新时代·技术新未来

深入浅出Python量化交易实战

段小手　著

清華大學出版社
北　京

深入浅出Python量化交易实战　四校　文前.indd　1　2021/12/1　20:46:05

内 容 简 介

本书主要以国内A股市场为例，借助第三方量化交易平台，讲述了KNN、线性模型、决策树、支持向量机、朴素贝叶斯等常见机器学习算法在交易策略中的应用，同时展示了如何对策略进行回测，以便让读者能够有效评估自己的策略。

另外，本书还讲解了自然语言处理（NLP）技术在量化交易领域的发展趋势，并使用时下热门的深度学习技术，向读者介绍了多层感知机、卷积神经网络，以及长短期记忆网络在量化交易方面的前瞻性应用。

本书没有从 Python 基础语法讲起，对于传统交易策略也只是一带而过，直接将读者带入机器学习的世界。本书适合对 Python 语言有一定了解且对量化交易感兴趣的读者阅读。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。举报：010-62782989，beiqinquan@tup.tsinghua.edu.cn。

图书在版编目(CIP)数据

深入浅出 Python 量化交易实战 / 段小手著 . —北京：清华大学出版社，2021.11
（新时代 • 技术新未来）
ISBN 978-7-302-58748-4

Ⅰ . ①深… Ⅱ . ①段… 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 (2021) 第 146402 号

责任编辑：刘 洋
封面设计：徐 超
版式设计：方加青
责任校对：王荣静
责任印制：曹婉颖

出版发行：清华大学出版社
网 址：http://www.tup.com.cn，http://www.wqbook.com
地 址：北京清华大学学研大厦 A 座 **邮 编：**100084
社 总 机：010-62770175 **邮 购：**010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：三河市金元印装有限公司
经 销：全国新华书店
开 本：187mm×235mm **印 张：**16.75 **字 数：**311 千字
版 次：2021 年 12 月第 1 版 **印 次：**2021 年 12 月第 1 次印刷
定 价：99.00 元

产品编号：090544-01

前　言

总有些人，不甘平凡

不管是在上学的时候，还是走向工作岗位之后，大家可能都会遇到这样的人——他们从不满足于平凡的现状，常常想着如何让自己的生活变得更好。于是他们比别人花更多的时间“泡”图书馆、查资料、加班，努力学习，认真思考，总想找到一条让自己的人生发生“质变”的路。

有句“鸡汤”是这么说的：上天总会奖励那些偷偷努力的人。虽说“鸡汤”适量“饮用”就好，不过说真的，就我们目前观察，努力的人大概率还是能够提升自己的生活品质的。不过大家要注意，这里说的是“大概率”，而不是说“肯定”。因为即使你很努力，但是努力的方向选得不太好，那很可能会事倍功半。

就像我们这本书的主角——小瓦，她有足够的动力去改变自己的生活，也为了这个目标很努力地在学习和思考，但她需要一个很好的努力方向来实现自己的人生目标，而本书的目的也是要帮助她厘清思路，使她掌握相关的技能。

当然，在培养起兴趣，并掌握了基本技能之后，小瓦未来还要再接再厉，勇攀高峰。常言道，“功夫不负有心人”。相信只要坚持不懈地钻研，小瓦最终会过上自己想要的生活（小瓦的名字来源于北欧神话中女武神 Valkyria 的名字，暗指她未来将会有很高的成就）。

本书会带给读者什么

这里要强调的是，本书并不是鼓励大家都去一头扎进量化交易当中，把它当作实现“一夜暴富”的捷径。实际上，本书写作的初衷是：在股票交易这个门槛较低，且数据相对比较完善的场景当中，让读者可以像小瓦一样，提高自己的数据分析能力，掌握机器学习技能。

再直白一点地说，如果读者朋友和小瓦有类似的动机和背景，那本书带给你的是一种

深入浅出Python量化交易实战 四校 文前.indd 1 2021/12/1 20:46:05

思维方式和比较通用的技能：你会更加重视数据在各种业务场景中的重要性；了解如何使用科学的方式去分析和解决业务问题，如何使用算法在多个维度的数据中找到最有用的信息，如何使用机器学习和深度学习技术对目标做出预测，以及如何处理文本数据等。

退一步讲，即使读者朋友没有使用量化交易技术进行投资，依靠上述技能仍然可以获得一份不错的收入（如找一份数据分析师的工作，或者是兼职帮别人做点事情），这样也算是“进可攻退可守”的策略了。

本书内容及体系结构

本书以完全没有量化交易经验的对象视角，从最基础的环境搭建开始进行讲解，并直接带读者进入多因子、机器学习的时代。本书后半部分更是紧贴国际前沿趋势，介绍了 NLP 技术在量化交易领域的应用，以便读者参考。

本书更多地采用启发式的方法，让读者朋友能够跟着相关内容不停思考和尝试，而不是简单地照搬现有策略。

第 1 章，先是简单介绍了小瓦的情况，以便让读者朋友们更有代入感；紧接着对人类的交易历史做了简明的阐述，主要是为了让读者朋友们可以了解交易技术的发展脉络，对量化交易的概念有基本的认知；随后直接带读者进入环境配置的环节，并使用真实的股票数据集进行实验，让喜欢实操的读者对量化交易有一个直观的感受，并产生兴趣。

第 2 章，通过对小瓦策略的回测，向读者朋友介绍了回测的基本概念和简单方法；随后简单介绍了一些历史上比较经典的交易策略——移动平均策略和海龟策略，并通过简单的回测让读者朋友对这两种策略的收益情况有一个大致的了解。

第 3 章，就开始让读者朋友们和小瓦一起接触机器学习的概念了。本章先是用通俗易懂的例子讲解了有监督学习和无监督学习的概念；然后对机器学习中的分类和回归进行了阐述；随后用真实的股票数据训练了一个简单的 KNN 模型，并基于 KNN 模型编写了简单的交易策略，最后对这个策略进行了简单的回测。

第 4 章，为了能够让小瓦和读者朋友们更加专注于策略的编写和回测，我们挑选了一个第三方量化交易平台，并基于该平台的研究环境，讲解了如何获取股票的概况数据、财务数据、股东数据、主力资金流入 / 流出数据等。从本章开始，本书附带的代码，需要在该平台上运行。

第 5 章，开始介绍时下流行的因子。为了给读者朋友带来更多启发，我们让小瓦发动

深入浅出Python量化交易实战 四校 文前.indd 2 2021/12/1 20:46:05

自己的聪明才智设计了一个专属于自己的“瓦氏因子”，并借此向读者朋友展示了因子的基本原理；随后使用代码，通过第三方量化交易平台获取了股票的市值因子、现金流因子、净利率因子等；最后别出心裁地使用了无监督学习的主成分分析（PCA）算法，把上述一些因子进行了“打包”，并借此进行了选股的实验。

第 6 章，介绍了对因子进行分析的方法。这里我们建立了一个实验用的投资组合，并以“成交量的 5 日指数移动平均”因子为例，介绍了如何对因子进行收益分析、因子的 IC 分析，以及因子换手率、因子相关性和因子预测能力的分析，以便让读者朋友对因子的评价方法有基本的了解，并可以掌握相关的方法。

第 7 章，开始将机器学习与多因子进行结合。介绍了机器学习中的线性模型算法，包括最基础的线性回归算法和使用正则化的岭回归算法，并使用实验数据集对两种算法进行了对比；紧接着使用了多个因子与线性模型结合，编写了策略；最后使用了第三方量化交易平台的回测功能对策略进行了回测，在这个过程中，读者朋友也可以掌握回测涉及的相关指标。

