大数据与人工智能技术丛书

Python 机器学习

——数据分析与评分卡建模 微课版

◎ 翟锟 胡锋 周晓然 编著

本书特色

- **零基础入门，注重实战**

  10个学习实例，3个完整的项目案例

- **视频教学，全程语音讲解**

  270分钟高品质配套教学视频

- **教学资源丰富**

  提供教学课件、源代码、数据集

270分钟 微课视频

清华大学出版社

技术丛书

Python机器学习

——数据分析与评分卡建模 微课版

◎ 翟锟 胡锋 周晓然 编著

清华大学出版社
北京

内 容 简 介

本书在 Python 数据分析与建模方面，既是一本入门书，也是一本提高书，它提炼总结了作者从 Python 小白到 Python 建模工程师的历程。如果读者有志于数据分析、建模领域，那么它一定会带给读者惊喜。书中代码具有很高的可移植性，可供读者直接使用。

全书共分为 8 章，从 Python 的环境搭建到基本语法结构，从趣味应用到分析与建模，最后以社交网络分析结束。

本书附有教学视频、源代码、课件等配套资源，适用于银行业或互联网金融行业中的风控人员，金融行业中的数据分析师（或想转行数据分析师的学习者），以及正在学习机器学习的从业人员。

图书在版编目（CIP）数据

Python 机器学习：数据分析与评分卡建模：微课版/翟锟，胡锋，周晓然编著. —北京：清华大学出版社，2019
（大数据与人工智能技术丛书）
ISBN 978-7-302-51684-2

Ⅰ. ①P…　Ⅱ. ①翟… ②胡… ③周…　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 264264 号

责任编辑：黄　芝
封面设计：刘　键
责任校对：时翠兰
责任印制：沈　露

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
课件下载：http://www.tup.com.cn，010-62795954
印 刷 者：北京富博印刷有限公司
装 订 者：北京市密云县京文制本装订厂
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张**：11.75　　**字　　数**：200 千字
版　　次：2019 年 5 月第 1 版　　**印　　次**：2019 年 5 月第1次印刷
印　　数：1～2000
定　　价：49.00 元

产品编号：080057-01

前言

Python自20世纪90年代初诞生以来，已逐渐被越来越多的开发者所接受甚至着迷。当然，这与其本身的简洁性、易读性以及可拓展性密不可分。Python具有丰富而强大的库，它能够很轻松地与其他语言的各种模块相结合。如果读者第一次见到用Python编写的语句，估计会被其整齐划一的设计风格所触动，Python的最大优点是易读、易维护。

本书以“零基础”为出发点，直接从实际应用案例入手，在内容方面更注重实战性。希望读者能在学习中不断思考，学以致用。

在本书的编写过程中，结合工作中的一些具体案例，整理成书。由于作者水平有限，疏漏在所难免，读者如发现问题，欢迎您及时指正。

特别声明，本书非系统性学习资料，内容的准备上有点跳跃，学习某些内容时还需要具备一些相关的基础知识，Google和百度会是阅读过程中的常伴。

为便于教学，本书有教学视频、源代码、课件等配套资源。

(1) 获取教学视频方式：读者可以先扫描本书封底的文泉云盘防盗码，再扫描书中相应的视频二维码，观看教学视频。

(2) 获取源代码、数据集方式：先扫描本书封底的文泉云盘防盗码，再扫描下方二维码，即可获取。

(3) 其他配套资源可以扫描封底课件二维码下载。

感谢本书的另外两名作者胡锋、周晓然，他们为此书同样牺牲了很多个人时间；感谢我的同事，他们在本书的写作过程中提供了很多的帮助；最后，感谢我的父母和妻子，把家庭生活安排得井井有条，让我能无后顾之忧，安心地编写此书！

希望这本书能够为正在学习或想要学习Python的读者提供帮助。

翟　锟

2019年1月

目录

第1章

Python开发环境搭建

Python是一种面向对象的解释型计算机程序设计语言，于1989年由荷兰人Guido van Rossum发明。1991年，第一个公开的Python版正式发行。

近年来，随着机器学习、深度学习的快速发展，Python的受欢迎程度越来越高，具体的排名顺序可以查看TIOBE官网：https://www.tiobe.com/tiobe-index/。2018年7月Python的排名是第4名，如图1-1所示。

Apr 2018	Apr 2017	Change	Programming Language	Ratings	Change
1	1		Java	15.777%	+0.21%
2	2		C	13.589%	+6.62%
3	3		C++	7.218%	+2.66%
4	5	^	Python	5.803%	+2.35%
5	4	v	C#	5.265%	+1.69%
6	7	^	Visual Basic .NET	4.947%	+1.70%
7	6	v	PHP	4.218%	+0.84%

图1-1　TIOBE 2018年7月编程语言排行榜

2018年3月，Python作者在邮件列表上宣布将于2020年1月1日终止支持Python 2.7版本。同时，由于越来越多的第三方库都不再维护Python 2版本，Python 3是大势所趋。本书主要讲述Python 3的用法，有关Python 2的用法部分，

会做特殊说明。

工欲善其事,必先利其器。在实际开发 Python 代码的过程中,我们常用的“利器”有哪些呢?

