

# Python 数据分析

---

Cloudera 大数据培训基地

重庆翰海睿智大数据科技有限公司

# 数值计算类库 SciPy

---

## 数值计算类库 SciPy

在本章中，您将了解到：

1. SciPy 库简介
2. 最小二乘拟合算法
3. 函数最小值算法
4. 非线性方程组算法
5. B-Spline 样条曲线插值算法
6. 数值积分算法
7. 常微分方程组算法

数值计算类库 SciPy

SciPy 数值计算类库

结论

## Scipy 库简介 (1)

SciPy 函数库在 NumPy 库的基础上增加了众多的数学、科学以及工程计算中常用的库函数。例如线性代数、常微分方程数值求解、信号处理、图像处理、稀疏矩阵等等。

模块	功能
scipy.cluster	矢量量化 / K-均值
scipy.constants	物理和数学常数
scipy.fftpack	傅里叶变换
scipy.integrate	积分程序
scipy.interpolate	插值

## Scipy 库简介 (2)

模块	功能
scipy.io	数据输入输出
scipy.linalg	线性代数程序
scipy.ndimage	n 维图像包
scipy.odr	正交距离回归
scipy.optimize	优化和拟合
scipy.signal	信号处理
scipy.sparse	稀疏矩阵
scipy.spatial	空间数据结构和算法
scipy.special	任何特殊数学函数
scipy.stats	统计

## 最小二乘拟合算法

如有一组噪音实验数据  $(x[i], y[i])$ ，确定它们之间的函数关系： $y = f(x)$ ，通过这些已知信息，需要确定函数中的一些参数项。例如，如果  $f$  是一个线型函数  $f(x) = k \cdot x + b$ ，参数  $k$  和  $b$  是需要确定的值。如果参数用  $p$  表示的话，那么我们就是要找到一组  $p$  值，使如下公式中的  $S$  函数最小。

$$S(p) = \sum_{i=1}^m [y_i - f(x_i, p)]^2$$

图 1: ‘最小二乘拟合’

这种算法被称之为最小二乘拟合。

scipy 中的子函数库 optimize 已经提供了实现最小二乘拟合算法的函数 `leastsq`



Scipy 的 stats 模块包含了多种概率分布的随机变量，随机变量分为连续的和离散的两种。

### 示例 2 统计函数分析随机数

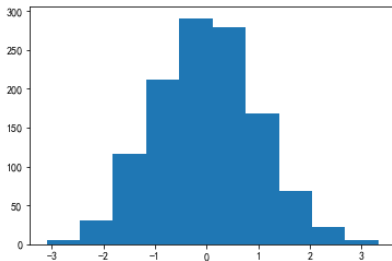


图 2: ‘分析随机数视图’

linalg 库中的 solve 函数可以用来对线性代数方程组进行求解。

基本调用形式: `linalg.solve(a, b)`

## 示例 3 求解线性方程组

线性方程组  $Ax = b$  即  $x = A^{-1}b$

- A:  $m \times n$  的方形矩阵
- x: 长为  $m$  的向量
- b: 长为  $m$  的向量

求解线性方程组  $AX = B$  即  $X = A^{-1}B$

- X:  $m \times n$  的矩阵
- B:  $m \times n$  的矩阵

interpolate 库提供了许多对数据进行插值运算的函数。

### 示例 4 直线和 B-Spline 对正弦波上的点进行插值

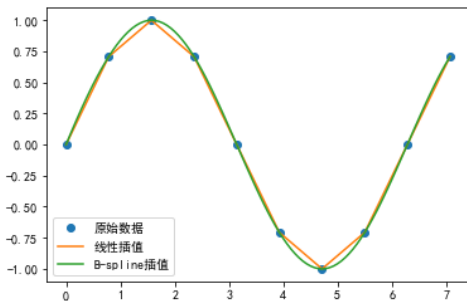


图 3: 'B-Spline 样条曲线'

# 数值积分算法

integrate 库提供了数值积分算法的函数。如 cumtrapz 累计计算积分。

## 示例 5 累计计算积分

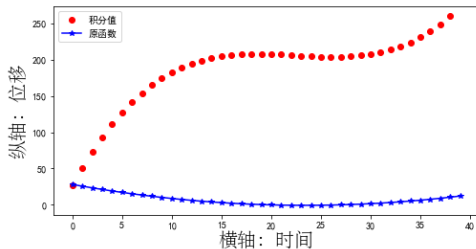


图 4: ‘累计计算积分’

integrate 库除了提供数值积分算法的函数，还有常微分方程组求解算法的函数 odeint。

### 示例 6 计算洛伦兹吸引子的轨迹

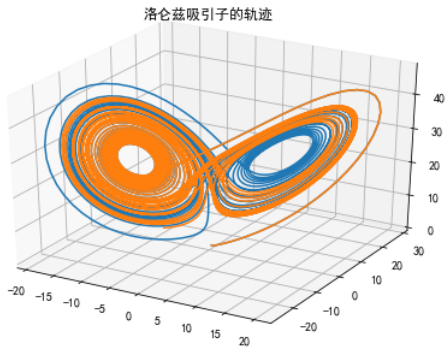


图 1 洛伦兹吸引子的轨迹

数值计算类库 SciPy

SciPy 数值计算类库

结论

## 基本要点

---

- 最小二乘拟合算法
- 函数最小值算法
- 非线性方程组算法
- B-Spline 样条曲线插值算法
- 数值积分算法
- 常微分方程组算法