第 8 章，将机器学习算法与多因子的结合更进一步。本章不仅仅介绍了决策树与随机森林算法，更是使用了决策树的判断特征重要性的功能对若干个因子的重要性进行了计算；随后，我们使用了决策树“认为”比较重要的因子，结合随机森林算法编写了交易策略；同样地，也再次对策略进行了回测。

第 9 章，将近年来普遍在量化交易中表现比较好的支持向量机（SVM）带到读者面前。本章从基本的原理开始介绍，之后提出了动态因子选择策略——每次运行程序都使用决策树算法选择重要性最高的因子，再用其来训练支持向量机模型，并形成策略和进行回测。可以说，到这一章，读者朋友们可以对传统机器学习算法在量化交易领域的应用有了初步的掌握。

第 10 章，会让小瓦和读者朋友一起，进入一个更新的世界——开始探讨更加前沿的尝试：将自然语言处理技术应用于量化交易当中。本章先介绍了国际上一些知名机构在自然语言处理和量化交易方向的一些成果，随后使用真实的新闻文本数据，介绍了中文的分词方法及应用，为后面的章节打下基础。

第 11 章在第 10 章的基础上，让小瓦和读者朋友一起学习文本向量化方法，并使用潜狄利克雷分布（LDA）进行话题建模技术的学习。通过本章的学习，小瓦将能够掌握如何使用机器学习算法，从大量文本中快速获取话题。

第 12 章，开始让小瓦和读者朋友一起，进行文本数据情感分析的实验。在本章中，我们还会继续使用到文本向量化方法；之后使用了在文本分类中非常常用的朴素贝叶斯算法。

深入浅出Python量化交易实战 四校 文前.indd 3 2021/12/1 20:46:05

在完成本章的学习之后，读者朋友将会和小瓦一样，掌握文本分类的基本方法。

第 13 章，我们让小瓦开始接触神经网络算法。作为近几年大热的人工智能算法，神经网络不论是在图像识别方面还是在文本分类方面，都有广泛的应用。本章主要向小瓦和读者朋友介绍对用户非常友好的深度学习框架——Keras，并介绍如何使用 Keras 内置的工具对文本数据进行处理；随后使用 Keras 搭建了多层感知机（全连接层）神经网络，对文本的情绪进行分类实验。

第 14 章，在小瓦已经掌握多层感知机的基础上，我们进一步介绍了卷积神经网络和长短期记忆网络的原理，并基于文本分类任务，分别训练了卷积神经网络和长短期记忆网络。经过本章的学习之后，小瓦和读者朋友可以掌握卷积神经网络和长短期记忆网络的原理和基本的训练方法。

第 15 章，提出了一些问题，并对小瓦未来学习和研究的方向给出了一些建议，也给读者朋友提供了一些参考。

本 书 特 色

1. 视角独特，让读者能够有较强的代入感

本书以一个女大学生的故事作为开篇，将她的动机和基础条件列出——想要改善自己的生活状态，并且有一些 Python 基础，对于量化交易只是听说过，但从来没有接触过。因此本书站在她的角度，内容安排由浅入深，层层递进，让读者能够轻松入门，不会觉得难于上手。

2. 内容新颖，核心内容单刀直入

本书没有花篇幅去讲解 Python 基础语法、数据类型等基础知识（因为市面上此类的书籍已经很多），而是本着实用性原则，给读者朋友呈现的都是核心内容，文字与代码精练，读者朋友容易上手。

3. 紧跟国际潮流，给读者更前沿的视野

本书的另一特点是，对传统的交易策略一带而过。作者重点对目前国内外热门的机器学习算法相关应用进行了系统的阐述。对于国际上先进的自然语言处理技术和神经网络的应用，作者更是不吝笔墨，为读者朋友做了系统的介绍。这样做的目的是让读者朋友能够了解国际上更加前沿的做法，并给读者朋友留下思考和探索的空间，能够让读者朋友在入门之后，有进一步努力的方向。

深入浅出Python量化交易实战 四校 文前.indd 4 2021/12/1 20:46:06

本书读者对象

- 学习 Python，但没找到应用场景的朋友；
- 对股票投资感兴趣，想尝试量化交易的朋友；
- 有意向进入投资机构，成为一名量化基金经理的朋友；
- 平时工作不是很忙，打算做点副业的朋友。

由于作者水平有限，本书难免会有疏漏之处，欢迎各位读者朋友指正。

作　者

深入浅出Python量化交易实战 四校 文前.indd 5 2021/12/1 20:46:06

深入浅出Python量化交易实战 四校 文前.indd 6 2021/12/1 20:46:06

目 录

第 1 章 小瓦的故事——从零开始

1.1 何以解忧，“小富”也行……1
1.1.1 那些年，那些交易 ……2
1.1.2 自动化交易和高频交易 ……2
1.1.3 因子投资悄然兴起 ……3
1.2 机器学习崛起……4
1.2.1 量化投资风生水起 ……4
1.2.2 没有数据是不行的 ……5
1.2.3 交易策略和阿尔法因子 ……5
1.3 要想富，先配库……6
1.3.1 Anaconda 的下载和安装……6
1.3.2 Jupyter Notebook 的基本使用方法 ……8
1.3.3 用真实股票数据练练手 …… 11
1.4 小结…… 15

第 2 章 小瓦的策略靠谱吗——回测与经典策略

2.1 对小瓦的策略进行简单回测…… 16
2.1.1 下载数据并创建交易信号 …… 16
2.1.2 对交易策略进行简单回测 …… 18
2.1.3 关于回测，你还需要知道的 …… 20
2.2 经典策略之移动平均策略…… 21
2.2.1 单一移动平均指标 …… 21
2.2.2 双移动平均策略的实现 …… 23
2.2.3 对双移动平均策略进行回测 …… 26
2.3 经典策略之海龟策略…… 28
2.3.1 使用海龟策略生成交易信号 …… 28
2.3.2 根据交易信号和仓位进行下单 …… 29
2.3.3 对海龟策略进行回测 …… 31
2.4 小结…… 34

第 3 章 AI 来了——机器学习在交易中的简单应用

3.1 机器学习的基本概念…… 35
3.1.1 有监督学习和无监督学习 …… 35

深入浅出Python量化交易实战 四校 文前.indd 7 2021/12/1 20:46:06

3.1.2 分类和回归 …… 37
3.1.3 模型性能的评估 …… 37
3.2 机器学习工具的基本使用方法 …… 37
3.2.1 KNN 算法的基本原理 …… 38
3.2.2 KNN 算法用于分类 …… 38
3.2.3 KNN 算法用于回归 …… 43
3.3 基于机器学习的简单交易策略 …… 47
3.3.1 获取股票数据 …… 47
3.3.2 创建交易条件 …… 49
3.3.3 使用分类算法制定交易策略 …… 50
3.4 小结 …… 54

第 4 章 多来点数据——借助量化交易平台

4.1 数据不够，平台来凑 …… 55
4.1.1 选择量化交易平台 …… 56
4.1.2 量化交易平台的研究环境 …… 57
4.1.3 在研究环境中运行代码 …… 58
4.2 借助财务数据筛选股票 …… 59
4.2.1 获取股票的概况 …… 60
4.2.2 获取股票的财务数据 …… 62
4.2.3 通过财务指标进行选股 …… 64
4.3 谁是幕后“大佬” …… 65
4.3.1 找到最大的股东 …… 66
4.3.2 大股东们增持了还是减持了 …… 67
4.3.3 资金净流入还是净流出 …… 69
4.4 小结 …… 71