本章只是简单地对 Windows 系统上相关 Python 开发软件的安装过程进行讲解,而 Mac OS 系统上的安装过程与之类似。Linux 系统上的安装,可以使用 Google 或者百度来查找相关的安装教程。在 Linux 系统上安装相关软件时,可能会遇到提示的错误信息情况,可在网上搜出相应的解决办法。

1.1 利器 1: Notepad 编辑器

Notepad(记事本)是代码编辑器或 Windows 中的小程序,用于文本编辑,是一款开源、小巧、免费的纯文本编辑器,具有运行便携、体积小、资源占用小,以及支持众多程序语言等优点。Notepad 支持的语言有 C++、C#、Java 等主流程序语言,HTML、XML、Python、JavaScript 等网页/脚本语言。Notepad 内置支持多达 27 种语法高亮度显示,还可实现语法折叠、宏等常用功能。Notepad 的官网链接地址为: https://notepad-plus-plus.org/,对应的界面如图 1-2 所示。

图 1-2 Notepad 官网首页界面图

安装好 Notepad 之后,其操作界面如图 1-3 所示。

Notepad 的优点主要包括以下 5 个方面。

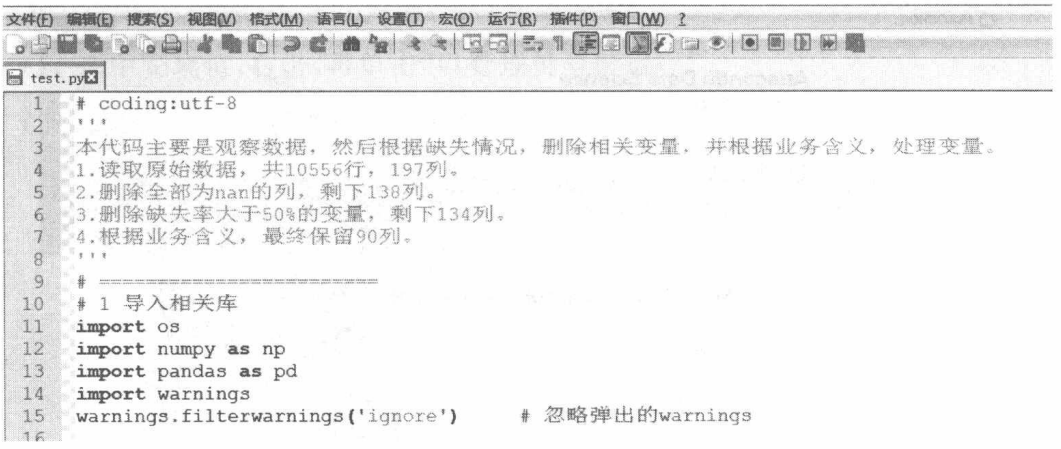

图 1-3 Notepad 界面图

(1) 语法高亮，具有字词自动完成功能，支持同时编辑多重文档；支持自定义语言。

(2) 自动检测文件类型，根据关键字显示节点，节点可自由折叠或打开，还可显示缩进引导线，使代码富有层次感。

(3) 在分窗口中又可打开多个子窗口，可以使用 F11 键快捷切换至全屏显示模式，支持鼠标滚轮改变文档显示比例。

(4) 可显示选中文本的字节数，并非普通编辑器所显示的字数；提供了一些实用工具，如邻行互换位置、宏功能等。

(5) 能够进行列块编辑，可快速地插入或者删除文本，能够成批替换文本。

1.2 利器 2：Anaconda

Anaconda 可以简单地理解为一个工具箱，其中包含了 conda、flask、nltk、pandas、pip 等 180 多个科学包及其依赖项，可以方便地实现包的安装、更新和卸载，比如机器学习包 scikit-learn、词云包 wordcloud 等。Anaconda 最大的好处在于它集成了 Jupyter Notebook 和 Spyder，这两个工具可以快速地让我们看到代码的运行结果，以便进行调试。

Anaconda 官网链接地址为：https://www.anaconda.com，官网首页如图 1-4 所示。Anaconda 下载链接地址为：https://www.anaconda.com/download/和 https://repo.continuum.io/archive/，如图 1-5 和图 1-6 所示。

