一种语法简洁而清晰，面向对象的高级程序设计语言，并且 Python 语言使用广泛，代码范例也很多，便于读者快速学习和掌握，十分容易上手。机器学习已应用于人们生活的方方面面，远超出大多数人的想象，如人脸识别、预测天气、垃圾邮件过滤和电商网站购物产品推荐等。随着各种数据以指数级增长，我们不仅需要使用更好的工具对这些数据进行分析，还要通过这些数据分析掌握其内涵，进而进行学习以提高人类应对未知世界的能力。

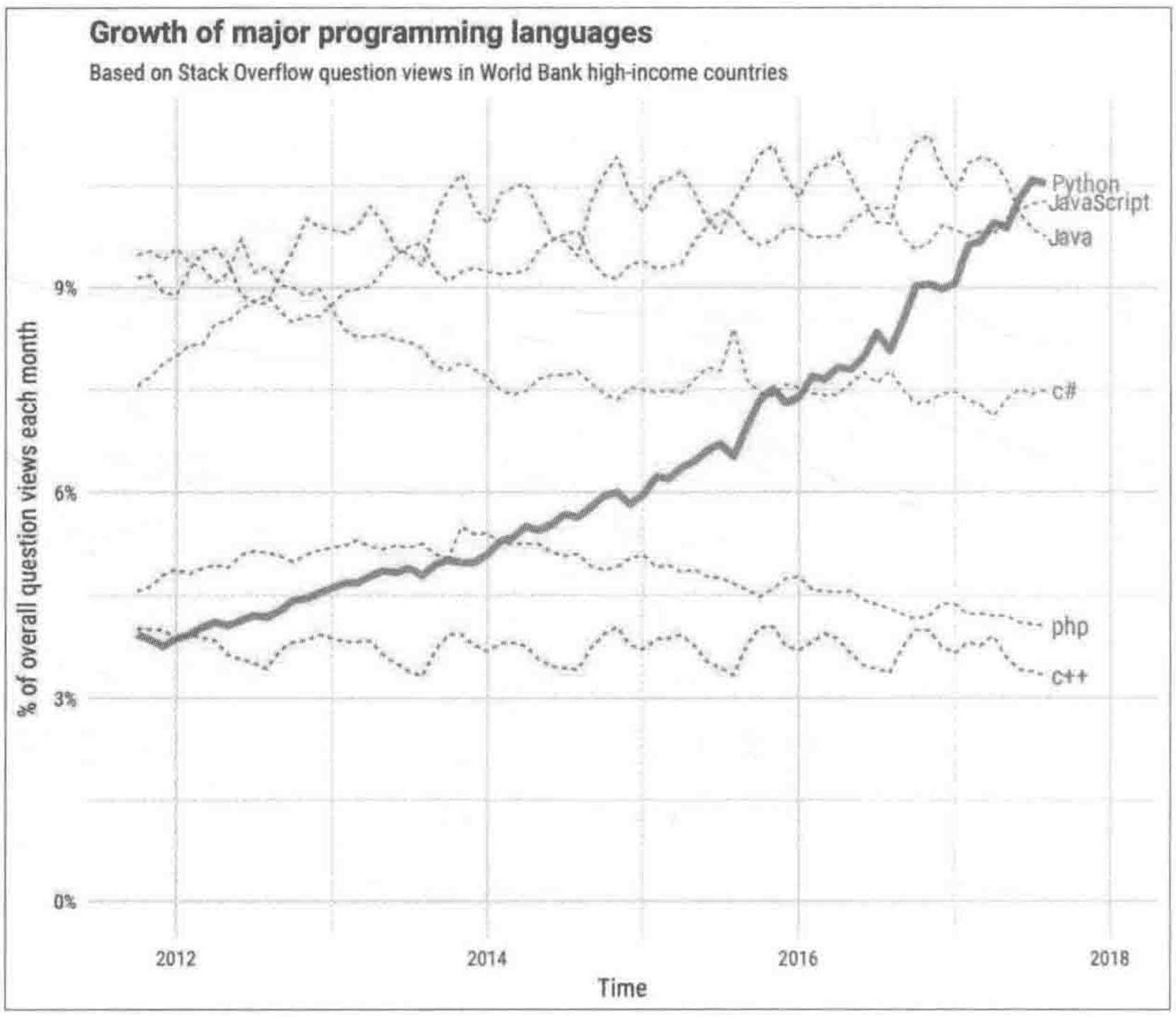
本章将介绍以下内容：

- Python 的发展趋势
- 人工智能时代学习 Python 的重要性
- 本书的技术体系
- 学习本书需要注意的事项

0.1 Python 的发展趋势

也许几年前提到 Python 语言，很多人知之甚少，然而最近一两年不管是程序员，还是普通的计算机爱好者，都逐渐接触到这门编程语言， Python 的关注量也占据榜首。

知名 IT 技术问答社区 Stack Overflow 最近公布了程序语言排行榜，排名结果如下图所示。



从 Stack Overflow 编程语言流行趋势中可以看到，Python 的热度在过去几年中一直在迅速增长，已经成为目前热度增长最快的编程语言。2017 年 6 月，Python 第一次成为高收入国家在 Stack Overflow 中访问量最大的编程语言。

因为 Python 语言的语法非常简单易懂，所以很多提及编程就恐慌的人少了一些担心，Python 语言不仅在学术上非常受欢迎，而且很多非计算机专业的人也在学习 Python。他们利用自己写的简单的小程序，让生活变得精彩起来。

0.2 人工智能时代学习 Python 的重要性

现在的社会是一个高速发展的社会，科技发达，信息流通，人们之间的