第 5 章 因子来了——基本原理和用法

5.1 “瓦氏因子”了解一下 …… 72
5.1.1 获取主力资金流向数据 …… 73
5.1.2 简易特征工程 …… 74
5.1.3 “瓦氏因子”的计算 …… 75
5.1.4 用添加“瓦氏因子”的数据训练模型 …… 76
5.1.5 “因子”都能干啥 …… 77
5.2 股票不知道怎么选？因子来帮忙 …… 78
5.2.1 确定股票池 …… 78
5.2.2 获取沪深两市的全部指数 …… 79
5.2.3 获取股票的市值因子 …… 80
5.2.4 获取股票的现金流因子 …… 81
5.2.5 获取股票的净利率因子 …… 82
5.2.6 获取股票的净利润增长率因子 …… 83
5.3 把诸多因子“打个包” …… 84
5.3.1 将 4 个因子存入一个 DataFrame …… 84
5.3.2 使用 PCA 提取主成分 …… 85
5.3.3 找到主成分数值最高的股票 …… 86
5.4 小结 …… 87

深入浅出Python量化交易实战 四校 文前.indd 8 2021/12/1 20:46:06

第 6 章　因子好用吗——有些事需要你知道

6.1　针对投资组合获取因子值 …… 88
6.1.1　建立投资组合并设定日期 …… 88
6.1.2　获取一个情绪因子 …… 90
6.1.3　获取全部的因子分析结果 …… 91
6.2　因子收益分析 …… 92
6.2.1　因子各分位统计 …… 92
6.2.2　因子加权多空组合累计收益 …… 94
6.2.3　做多最大分位做空最小分位收益 …… 96
6.2.4　分位数累计收益对比 …… 97
6.3　因子 IC 分析 …… 98
6.3.1　因子 IC 分析概况 …… 99
6.3.2　因子 IC 时间序列图 …… 99
6.3.3　因子 IC 正态分布 Q-Q 图和月度均值 …… 101
6.4　因子换手率、因子自相关性和因子预测能力分析 …… 102
6.4.1　因子换手率分析 …… 103
6.4.2　因子自相关性分析 …… 104
6.4.3　因子预测能力分析 …… 106
6.5　小结 …… 107

第 7 章　当因子遇上线性模型

7.1　什么是线性模型 …… 108
7.1.1　准备用于演示的数据 …… 108
7.1.2　来试试最简单的线性回归 …… 110
7.1.3　使用正则化的线性模型 …… 113
7.2　用线性模型搞搞交易策略 …… 115
7.2.1　准备因子 …… 115
7.2.2　训练模型 …… 117
7.2.3　基于模型的预测进行选股 …… 118
7.3　能不能赚到钱 …… 119
7.3.1　平台的策略回测功能 …… 120
7.3.2　把研究成果写成策略 …… 121
7.3.3　回测 …… 124
7.4　小结 …… 126

第 8 章　因子遇到决策树与随机森林

8.1　什么是决策树和随机森林 …… 127
8.1.1　线性模型不适用的数据样本 …… 127
8.1.2　决策树的用法和原理 …… 129
8.1.3　随机森林的用法和原理 …… 130
8.2　哪些因子重要，决策树能告诉你 …… 132
8.2.1　多来点因子 …… 132
8.2.2　设定目标并训练模型 …… 135
8.2.3　哪些因子重要 …… 137
8.3　用重要因子和随机森林来制订策略 …… 138

深入浅出Python量化交易实战 四校 文前.indd 9　2021/12/1 20:46:06

8.3.1 回测函数的初始化 ……138
8.3.2 盘前的准备工作……139
8.3.3 策略中的机器学习部分 ……141
8.3.4 定义买入股票和卖出股票的列表 ……142
8.3.5 定义买入操作和卖出操作 ……144
8.3.6 对策略进行回测……145
8.4 小结……146

第 9 章 因子遇到支持向量机

9.1 什么是支持向量机……147
9.1.1 支持向量机的基本原理 ……147
9.1.2 线性内核有时“很着急” ……149
9.1.3 RBF 内核“闪亮登场” ……150
9.2 动态因子选择策略……152
9.2.1 设置回测环境……152
9.2.2 开盘前准备……153
9.2.3 机器学习的部分……155
9.2.4 买入和卖出的操作 ……157
9.3 策略的回测详情……158
9.3.1 策略收益概述……159
9.3.2 策略交易详情……159
9.3.3 持仓和收益详情……161
9.4 使用策略进行模拟交易……162
9.4.1 模拟交易……163
9.4.2 查看模拟交易详情 ……164
9.4.3 模拟交易的持仓与下单 ……165
9.5 小结……166

第 10 章 初识自然语言处理技术

10.1 我们的想法是否靠谱 ……167
10.1.1 思考几个问题……167
10.1.2 参考一下“大佬”们的做法……168
10.1.3 说了那么多，什么是 NLP……169
10.2 获取文本数据并简单清洗 ……170
10.2.1 获取新闻联播文本数据 ……170
10.2.2 对文本数据进行简单清洗 ……172
10.3 中文分词，“结巴”来帮忙 ……173
10.3.1 使用“结巴”进行分词……174
10.3.2 使用“结巴”进行列表分词……174
10.3.3 建立停用词表 ……175
10.3.4 去掉文本中的停用词……176
10.3.5 使用“结巴”提取关键词……178
10.4 小结 ……180

第 11 章 新闻文本向量化和话题建模

11.1 让机器“读懂”新闻 ……181
11.1.1 准备文本数据……181

深入浅出Python量化交易实战 四校 文前.indd 10 2021/12/1 20:46:06

11.1.2 使用 CountVectorizer 将文本转化为向量 ····· 183
11.1.3 使用 TfidfVectorizer 将文本转化为向量 ····· 185
11.2 让机器告诉我们新闻说了啥 ····· 186
11.2.1 什么是话题建模 ····· 186
11.2.2 什么是 LDA 模型 ····· 187
11.3 话题建模实战 ····· 188
11.3.1 加载数据并进行分词 ····· 188
11.3.2 将分词结果合并保存 ····· 190
11.3.3 使用 LDA 进行话题建模 ····· 191
11.3.4 对模型进行改进 ····· 192
11.4 小结 ····· 194

第 12 章 股评数据情感分析

12.1 机器懂我们的情感吗 ····· 195
12.1.1 了解分好类的语料 ····· 196
12.1.2 将文件上传到量化交易平台 ····· 197
12.2 用语料制作数据集 ····· 198
12.2.1 将正面情绪语料存储为列表 ····· 198
12.2.2 将负面情绪语料存储为列表 ····· 200
12.2.3 给数据“打上标签” ····· 201
12.2.4 合并正负面情绪语料 ····· 202
12.3 隆重推出“朴素贝叶斯” ····· 203
12.3.1 “朴素贝叶斯”又是什么 ····· 204
12.3.2 为贝叶斯模型准备数据 ····· 205
12.3.3 开始训练贝叶斯模型并评估其性能 ····· 206
12.4 小结 ····· 208