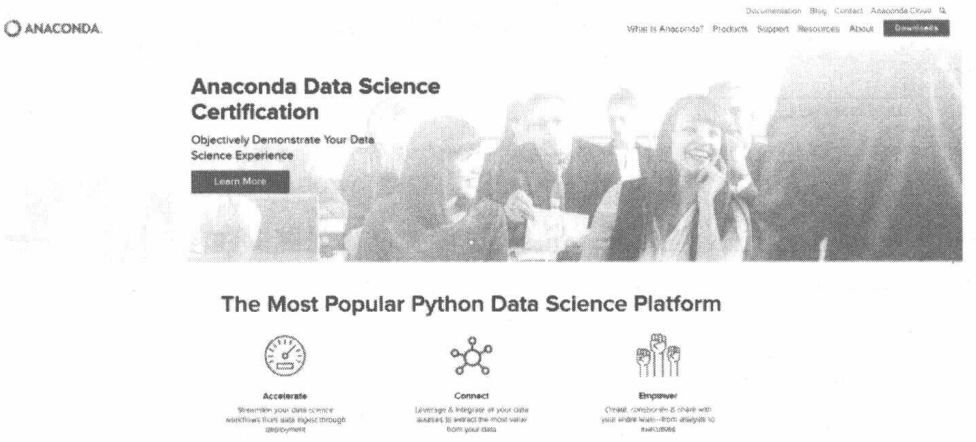

图 1-4　Anaconda 官网首页界面图

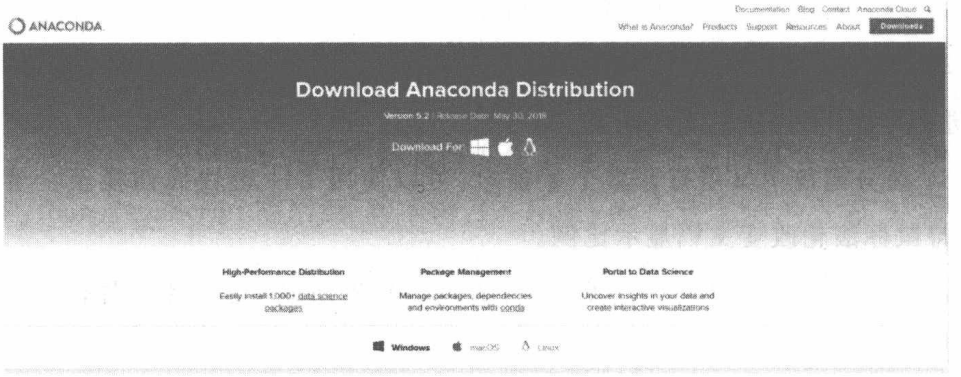

图 1-5　Anaconda 下载界面图(1)

Anaconda installer archive

Filename	Size	Last Modified	MD5
Anaconda2-5.2.0-Linux-ppc64le.sh	269.6M	2018-05-30 13:04:31	479633a95906ea6d41056ebe84a4c47b
Anaconda2-5.2.0-Linux-x86.sh	488.7M	2018-05-30 13:05:30	758e172a824f467ea6b55d3d076c132f
Anaconda2-5.2.0-Linux-x86_64.sh	603.4M	2018-05-30 13:04:33	5c034a4ab36ec9b6ae01fa13d8a04462
Anaconda2-5.2.0-MacOSX-x86_64.pkg	616.8M	2018-05-30 13:05:32	2836c839d29be8d9569a715f4c631a3b
Anaconda2-5.2.0-MacOSX-x86_64.sh	527.1M	2018-05-30 13:05:34	b1f3fcf58955830b65613a4a8d75c3cf
Anaconda2-5.2.0-Windows-x86.exe	443.4M	2018-05-30 13:04:17	4a3729b14c2d3fccd3a050821679c702
Anaconda2-5.2.0-Windows-x86_64.exe	564.0M	2018-05-30 13:04:16	595e427e4b625b6eab92623a28dc4e21
Anaconda3-5.2.0-Linux-ppc64le.sh	288.3M	2018-05-30 13:05:40	cbd1d5435ead2b0b97dba5b3cf45d694
Anaconda3-5.2.0-Linux-x86.sh	507.3M	2018-05-30 13:05:46	81d5a1648e3aca4843f88ca3769c0830
Anaconda3-5.2.0-Linux-x86_64.sh	621.6M	2018-05-30 13:05:43	3e58f494ab9fbe12db4460dc152377b5
Anaconda3-5.2.0-MacOSX-x86_64.pkg	613.1M	2018-05-30 13:07:00	9c35bf27e9986701f7d80241616c665f
Anaconda3-5.2.0-MacOSX-x86_64.sh	523.3M	2018-05-30 13:07:03	b5b789c01e1992de55ee911754c310d4
Anaconda3-5.2.0-Windows-x86.exe	506.3M	2018-05-30 13:04:19	285387e7b6ea81edba98c011922e235a
Anaconda3-5.2.0-Windows-x86_64.exe	631.3M	2018-05-30 13:04:18	62244c0382b8142743622fdc3526eda7
Anaconda2-5.1.0-Linux-ppc64le.sh	267.3M	2018-02-15 09:08:49	e894dcc547a1c7d67deb04f6bba7223a
Anaconda2-5.1.0-Linux-x86.sh	431.3M	2018-02-15 09:08:51	e26fb9d3e53049f6e32212270af6b987
Anaconda2-5.1.0-Linux-x86_64.sh	533.0M	2018-02-15 09:08:50	5b1b5784cae93cf696e11e66983d8756
Anaconda2-5.1.0-MacOSX-x86_64.pkg	588.0M	2018-02-15 09:08:52	4f9c197dfe6d3dc7e50a8611b4d3cfa2
Anaconda2-5.1.0-MacOSX-x86_64.sh	505.9M	2018-02-15 09:08:53	e9845ccf67542523c5be09552311666e
Anaconda2-5.1.0-Windows-x86.exe	419.8M	2018-02-15 09:08:55	a09347a53e04a15ee965300c2b95dfde

