

Python Data Science Handbook

Python数据科学手册

Jake VanderPlas

原作者 [Jake VanderPlas](#)

译者 [wangyingsm@github.com](#)

This is the Jupyter notebook version of the [Python Data Science Handbook](#) by Jake VanderPlas; the content is available [on GitHub](#).^{*} The text is released under the [CC-BY-NC-ND license](#), and code is released under the [MIT license](#). If you find this content useful, please consider supporting the work by [buying the book!](#)

这是Jake VenderPlas所著的[Python数据科学手册](#)的Jupyter notebook版本；本内容在[GitHub](#)上。文字发行协议遵循[CC-BY-NC-ND license](#)协议，而代码发行遵循[MIT license](#)。如果你认为这些内容很有用，请考虑通过[购买本书](#)支持作者。

目录

译者注：原版英文目录见下方

[序言](#)

[1. IPython：超越Python解释器](#)

- [IPython帮助和文档](#)
- [IPython Shell中的键盘快捷键](#)
- [IPython魔术命令](#)
- [输入输出历史](#)
- [IPython和Shell命令](#)
- [错误和调试](#)
- [性能测算和计时](#)
- [更多IPython资源](#)

[2. Numpy介绍](#)

- [理解Python中的数据类型](#)
- [Numpy数组基础](#)
- [使用Numpy计算：通用函数](#)
- [聚合：Min, Max, 以及其他](#)
- [在数组上计算：广播](#)
- [比较，遮盖和布尔逻辑](#)
- [高级索引](#)
- [数组排序](#)
- [格式化数据：NumPy里的结构化数组](#)

[3. 使用Pandas进行数据处理](#)

- [Pandas对象简介](#)
- [数据索引和选择](#)
- [在Pandas中操作数据](#)
- [处理空缺数据](#)
- [层次化的索引](#)
- [组合数据集：Concat 和 Append](#)
- [组合数据集：Merge 和 Join](#)
- [聚合与分组](#)
- [数据透视表](#)
- [向量化的字符串操作](#)
- [在时间序列上操作](#)
- [高性能Pandas: `eval\(\)` 和 `query\(\)`](#)
- [更多资源](#)

[4. 使用matplotlib展示数据](#)

- [简单的折线图](#)
- [简单的散点图](#)
- [误差可视化](#)
- [密度和轮廓图](#)
- [直方图, 分桶和密度](#)
- [自定义图表图例](#)
- [自定义颜色条](#)
- [多个子图表](#)
- [文本和标注](#)
- [自定义刻度](#)
- [自定义matplotlib：配置和样式单](#)
- [在matplotlib中创建三维图表](#)
- [使用Basemap创建地理位置图表](#)
- [使用Seaborn进行可视化](#)
- [更多资源](#)

[5. 机器学习](#)

- [什么是机器学习？](#)
- [Scikit-Learn简介](#)
- [超参数及模型验证](#)
- [特征工程](#)
- [深入：朴素贝叶斯分类](#)
- [深入：线性回归](#)
- [深入：支持向量机](#)
- [深入：决策树和随机森林](#)
- [深入：主成分分析](#)
- [深入：流形学习](#)
- [深入：k-均值聚类](#)
- [深入：高斯混合模型](#)
- [深入：核密度估计](#)
- [应用：脸部识别管道](#)
- [更多机器学习资源](#)

[附录：生成图像的代码](#)

Table of Contents

[Preface](#)

[1. IPython: Beyond Normal Python](#)

- [Help and Documentation in IPython](#)
- [Keyboard Shortcuts in the IPython Shell](#)
- [IPython Magic Commands](#)
- [Input and Output History](#)
- [IPython and Shell Commands](#)
- [Errors and Debugging](#)
- [Profiling and Timing Code](#)
- [More IPython Resources](#)

[2. Introduction to NumPy](#)

- [Understanding Data Types in Python](#)
- [The Basics of NumPy Arrays](#)
- [Computation on NumPy Arrays: Universal Functions](#)
- [Aggregations: Min, Max, and Everything In Between](#)
- [Computation on Arrays: Broadcasting](#)
- [Comparisons, Masks, and Boolean Logic](#)
- [Fancy Indexing](#)
- [Sorting Arrays](#)
- [Structured Data: NumPy's Structured Arrays](#)

[3. Data Manipulation with Pandas](#)

- [Introducing Pandas Objects](#)
- [Data Indexing and Selection](#)
- [Operating on Data in Pandas](#)
- [Handling Missing Data](#)
- [Hierarchical Indexing](#)
- [Combining Datasets: Concat and Append](#)
- [Combining Datasets: Merge and Join](#)
- [Aggregation and Grouping](#)
- [Pivot Tables](#)
- [Vectorized String Operations](#)
- [Working with Time Series](#)
- [High-Performance Pandas: `eval\(\)` and `query\(\)`](#)
- [Further Resources](#)

[4. Visualization with Matplotlib](#)

- [Simple Line Plots](#)
- [Simple Scatter Plots](#)
- [Visualizing Errors](#)
- [Density and Contour Plots](#)
- [Histograms, Binnings, and Density](#)
- [Customizing Plot Legends](#)
- [Customizing Colorbars](#)
- [Multiple Subplots](#)
- [Text and Annotation](#)
- [Customizing Ticks](#)
- [Customizing Matplotlib: Configurations and Stylesheets](#)
- [Three-Dimensional Plotting in Matplotlib](#)
- [Geographic Data with Basemap](#)
- [Visualization with Seaborn](#)
- [Further Resources](#)

[5. Machine Learning](#)

- [What Is Machine Learning?](#)
- [Introducing Scikit-Learn](#)
- [Hyperparameters and Model Validation](#)
- [Feature Engineering](#)
- [In Depth: Naive Bayes Classification](#)
- [In Depth: Linear Regression](#)
- [In-Depth: Support Vector Machines](#)
- [In-Depth: Decision Trees and Random Forests](#)
- [In Depth: Principal Component Analysis](#)
- [In-Depth: Manifold Learning](#)
- [In Depth: k-Means Clustering](#)
- [In Depth: Gaussian Mixture Models](#)
- [In-Depth: Kernel Density Estimation](#)
- [Application: A Face Detection Pipeline](#)
- [Further Machine Learning Resources](#)

[Appendix: Figure Code](#)