第 13 章 咱也“潮”一把——深度学习来了

13.1 开始研究前的准备 ····· 209
13.1.1 翻翻工具箱，看看有什么 ····· 210
13.1.2 为神经网络准备数据 ····· 211
13.2 使用 Keras 对文本进行预处理 ····· 213
13.2.1 使用 Tokenizer 提取特征 ····· 213
13.2.2 将文本转化为序列 ····· 214
13.2.3 填充序列与转化矩阵 ····· 216
13.3 使用 Keras 构建简单神经网络 ····· 217
13.3.1 先动手“撸”一个多层感知机 ····· 217
13.3.2 念叨一下多层感知机的原理 ····· 218
13.3.3 再来说说激活函数 ····· 220
13.3.4 Dropout 层又是干吗的 ····· 221
13.3.5 训练一下，看看效果如何 ····· 222
13.4 小结 ····· 224

深入浅出Python量化交易实战 四校 文前.indd 11 2021/12/1 20:46:07

第 14 章　再进一步——CNN 和 LSTM

14.1　先动手“撸”一个卷积神经网络 ……225
14.1.1　准备好库和数据集 ……225
14.1.2　处理数据与搭建模型 ……227
14.2　卷积神经网络模型详解 ……229
14.2.1　嵌入层是干啥用的 ……230
14.2.2　卷积层是干啥用的 ……231
14.2.3　最大池化层是干啥用的 ……233
14.2.4　训练模型看看效果 ……234
14.3　长短期记忆网络 ……236
14.3.1　搭建一个简单的长短期记忆网络 ……236
14.3.2　关于长短期记忆网络 ……237
14.3.3　训练模型及评估 ……238
14.3.4　保存模型并在回测中调用 ……240
14.4　小结 ……241

第 15 章　写在最后——小瓦的征程

15.1　可以一夜暴富了吗 ……242
15.1.1　使用第三方量化平台是个好主意吗 ……243
15.1.2　机器学习到底有没有用 ……243
15.1.3　要“吊死”在 A 股“这棵树”上吗 ……244
15.2　将来要做什么 ……245
15.2.1　学习一些数据库知识 ……245
15.2.2　多看看不同的投资标的 ……247
15.2.3　打开国际化的视野 ……249
15.3　小结 ……252

深入浅出Python量化交易实战 四校 文前.indd 12　2021/12/1 20:46:07

第 1 章　小瓦的故事——从零开始

本书源于一个真实的故事，故事的主角是一位名叫小瓦的姑娘。小瓦出生在一个普通的家庭，父母都是老实淳朴的普通人，靠着并不丰厚的收入把小瓦养育成人。18 岁那年，小瓦考上了一所不好不坏的大学，所学专业是一个就业前景算不上理想的专业。再加上她本身也谈不上出色，说她是一个现实版的“灰姑娘”也不为过。综上所述，小瓦应该是一个有点危机感的孩子，实际上她也确实有改变现状的想法。因此，我们的任务就是帮助小瓦实现她的愿望。当然，千里之行，始于足下。在本章中，我们先要了解一些基础知识。本章的主要内容如下。

- 简要回顾历史上的交易。
- 从自动化交易到因子投资。
- 机器学习技术的崛起。
- 环境配置及常用工具的基本使用。

1.1　何以解忧，“小富”也行

和其他女孩子一样，小瓦也有爱美的天性。她也想像其他女生一样把自己打扮得漂漂亮亮，并且顺利完成学业，如果条件允许，她还想继续深造。然而，其家庭的现实情况支撑不了她的梦想，所以平日里小瓦的生活还是很简朴的。小瓦希望找到一个方法，能够让她的状况变得好一些，至少可以通过自己的努力为父母减轻一些经济压力。在现实中，能够通过自己的努力，先“小富”起来，这已经是很不错的了。

想要实现“小富”，其实途径还是蛮多的。如打打零工，或者做点小生意，都是可以的。不过小瓦有自己的优势——她因为所学的专业担心自己毕业以后不好找工作，于是自学了 Python 的入门课程，并掌握了一些 Python 的

深入浅出Python量化交易实战 四校 正文.indd 1　　2021/12/1 20:49:58

基础语法和常用工具，如数据分析工具 pandas 和可视化工具 matplotlib 等。基于这样的背景，我们可以尝试帮助小瓦利用她已经掌握的知识来做点更有技术含量的事情——量化交易。这样的话，即便小瓦最终没有成为一代“股神”，也可以掌握更多的知识和技能，为其日后找工作增加一点儿优势。

考虑到小瓦所学专业既不是计算机相关专业，也不是金融相关专业，我们就先让小瓦了解一些基础知识。

1.1.1 那些年，那些交易

相信大家对“交易”这个词并不陌生。早在古罗马时代，人类就开始进行各种各样的交易了。对历史感兴趣的读者朋友可能会听说过古罗马广场。在那里，人们除了进行实物商品的交易之外，还会进行交换货币、债券及其他形式的投资。人们进行交易的目的就是获利，因此人们在某种商品价格较低时买入商品，再等到价格较高时卖出商品，以此实现盈利的目的。

1602 年，世界上第一个股份制公司——荷兰东印度公司诞生，并在 1606 年发行了世界上第一只股票。1609 年，世界上第一个股票交易所在阿姆斯特丹诞生，之后在这个股票交易所的基础上，世界上第一家现代意义的银行——阿姆斯特丹银行成立。至于后来荷兰东印度公司借助资本的力量，建立了怎样的霸业，可以留给读者朋友自行搜索。本书不展开叙述。

中国的资本市场起步相对较晚。中华人民共和国成立以后的第一只股票是上海飞乐音响股份有限公司发行的“飞乐音响”，于 1984 年发行。中国在 1989 年才开始进行股票交易市场的试点。1990 年，深圳证券交易所和上海证券交易所开始试营业。

1.1.2 自动化交易和高频交易

时光荏苒，岁月变迁，曾经的历史我们一带而过，现在该把小瓦拉回现代社会了。在过去的几十年间，投资行业可以说是发生了翻天覆地的变化——那些股民们扯着嗓子喊单的日子一去不复返，取而代之的是高速运转的计算机设备和各种各样的应用软件。尤其各大金融机构，为了保持自己的竞争力，以便应对日益复杂的市场环境，一直都走在最新科技前沿。

深入浅出Python量化交易实战 四校 正文.indd 2 2021/12/1 20:49:58

说到最新技术，就要回顾一下电子交易的概念——毕竟这是现代交易的根基。自 20 世纪 60 年代以来，计算机和互联网的诞生使交易不再受物理空间的限制，交易的品种增加，交易量也得到了极大的提升。随着电子交易的兴起，自动化交易也得到了迅速发展。到了 2000 年左右，自动化交易作为一种卖方工具出现，其目的是实现一种低成本、高收益的交易。在自动化交易中，订单被分散，以避免订单过大而影响市场。后来这些工具扩展到买方，并在交易中，加入对成本和流动性的考虑，以及期望预测短期价格和成交量，这使得这个领域涉及的技术越来越复杂。

读者朋友也听说过高频交易（High-Frequency Trading，HFT）这个术语。高频交易也是建立在电子交易的基础之上的。高频交易是指在微秒范围内以极低延迟执行的金融工具的自动交易。在过去的十余年中，使用高频交易的成交量大幅增长。据估计，2010 年前后，高频交易成交量约占美国股市交易量的 55%，占欧洲股市交易量的 40%。在外汇期货市场中，通过高频交易实现的成交量更是达到了惊人的 80% 左右！