图 1-6　Anaconda 下载界面图(2)

本节以 Anaconda3 5.0.0 为例，进行安装说明。

(1) 选中安装包，右击，再单击“以管理员身份运行”。

(2) 单击 Next 按钮，如图 1-7 所示。

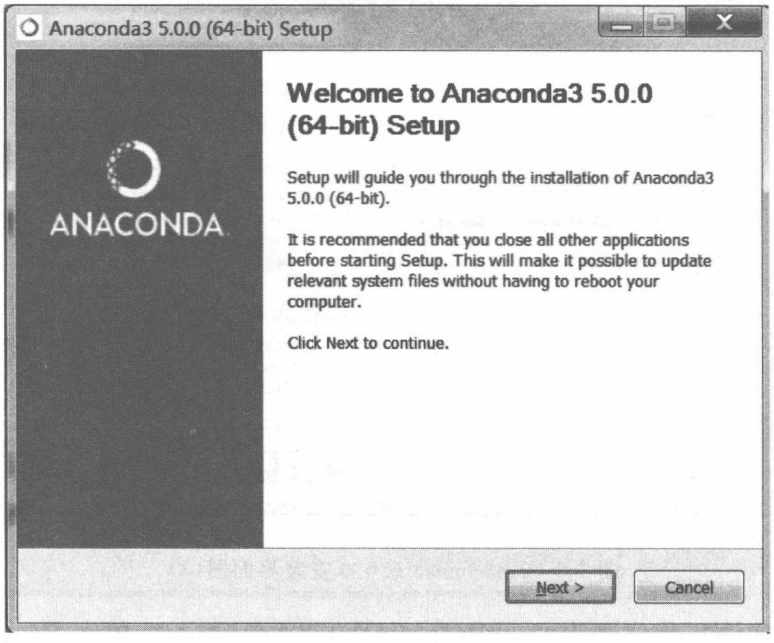

图 1-7 Anaconda3 5.0.0 安装界面图(1)

(3) 单击 I Agree 按钮，如图 1-8 所示。

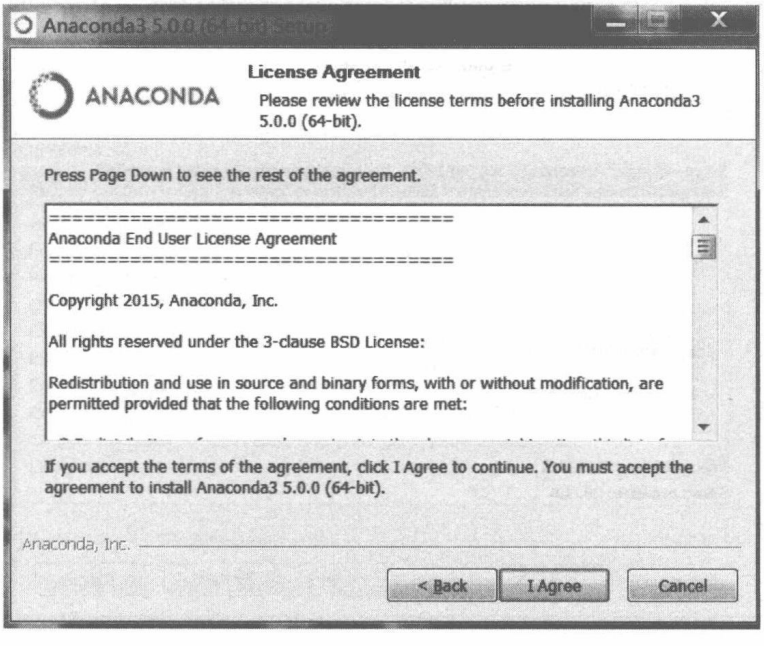

图 1-8 Anaconda3 5.0.0 安装界面图(2)

(4) 勾选 All Users(requires admin privileges),单击 Next 按钮,如图 1-9 所示。

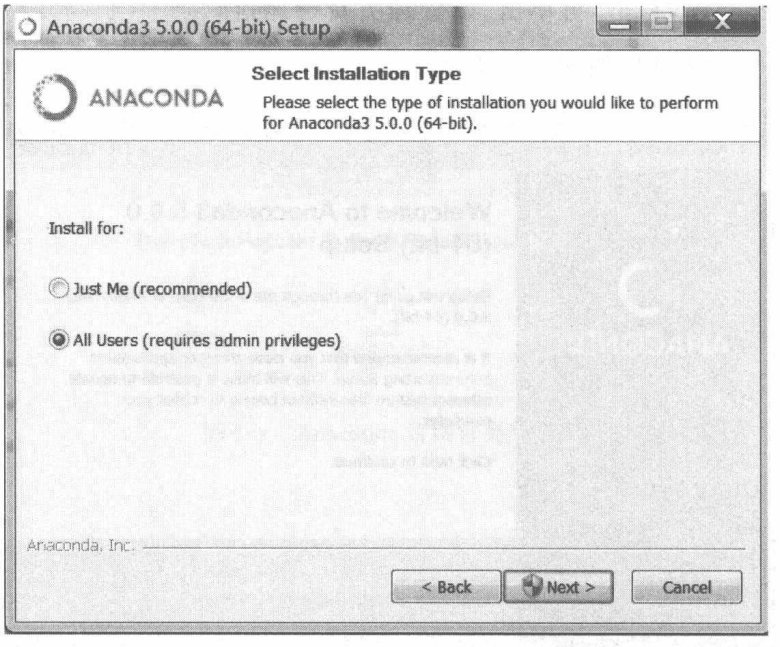

图 1-9 Anaconda3 5.0.0 安装界面图(3)

