

熟悉Python数据分析的工具

了解Python数据分析的优势

- **语法简单精练**。对于初学者来说，比其他编程语言更容易上手。
- **有强大的库**。可以只使用Python这一种语言去构建以数据为中心的应用程序。
- **功能强大**。Python是一个混合体，丰富的工具集使它介于传统的脚本语言和系统语言之间。Python不仅具备所有脚本语言简单和易用的特点，还提供了编译语言所具有的高级软件工程工具。
- **不仅适用于研究和原型构建，同时也适用于构建生产系统**。研究人员和工程技术人员使用同一种编程工具，给企业带来非常显著的组织效益，并降低企业的运营成本。
- **Python是一门胶水语言**。Python程序能够以多种方式轻易的与其他语言的组件“粘接”在一起。

Python数据分析常用类库

1. Ipython – 科学计算标准工具集的组成部分

- 是一个增强的Python shell，目的是提高编写、测试、调试Python代码的速度。
- 主要用于交互式数据并行处理，是分布式计算的基础架构。
- 提供了一个类似于Mathematica的HTML笔记本，一个计与Qt构架的GUI控制台，具有绘图、多行编辑以及语法高亮显示等功能。

Python数据分析常用类库

NumPy (Numerical Python) -- Python科学计算的基础包

- 快速高效的多维数组对象 ndarray。
- 对数组执行元素级的计算以及直接对数组执行数学运算的函数。
- 读写硬盘上基于数组的数据集工具。
- 线性代数运算、傅里叶变换以及随机数生成的功能。
- 将C、C++、Fortran代码集成到Python的工具。

Python数据分析常用类库

SciPy – 专门解决科学计算中各种标准问题域的模块的集合

- SciPy主要包含了8个模块，不同的子模块有不同的应用，如插值、积分、优化、图像处理和特殊函数等。
- `scipy.integrate` – 数值积分例程和微分方程求解器
- `scipy.linalg` -- 扩展了由`numpy.linalg` 提供的线性代数例程和矩阵分解功能
- `scipy.optimize` – 函数优化器（最小化器）以及根查找算法
- `scipy.signal` – 信号处理工具
- `scipy.sparse` – 稀疏矩阵和稀疏线性系统求解器
- `scipy.special` – `specfun`（这是一个实现了许多常用数学函数的Fortran）
- `scipy.stats` – 检验连续和离散概率分布、各种统计检验方法以及更好的
- `scipy.weave` – 利用内联 C++ 代码加速数组计算的工具

Python数据分析常用类库

Pandas -- 数据分析核心库

- 提供了一系列能够快速、便捷的处理结构化数据的数据结构和函数。
- 高性能的数组计算功能以及电子表格和关系型数据库（如 SQL）灵活的数据处理功能。
- 复杂精细的索引功能、以便便捷的完成重塑、切片和切块、聚合及选取数据子集等操作。

Python数据分析常用类库

Matplotlib – 绘制数据图表的Python库

- Python的2D绘图库，非常适合创建出版物上用的图表。
- 操作比较容易，只需几行代码即可生成直方图、功率谱图、条形图、错误图和散点图等图形。
- 提供了pylab的模块，其中包括了NumPy和pyplot中许多常用的函数，方便用户快速计算和绘图。
- 交互式的数据绘图环境，绘制的图表也是交互式的。

Python数据分析常用类库

scikit-learn – 数据挖掘和数据分析工具

- 简单有效，可供用户在各种环境下重复使用。
- 封装了一些常用的算法方法。
- 基本模块主要有数据预处理、模型选择、分类、聚类、数据降维和回归6个，在数据量不大的情况下，scikit-learn可以解决大部分问题。

Python数据分析常用类库

Spyder – 交互式Python语言开发环境

- 提供高级的代码编辑、交互测试和调试等特性。
- 包含数值计算环境。
- 可用于将调试控制台直接集成到图形用户界面的布局中。
- 模仿MATLAB“工作空间”，可以很方便的观察和修改数组的值。

了解Python的Anaconda发行版

- 预装了大量常用的Packages。
- 完全开源和免费。
- 额外的加速和优化是收费的，但对于学术用途，可以申请免费的License。
- 对全平台和几乎所有Python版本支持。

掌握Jupyter Notebook的基本功能

- 打开并新建一个Notebook
- 打开Jupyter Notebook
- “Text File” 为纯文本型
- “Folder”为文件夹
- “Python 3” 表示Python运行脚本

掌握Jupyter Notebook的基本功能

Notebook文档由一系列单元构成，主要有两种形式的单元

- **代码单元** – 这里是读者编写代码的地方。
- **Markdown单元** – 在这里对文本进行编辑。
 - Markdown是一种可以使用普通文本编辑器编写的标记语言，通过简单的标记语法，它可以使普通文本内容具有一定的格式。

Markdown--标题

- 标题是标明文章和作品等内容的简短语句，一个“#”字符代表一级标题，以此类推。

一级标题

二级标题

三级标题

四级标题

五级标题

六级标题

一级标题

二级标题

三级标题

四级标题

五级标题

六级标题

Markdown –列表

列表是一种由数据项构成的有限序列，即按照一定的线性顺序排列而成的数据项的集合。

- **无序列表**：使用星号、加号或者剑豪作为列表标记
- **有序列表**：使用数字“， ” “（一个空格）”。

Markdown – 加粗/斜体

- 前后有**两个星号**或者**下划线**表示加粗,
- 前后由**三个星号**或者**下划线**表示斜体。

Markdown – 数学公式编辑

- LaTeX是写科研论文的必备工具，Markdown单元中也可以使用LaTeX来插入数学公式。
- 在文本行中插入数学公式，应在公式前后分别加上一个“\$”符号。
- 如果插入一个数学区块，则在公式前后分别加入两个“\$\$”符号。