1.1.3 因子投资悄然兴起

对于小瓦来说，“因子投资”这个词可能就有一点陌生了。不过没有关系，要理解因子投资的概念并不难。首先我们了解一下资本资产定价模型（Capital Asset Pricing Model，CAPM）。这个概念早在 20 世纪 60 年代就已经被提出，其非常重要的一个组成部分就是贝塔系数（beta）。贝塔系数用于表示某项资产的系统性风险。资本资产定价模型理论的先进性在于，它认识到一项资产的风险并非孤立地取决于该资产，而是该资产相对于其他资产及整个市场的走势。然而，它的局限性也非常明显——由于只使用贝塔系数这一个因子（factor），该模型对资产回报率的预测并不准确。于是，人们开始发掘更多额外的风险因子，以便可以更准确地预测资产能够带来的回报率。基于这种理念所进行的投资称为因子投资（factor investment）。

在因子投资中，因子的定义就是那些可以量化的信号、特征或其他变量。这些因子与资产回报率呈现明显的相关性，并且在未来还要保持这种相关性。

与 CAPM 模型相比，多因子模型要考虑的因素要多得多。例如，1972 年，史蒂芬·罗斯提出的套利定价理论（Arbitrage Pricing Theory，APT）就认为证券的回报率与一组因子线性相关。到了 1992 年，著名经济学家尤金·法马和金融学家肯尼思·弗伦奇提出了举世闻名的 Fama-French 三因子模型，指出证券回报率可以用市场资产组合、市值及账面市值

深入浅出Python量化交易实战 四校 正文.indd 3 2021/12/1 20:49:58

比这三个因子来进行解释。不久，卡哈特改进了三因子模型，添加了市场动量因子，提出了 Carhart 四因子模型。

经过若干年的发展，因子投资已经从一种方法论演变成了一种产品——智能贝塔基金，并且飞速增长。有资料显示，早在 2019 年，智能贝塔基金的资产管理规模就达到了 8800 亿美元。这种基于多因子模型进行投资的成功推动了机器学习技术在金融行业的飞速发展。

1.2 机器学习崛起

说起当今全球知名的几家使用机器学习算法进行交易的基金公司，不得不提到诸如文艺复兴科技公司、德邵基金公司、堡垒投资集团和 Two Sigma 公司等。其中，文艺复兴科技公司是由数学家詹姆斯·西蒙斯在 1982 年创建的，现在已经成为全球较大的量化投资公司。它的奖章基金自成立以来，年均收益率较稳定。另外三家基金公司也因为使用基于算法的交易策略之后，业绩表现亮眼，从而跻身全球前列。

与此同时，越来越多的基金公司转向机器学习技术。人工智能系统可以从大量数据中学习，并且持续进化。从对冲基金到共同基金，越来越多的机构使用算法制定交易策略。这也催生出了一个新的物种——量化基金（Quant funds）。

1.2.1 量化投资风生水起

随着新技术的普及，以往靠基本面分析进行投资的公司也在引入机器学习技术，并开展量化交易。例如，管理资产规模较大的 Point72 公司就已经有相当比例的投资组合是使用人与算法结合的方法来进行管理的。据摩根士丹利公司所进行的调查来看，截至 2020 年，在他们从事投资的客户中，有意愿或者已经在使用机器学习技术进行量化投资的客户数量，在短短两年间提高了 20%。

黑石集团为了能够立于不败之地，大举投资量化交易企业 SAE，希望借这场豪赌击败竞争对手。无独有偶，富兰克林·邓普顿基金集团也收购了一家名为“随机森林资本”的公司（一家以算法和数据为导向的技术型企业）。

在中国，很多金融机构也已经觉醒。相信在未来几年，我们也能看到优秀的本土量化投资公司。

深入浅出Python量化交易实战 四校 正文.indd 4 2021/12/1 20:49:59

1.2.2 没有数据是不行的

所谓机器学习，其实就是通过若干算法，使用数据训练模型并做出预测的过程。那么数据的重要性就不言而喻了。假如模型是马路上跑的汽车，那么数据就是让汽车正常运转的燃料。常规的数据包括经济统计数据、市场交易数据和上市公司财报。如今，人们使用的数据范围更广，甚至包括卫星图像、信用卡销售、股民情绪分析、手机地理位置定位和爬虫抓取等来源。理论上来讲，我们这里说的数据包括任何可以使用机器学习提取交易信号的信息。

举一个例子，如果在某家上市公司公布财报数据之前，我们可以获取该公司在招聘网站上发布的招聘岗位数量，就可以先于财报数据发布了解到该公司的运营状况。假如该公司的招聘人数在上升，则可能说明该公司业绩良好，自然其股票的价格也可能上涨；反之，假如该公司的招聘人数锐减，则说明该公司的经营可能有困难，则可能会导致该公司股价下跌。

当然，最直接有效的数据还是那些能够直接体现用户消费的数据，如支付数据。在后面的章节，我们会具体来讨论如何使用外部数据，并将其添加到模型的训练当中。

1.2.3 交易策略和阿尔法因子

说完了算法和数据，下面我们就要讲讲如何将二者结合起来并应用到投资中。我们先通过各种数据源提取出有效信息，并且通过特征工程（feature engineering）将数据转换为阿尔法因子（alpha factor），再将这些因子拿来训练模型，使模型可以对交易品未来的趋势或价格变动做出预测，并触发买单或者卖单。例如，模型预测次日股价大涨，则下单买入，反之则卖出。

在这个过程中，阿尔法因子（以下统称为 alpha 因子）的确定是一个复杂的工程。在这一步中，我们需要探索输入的数据与目标收益之间的关系，并进行复杂的特征工程，还要不断测试及对模型进行调优，以此来优化模型的预测能力。

当然，经过数十年的学术研究，金融学家已经帮我们总结好 alpha 因子。如果要给小瓦完全解释明白这背后的理论，我们就要花很大的篇幅去讲解市场金融理论和投资者行为学等。不过这些不是主线“剧情”，因此本书不展开介绍了。

深入浅出Python量化交易实战 四校 正文.indd 5 2021/12/1 20:49:59

1.3 要想富，先配库

我们在前面小节中唠叨了很多内容，相信很多读者朋友也有和小瓦有一样的想法：说了这么多，我们到底怎么开始呢？下面就给小瓦安排一些动手的环节。过去人们常说“要想富，先修路”。也就是说，只有搞好基础设施建设，才能带动经济发展。在量化交易这个环节，基础设施就是我们使用的工具了。下面咱们就带小瓦一起，把需要的软件环境配置好。

1.3.1 Anaconda的下载和安装

鉴于小瓦唯一会使用的编程语言是 Python，我们选择可以支持 Python 的工具。目前比较主流的编辑器包括 PyCharm、Jupyter Notebook 等。考虑到小瓦现在主要使用的操作系统是 Windows，所以强烈建议小瓦直接安装 Anaconda。原因很简单，Anaconda 内置了 Python 解释器和 Jupyter Notebook，而且习惯使用 PyCharm 的读者朋友也可以直接调用 Anaconda 中的 Python 解释器，非常方便；更重要的是，Anaconda 内置了很多常用的数据科学库，免去了我们手动安装这些库的过程（要知道如果自己在 Windows 系统中手动安装这些库是非常痛苦的）！