(5) 程序默认安装位置为 C:\ProgramData\Anaconda3,单击 Browse 可选择自定义安装目录,本安装教程的自定义安装路径为 D:\DevTools\Anaconda3,单击 Next 按钮,如图 1-10 所示。

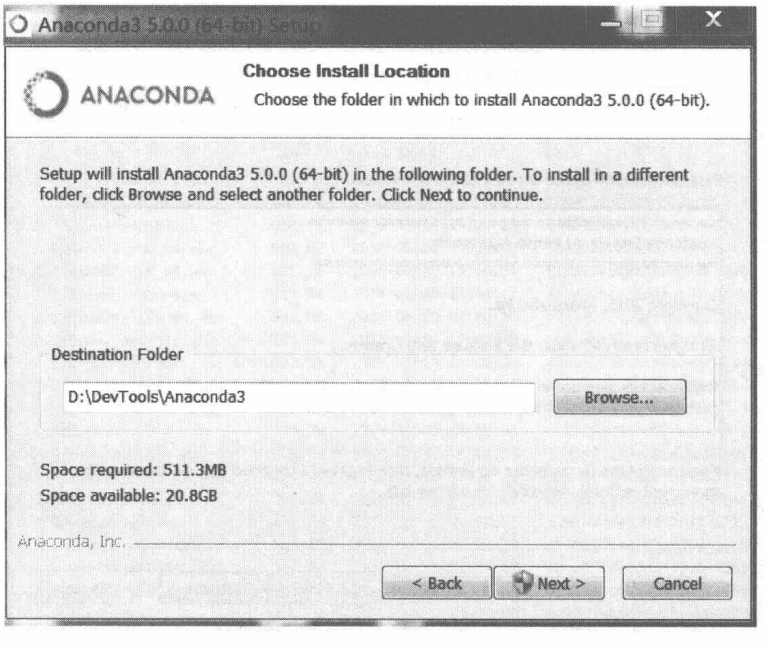

图 1-10 Anaconda3 5.0.0 安装界面图(4)

(6) 同时勾选 Add Anaconda to the system PATH environment variable 和 Register Anaconda as the system Python 3.6,单击 Install 按钮,如图 1-11 所示。

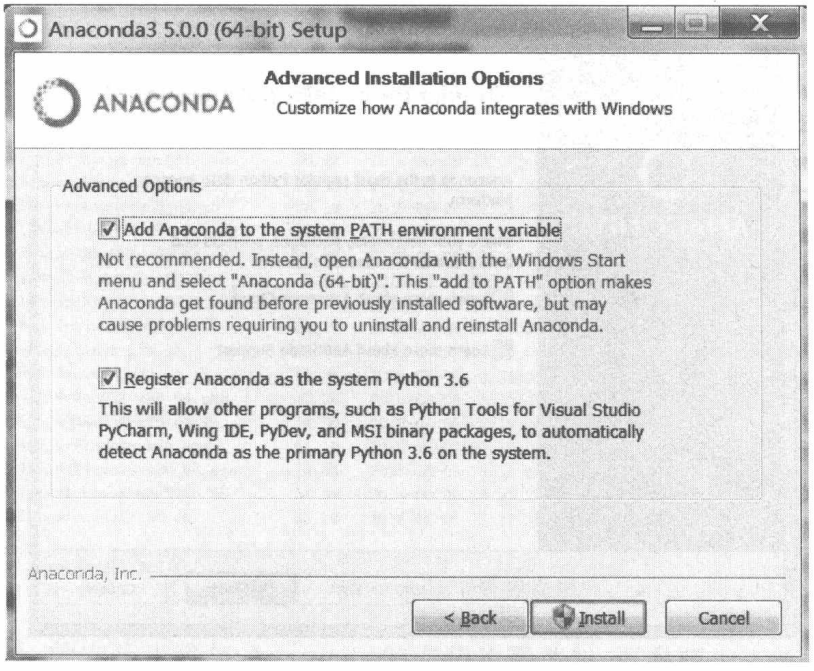

图 1-11 Anaconda3 5.0.0 安装界面图(5)

