

人工智能程序设计

M2 科学计算与数据分析基础

SciPy生态系统简介

张 莉





人工智能程序设计

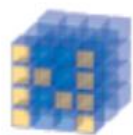
软件生态系统SCIPY

南京大学本科《人工智能程序设计》课程专用

SciPy

特征

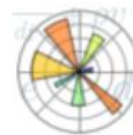
- 基于Python的软件生态系统（ecosystem）
- 开源
- 主要为数学、科学和工程服务



NumPy
Base N-dimensional
array package



SciPy library
Fundamental library for
scientific computing



Matplotlib
Comprehensive 2-D
plotting



IPython
Enhanced interactive
console



SymPy
Symbolic mathematics

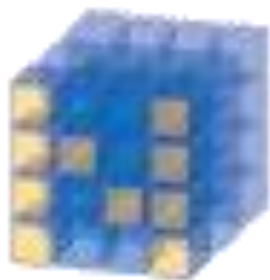


pandas
Data structures &
analysis

<https://scipy.org/>

NumPy

特征



- 高性能科学计算和数据分析的基础包
- 强大的ndarray对象
- 精巧的函数和ufunc函数
- 适合线性代数和随机数处理等科学计算

Source


```
>>> import numpy as np
>>> np.ones((4, 4))
>>> np.arange(1, 11)*3
```

SciPy library

特征

- 基于NumPy，是科学计算核心库
- 有效计算numpy矩阵，让NumPy和SciPy library协同工作
- 致力于科学计算中常见问题的各个工具箱，其不同子模块有不同的应用，如插值、积分、优化和图像处理等

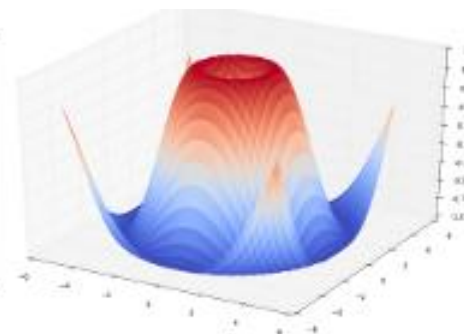
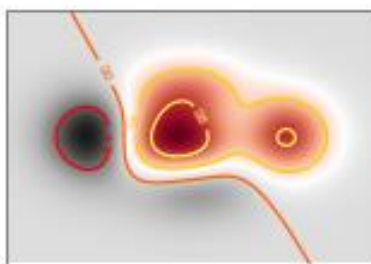
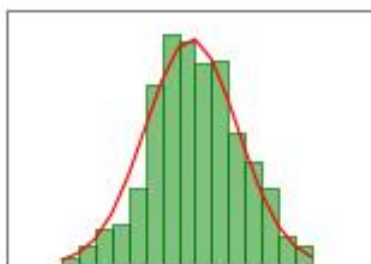
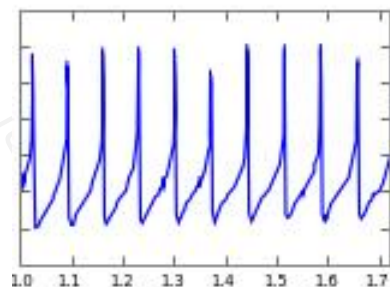



>>> import numpy as np
>>> from scipy import linalg
>>> arr = np.array([[1, 2], [3, 4]])
>>> linalg.det(arr)
-2.0

Matplotlib

特征

- 基于NumPy
- 二维绘图库，简单快速地生成曲线图、直方图和散点图等形式的图
- 常用的pyplot是一个简单提供类似MATLAB接口的模块



SymPy

特征



- 数学符号Python计算库，全功能的计算机代数系统，多项式求值、求极限、解方程、求积分、微分方程、级数展开等计算问题
- 代码简洁、易于理解和扩展



```
>>> from sympy import *  
>>> E**(1*pi)+1 # 内置符号  
>>> factorial(4) # 初等运算  
>>> x, y = symbols('x y') # 表达式求值  
>>> expr = x + 2*y  
>>> expr_temp = expr*x+1  
>>> limit(expr_temp, x, +oo) # 运算
```

pandas

特征



- 基于 SciPy library和 NumPy
- 高效的Series和DataFrame数据结构
- 强大的可扩展数据操作与分析的Python库
- 高效处理大数据集的切片等功能
- 提供优化库功能读写多种文件格式，如CSV、HDF5



```
...  
>>> df[2 : 5]  
>>> df.head(4)  
>>> df.tail(3)
```


人工智能程序设计

2 SCIPY中的数据结构

Python常用的数据结构



其他数据结构?



- **SciPy中的数据结构**

Python原有数据结构的变化

- ndarray (N维数组)
- Series (变长字典)
- DataFrame (数据框)