下面我们就来教会小瓦如何下载和安装 Anaconda，对这部分已经有了解的读者朋友可以跳过本小节，直接阅读后面的内容。

首先，小瓦需要访问 Anaconda 官方网站：http://www.anaconda.com，单击右上角的 Download 按钮，如图 1.1 所示。

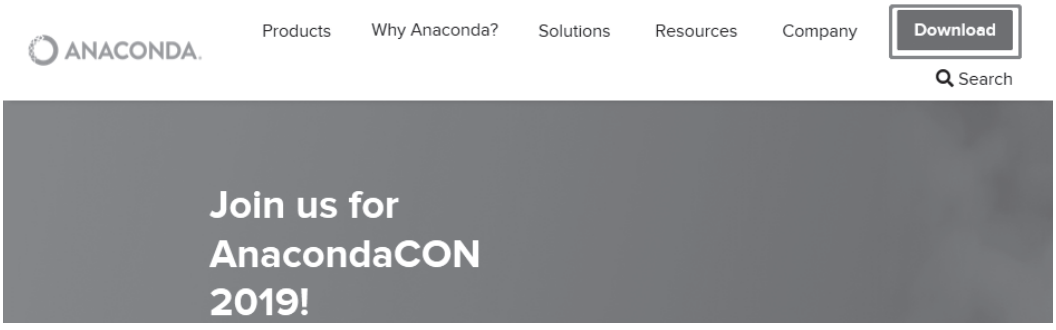

图 1.1 Anaconda 官网

深入浅出Python量化交易实战 四校 正文.indd 6 2021/12/1 20:49:59

单击 Download 按钮之后，会进入版本选择页面。小瓦使用的是 Windows 系统，所以这里选择了 Python 3.x 的 Windows 64 位版本，如图 1.2 所示。

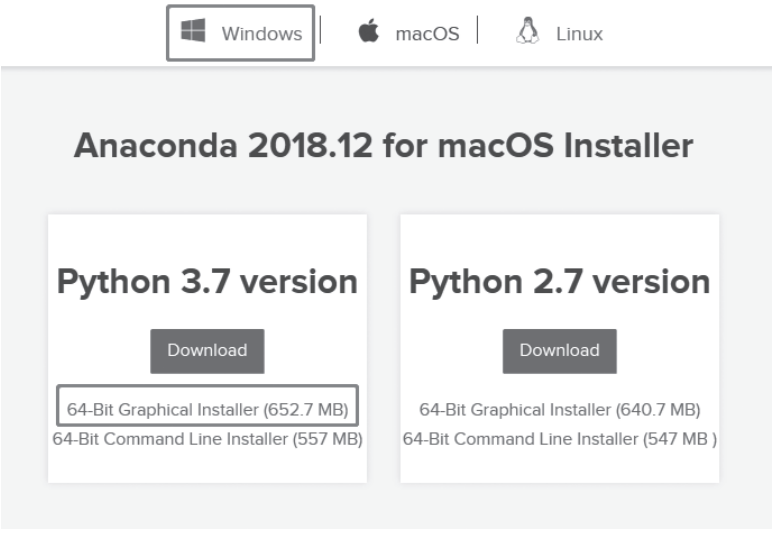

图 1.2　选择正确的 Anaconda 版本

这里，读者朋友可以根据自己操作系统的情况来选择版本：如果你使用的是苹果电脑，就选择 macOS 版本的 Anaconda；如果是 Ubuntu 或 Linux 系统，就选择 Linux 版本的 Anaconda。

等 Anaconda 的安装文件下载完毕后，双击“安装文件”即可开始安装程序。在安装过程中，各选项均保持默认设置即可。安装完成后，小瓦的“开始”菜单里就会出现图 1.3 所示的新图标。

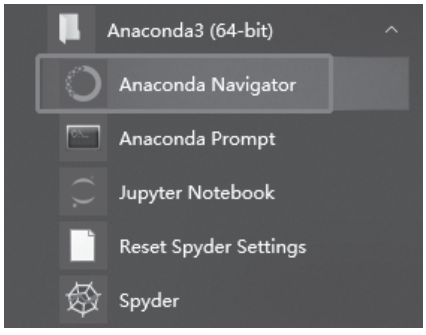

图 1.3　“开始”菜单中的新图标

深入浅出Python量化交易实战 四校 正文.indd 7 2021/12/1 20:49:59

Anaconda3（64-bit）目录下有若干个快捷方式。小瓦可以单击 Anaconda Navigator 来启动程序，也可以单击 Jupyter Notebook 来单独启动 Notebook 程序；如果需要安装新的库，可以启动 Anaconda Prompt 并使用“pip install”命令来进行安装。