(7) 安装过程需要 4～5 分钟,单击 Next 按钮,如图 1-12 所示。

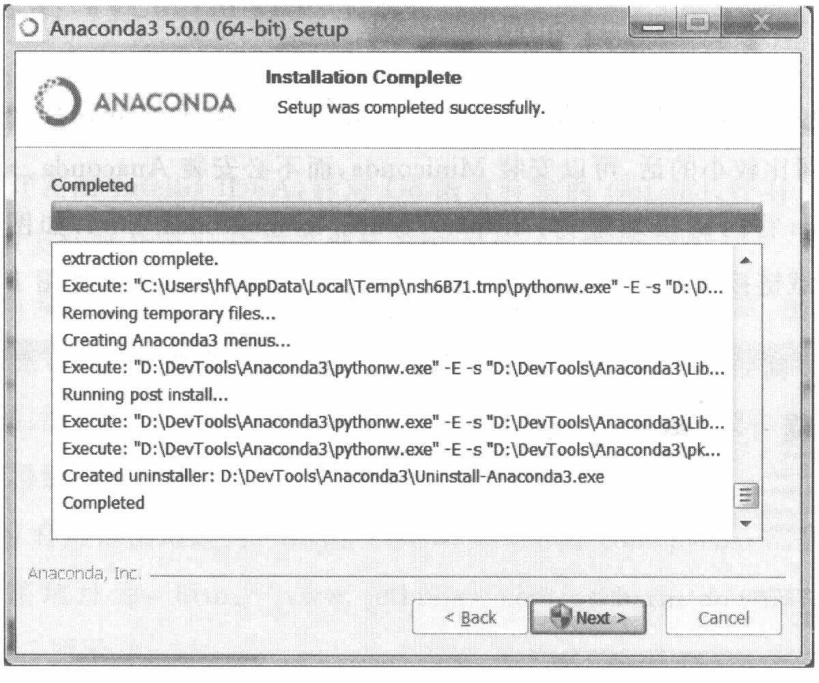

图 1-12 Anaconda3 5.0.0 安装界面图(6)

(8) 取消勾选 Learn more about Anaconda Cloud 和 Learn more about Anaconda Support,然后单击 Finish 按钮,至此完成了 Anaconda 的安装,如图 1-13 所示。

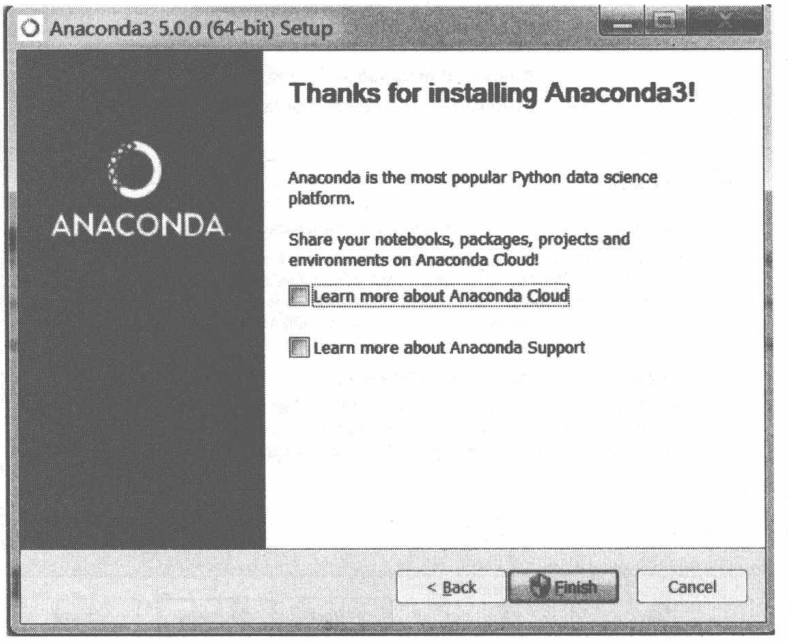

图 1-13 Anaconda3 5.0.0 安装界面图(7)

1.3 利器 3: Miniconda

Miniconda 相当于迷你版的 Anaconda,其功能比 Anaconda 稍微少一些。如果计算机存储空间比较小的话,可以安装 Miniconda,而不必安装 Anaconda。

Miniconda 官网链接地址为: https://conda.io/miniconda.html,如图 1-14 所示。Miniconda 下载链接地址为: https://repo.continuum.io/miniconda/,如图 1-15 所示。

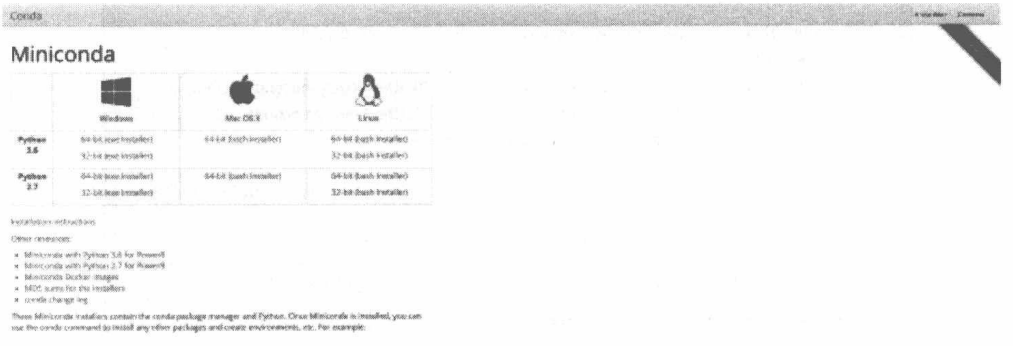

图 1-14 Miniconda 官网首页界面图

Miniconda installer archive

Filename	Size	Last Modified	MD5
Miniconda2-4.5.4-Linux-ppc64le.sh	36.9M	2018-06-06 23:07:18	3e26ee6447c8025609eale410f768417
Miniconda2-4.5.4-Linux-x86.sh	35.5M	2018-06-06 22:27:33	a638ae058a0ce15c5b289d151c488045
Miniconda2-4.5.4-Linux-x86_64.sh	38.1M	2018-06-06 22:24:38	8a1c02f6941d8778f8afad7328265cf5
Miniconda2-4.5.4-MacOSX-x86_64.pkg	34.5M	2018-06-06 23:12:27	6040ee82686b36f9bffa32d5f2a1341b
Miniconda2-4.5.4-MacOSX-x86_64.sh	29.8M	2018-06-06 23:12:26	35f4ca99d33ed56f68745eeaf1449274
Miniconda2-4.5.4-Windows-x86.exe	51.8M	2018-06-07 00:09:59	55502ce28ba3a16aa12954522d3624d2
Miniconda2-4.5.4-Windows-x86_64.exe	55.9M	2018-06-06 23:52:04	d285b7451deb4a33037033307c153684
Miniconda2-latest-Linux-ppc64le.sh	36.9M	2018-06-06 23:07:18	3e26ee6447c8025609eale410f768417
Miniconda2-latest-Linux-x86.sh	35.5M	2018-06-06 22:27:33	a638ae058a0ce15c5b289d151c488045
Miniconda2-latest-Linux-x86_64.sh	38.1M	2018-06-06 22:24:38	8a1c02f6941d8778f8afad7328265cf5
Miniconda2-latest-MacOSX-x86_64.pkg	34.5M	2018-06-06 23:12:27	6040ee82686b36f9bffa32d5f2a1341b
Miniconda2-latest-MacOSX-x86_64.sh	29.8M	2018-06-06 23:12:26	35f4ca99d33ed56f68745eeaf1449274
Miniconda2-latest-Windows-x86.exe	51.8M	2018-06-07 00:09:59	55502ce28ba3a16aa12954522d3624d2
Miniconda2-latest-Windows-x86_64.exe	55.9M	2018-06-06 23:52:04	d285b7451deb4a33037033307c153684
Miniconda3-4.5.4-Linux-ppc64le.sh	54.9M	2018-06-06 23:07:24	05c1e073f262105179cf57920dfc4d43
Miniconda3-4.5.4-Linux-x86.sh	53.7M	2018-06-06 22:27:35	0fcc79d640d82b7d36ea39654a82dd9d
Miniconda3-4.5.4-Linux-x86_64.sh	55.8M	2018-06-06 22:24:39	a946ea1d0c4a642ddf0c3a26a18bb16d
Miniconda3-4.5.4-MacOSX-x86_64.pkg	40.2M	2018-06-06 23:12:28	242882072fda2ada8551239e9041f5e9
Miniconda3-4.5.4-MacOSX-x86_64.sh	34.9M	2018-06-06 23:12:26	164ec263c4070db642ce31bb45d68813
Miniconda3-4.5.4-Windows-x86.exe	51.1M	2018-06-07 00:10:06	bc2f687a9e92455a099242929df8471d
Miniconda3-4.5.4-Windows-x86_64.exe	54.8M	2018-06-06 23:52:12	1c73051ccd997770288275ee6474b423
Miniconda3-latest-Linux-ppc64le.sh	54.9M	2018-06-06 23:07:24	05c1e073f262105179cf57920dfc4d43
Miniconda3-latest-Linux-x86.sh	53.7M	2018-06-06 22:27:35	0fcc79d640d82b7d36ea39654a82dd9d

图 1-15 Miniconda 下载界面图

Miniconda 的安装步骤与 Anaconda 的安装步骤类似，可参照 1.2 节中 Anaconda 的安装步骤，这里不再赘述。

1.4 利器 4：PyCharm IDE 工具

学过 Java、Go、Web 开发的，可能都知道 JetBrains 公司，该公司的主要产品有针对 Java 语言开发的 IntelliJ IDEA，针对 Go 语言开发的 GoLand，针对 Web 开发的 WebStorm，以及针对 Python 语言开发的 IDE 工具——PyCharm。

PyCharm 是一款 Python IDE，可以显著提高 Python 的开发效率，比如可以提高调试、语法高亮、Project 管理、代码跳转、智能提示、单元测试、版本控制等开发速率。在编写代码的过程中，PyCharm 可快速实现错误高亮，智能检测以及一键式代码快速补全建议，使得编码更加优化。

PyCharm 官网链接地址为：http://www.jetbrains.com/pycharm，如图 1-16 所示。下载链接地址为：http://www.jetbrains.com/pycharm/download/previous.html，如图 1-17 所示。