在后面的内容中，小瓦使用最多的就是 Jupyter Notebook 了，所以下面我们简单来介绍一下 Jupyter Notebook 的基本操作。

1.3.2 Jupyter Notebook的基本使用方法

小瓦在单击图 1.3 中的 Anaconda Navigator 图标之后，就可以看到如图 1.4 所示的界面。

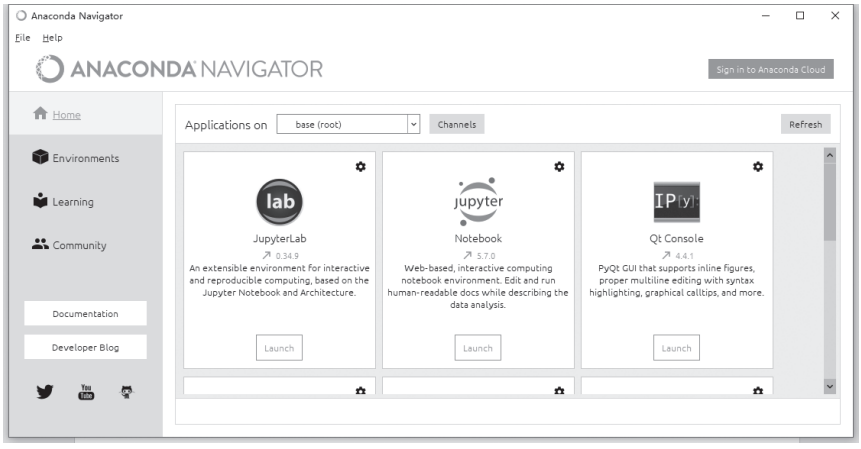

图 1.4 Anaconda Navigator 界面

在图 1.4 中，单击 Jupyter Notebook 图标下方的 Launch 按钮，就可以启动 Jupyter Notebook，如图 1.5 所示。

图 1.5 Jupyter Notebook 的启动按钮

深入浅出Python量化交易实战 四校 正文.indd 8 2021/12/1 20:49:59

单击 Launch 按钮，系统默认浏览器会自动弹出，并进入 Jupyter Notebook 的初始页面，如图 1.6 所示。

图 1.6　Jupyter Notebook 的初始页面

在进入图 1.6 所示的页面之后，通过单击右上方的 New 按钮可以新建一个 Notebook 文件，如图 1.7 所示。

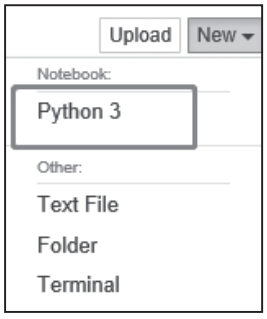

图 1.7　新建 Python 3 的 Notebook 文件

单击 New 下拉按钮之后，会出现一个下拉菜单，我们在下拉菜单中选择 Python 3，就会打开一个新的浏览器选项卡，如图 1.8 所示。

深入浅出Python量化交易实战 四校 正文.indd 9 2021/12/1 20:49:59

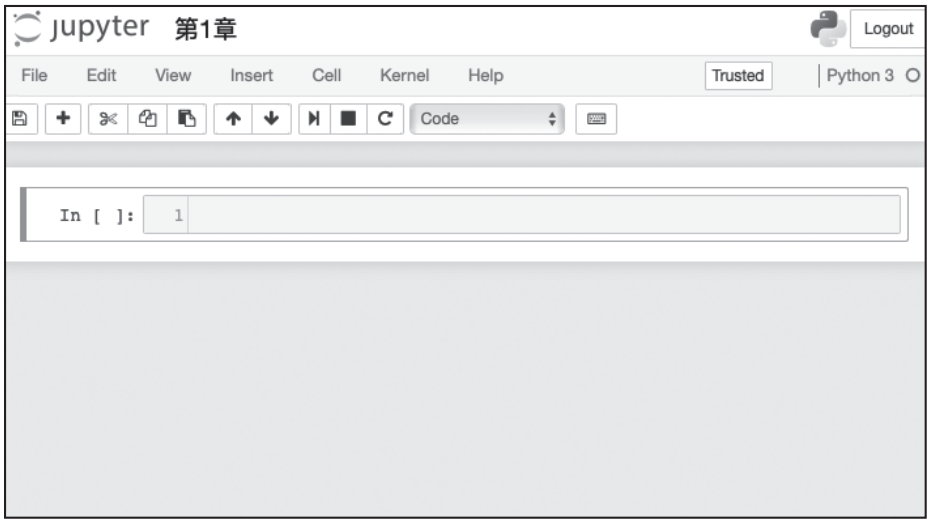

图 1.8　新建的 Python 3 Notebook 文件

现在新的 Python 3 Notebook 文件就创建成功了。接下来我们来输入第一行代码：

```
print('小瓦加油！你会成功的！')
```

按 Shift+Enter 组合键，就可以在 Jupyter Notebook 中执行代码，如图 1.9 所示。

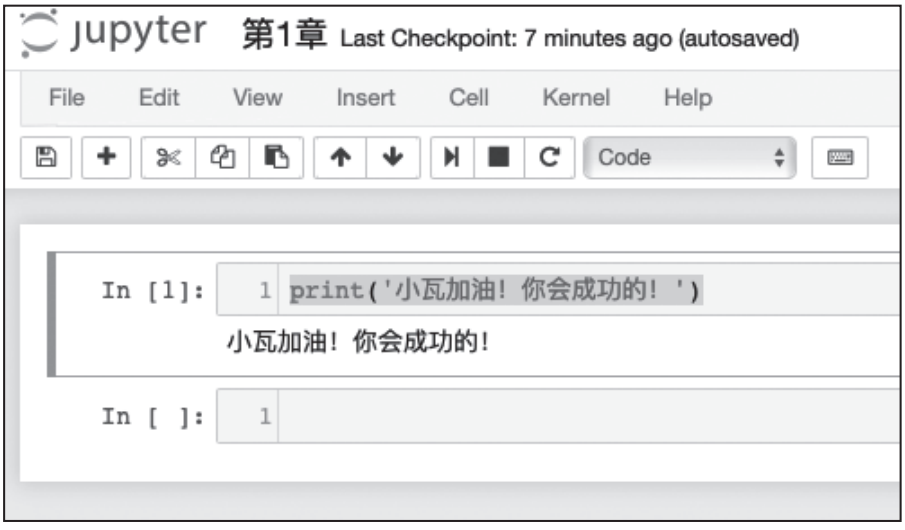

图 1.9　使用 Jupyter Notebook 执行的第一行代码

深入浅出Python量化交易实战 四校 正文.indd 10　2021/12/1 20:49:59

到此，小瓦对于在 Jupyter Notebook 中创建 Python 3 文件，以及如何输入和执行代码已经掌握了。下面我们给小瓦增加一点难度——针对真实的股票数据集进行一些简单的操作。

1.3.3 用真实股票数据练练手

在这一小节中，小瓦就开始使用真实数据集来进行练习了。这里需要用到一个获取金融数据的库——pandas_datareader。需要注意的是，编写本书时，Anaconda 并没有内置这个库，因此小瓦需要单独安装这个库。这个库的安装非常简单，只要在 Jupyter Notebook 的单元格中输入下面这行代码就可以了：

```
!pip3 install pandas_datareader --upgrade
```

运行这行代码，程序会自动下载这个库并完成安装。

注意：由于写作本书所使用的环境是 Python 2.x 和 Python 3.x 并存的，因此此处指定使用 pip3 install。如果读者朋友的电脑只安装了 Python 3.x，则直接使用 pip install 即可。

如果想检查 datareader 是否安装成功，则可以输入下面的代码：

```
#导入datareader并重命名为web
import pandas_datareader as web
#查看datareader的版本信息
web.__version__
```

运行代码，可以看到程序给我们返回 datareader 的版本号如下：

```
'0.8.1'
```

1. 下载股票数据

如果读者朋友也得到了这个结果，说明 datareader 已经安装成功。接下来我们使用 datareader 获取一只股票的数据，输入代码如下：

```
#指定获取股票数据的起始日期和截止日期
#这里就用2020年1月1日至3月18日的数据
start_date = '2020-01-01'
end_date = '2020-03-18'
#创建数据表，这里选择下载的股票代码为601318
#数据源设置为yahoo
```


深入浅出Python量化交易实战 四校 正文.indd 11 2021/12/1 20:49:59

```
#并把我们把设定的开始日期和截止日期作为参数传入
data = web.data.DataReader('601318.ss','yahoo',
                           start_date,
                           end_date)
#下面来检查一下数据表的前5行
data.head()
```

运行代码，会得到如表 1.1 所示的结果。

表 1.1　使用 datareader 获取的股票交易数据

Date（日期）	High（最高价）	Low（最低价）	Open（开盘价）	Close（收盘价）	Volume（成交量）	Adj Close（调整后的收盘价）
2020-01-02	86.790001	85.879997	85.900002	86.120003	77825207	86.120003
2020-01-03	86.879997	85.900002	86.809998	86.199997	59498001	86.199997
2020-01-06	86.870003	85.500000	85.919998	85.599998	63644804	85.599998
2020-01-07	86.459999	85.669998	86.010002	86.150002	45218832	86.150002
2020-01-08	85.000000	85.000000	85.000000	85.000000	62805311	85.000000

【结果分析】如果读者朋友也得到了与表 1.1 相同的结果，说明数据下载成功。在表 1.1 中，第一列 Date 是交易日的时期，后面依次是当日股票的最高价、最低价、开盘价、收盘价、成交量和调整后的收盘价。由于 1 月 1 日是元旦，这一天是没有交易数据的。

2. 最简单的数据处理

下面我们来做一点简单的处理工作。大家知道，股票每日的涨跌是用当日的收盘价减去前一个交易日的收盘价来计算的。例如，2020 年 1 月 3 日的收盘价约为 86.20 元，1 月 2 日的收盘价约为 86.12 元，则 1 月 3 日当天，该股票较前一个工作日上涨了 8 分钱。这样的话，可以在数据表中添加一个字段，用来表示当日股价较前一日的变化。输入代码如下：

```
#将新的字段命名为diff，代表difference
#用.diff()方法来计算每日股价变化情况
data['diff'] = data['Close'].diff()
#检查一下前5行
data.head()
```