其具体的安装步骤，详见附录 A。

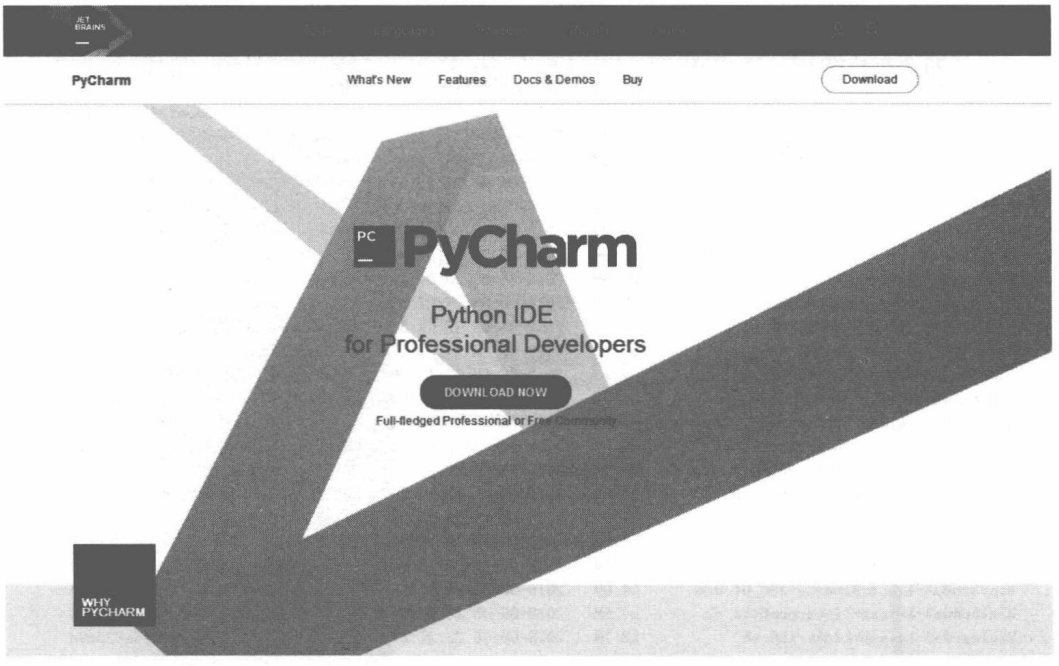

图 1-16　PyCharm 官网首页界面图

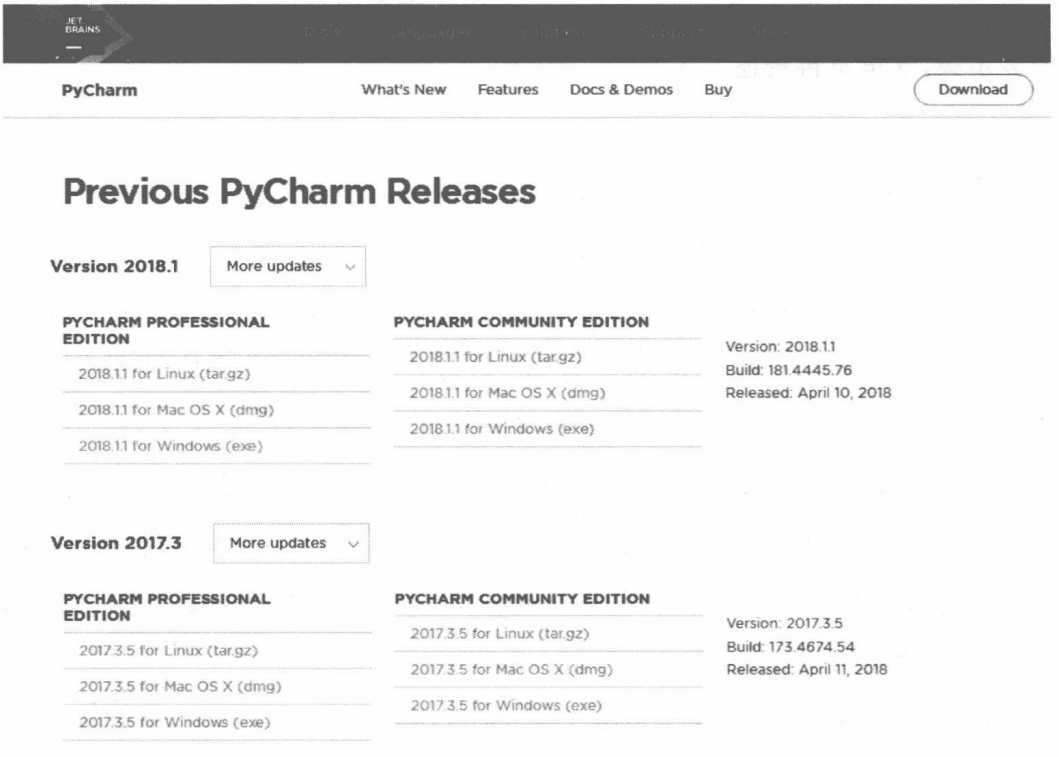

图 1-17　PyCharm 下载界面图