运行代码，会得到如表 1.2 所示的结果。

深入浅出Python量化交易实战 四校 正文.indd 12　　2021/12/1 20:49:59

表 1.2 添加 diff 字段的数据表

Date（日期）	High（最高价）	Low（最低价）	Open（开盘价）	Close（收盘价）	Volume（成交量）	Adj Close（调整后的收盘价）	diff（涨跌）
2020-01-02	86.790001	85.879997	85.900002	86.120003	77825207	86.120003	NaN
2020-01-03	86.879997	85.900002	86.809998	86.199997	59498001	86.199997	0.079994
2020-01-06	86.870003	85.500000	85.919998	85.599998	63644804	85.599998	–0.599998
2020-01-07	86.459999	85.669998	86.010002	86.150002	45218832	86.150002	0.550003
2020-01-08	85.000000	85.000000	85.000000	85.000000	62805311	85.000000	–1.150002

注：NaN 表示无数据。

【结果分析】从表 1.2 中可以看到，我们成功地添加了 diff 字段，以表示该股票每日的价格变化情况。例如，在 1 月 6 日当天，该股票收盘价较上一交易日下跌了约 6 角；而 1 月 7 日当天，该股收盘价较上一交易日上涨了约 5 角 5 分。通过添加这个字段，我们即可直观地看出股票每日的价格变动。

注意：为了便于展示，表 1.2 中省略了部分字段。

3. 设计最简单的交易策略

到这里，小瓦提出了一个问题。我们可以设置一个最简单的交易策略：如果当日股价下跌，我们就在下一个交易日开盘前挂单买入；反之，如果当日股价上涨，我们就在下一个交易日开盘前挂单卖出。循环进行这个步骤，我们不就可以赚钱了吗？

说实话，这个策略听起来还挺不错。我们先不论其是否可行，就单纯地用代码来试试。要实现这个策略，首先我们来创建一个新的字段——Signal（交易信号）。如果 diff 字段大于 0，则 Signal 标记为 1；如果 diff 字段小于或等于 0，则 Signal 标记为 0。输入代码如下：

```
#此处会用到numpy，故导入
import numpy as np
#创建交易信号字段，命名为Signal
#如果diff值大于0，则Signal为1，否则为0
data['Signal'] = np.where(data['diff'] > 0, 1, 0)
#检查是否成功
data.head()
```

运行代码，我们会得到如表 1.3 所示的结果。

深入浅出Python量化交易实战 四校 正文.indd 13 2021/12/1 20:49:59

表 1.3　添加了 Signal 字段的数据表

Date（日期）	High（最高价）	Low（最低价）	Open（开盘价）	Close（收盘价）	Volume（成交量）	Adj Close（调整后的收盘价）	diff（涨跌）	Signal（交易信号）
2020-01-02	86.790001	85.879997	85.900002	86.120003	77825207	86.120003	NaN	0
2020-01-03	86.879997	85.900002	86.809998	86.199997	59498001	86.199997	0.079994	1
2020-01-06	86.870003	85.500000	85.919998	85.599998	63644804	85.599998	–0.599998	0
2020-01-07	86.459999	85.669998	86.010002	86.150002	45218832	86.150002	0.550003	1
2020-01-08	85.000000	85.000000	85.000000	85.000000	62805311	85.000000	–1.150002	0

【结果分析】从表 1.3 中可以看到，使用 np.where() 可以让程序判断每日股价是上涨还是下跌：如果上涨，则交易信号为 1，代表卖出；否则交易信号为 0，代表买入。这样我们就得到了最简单的交易信号。

4. 交易信号可视化

在得到了交易信号之后，我们可以再来实现最简单的可视化，这样我们就可以在图像中直观地看到什么时候买入，什么时候卖出了。可视化的代码部分如下：

```
#导入画图工具matplotlib
import matplotlib.pyplot as plt
#设置画布的尺寸为10*5
plt.figure(figsize = (10,5))
#使用折线图绘制出每天的收盘价
data['Close'].plot(linewidth=2, color='k', grid=True)
#如果当天股价上涨，标出卖出信号，用倒三角表示
plt.scatter(data['Close'].loc[data.Signal==1].index,
        data['Close'][data.Signal==1],
        marker = 'v', s=80, c='g')
#如果当天股价下跌给出买入信号，用正三角表示
plt.scatter(data['Close'].loc[data.Signal==0].index,
        data['Close'][data.Signal==0],
        marker = '^', s=80, c='r')
#将图像进行展示
plt.show()
```

运行代码，我们会得到如图 1.10 所示的结果。

深入浅出Python量化交易实战 四校 正文.indd 14　　2021/12/1 20:50:00

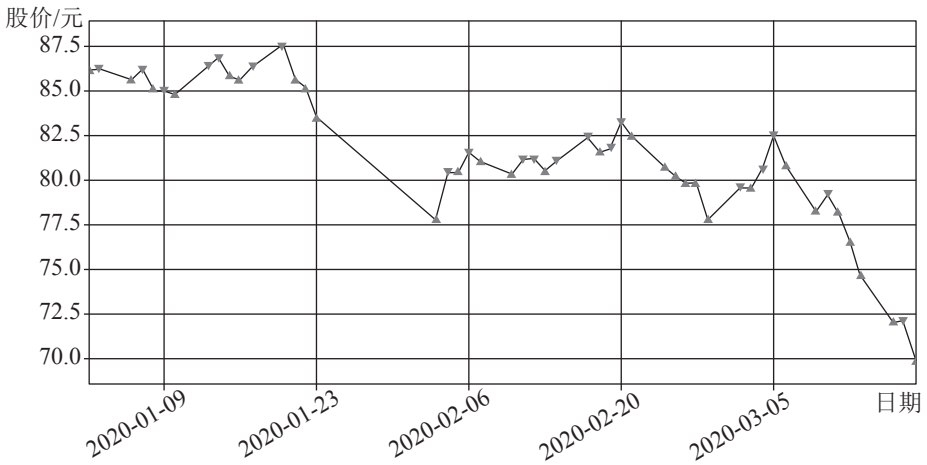

图 1.10　交易信号的可视化

【结果分析】从图 1.10 中可以看到，倒三角所处的位置是股票收盘价较上一个交易日上涨的时刻，代表卖出；正三角所处的位置是股票收盘价较上一个交易日下跌的时刻，代表买入。直观来看，在这个时间范围内，股价整体呈现出下跌天数大于上涨天数的情况。假设小瓦使用这个交易策略，且每次买入和卖出的股票数量相同的话，那么她持有这只股票的仓位会越来越高。

当然，我们不会真的让小瓦使用这个简单的策略去进行实盘交易，只是通过这些步骤帮助她熟悉一下使用 Python 进行一些简单的交易数据处理与可视化的方法。至于某种策略会带来怎样的回报，还要通过回测（backtesting）来进行评估。关于回测的具体方法，我们会在后面的章节进行介绍。

1.4　小结

作为本书的第一章，我们简要介绍了主人公小瓦姑娘的情况。面对有比较强烈的学习动机，且有一点 Python 语言基础，但对算法和金融知识所知较少的小瓦，我们首先简单介绍了交易的历史，也稍微阐述了现代交易的发展现状，目的是让小瓦先对这个领域有一定的认知，并产生兴趣；最后带着小瓦在 Anaconda 中，使用 Jupyter Notebook 做了一点简单的练习，让她稍微感受一下写代码的乐趣。当然，笔者鼓励读者朋友们跟小瓦一起进行操作，在动手的过程中掌握更多的知识。

深入浅出Python量化交易实战 四校 正文.indd 15　2021/12/1 20:50:00