科学编程基础

5. 科学计算模块

余恒

北师大天文系

Numpy

- Numpy是Python的扩展模块，提供了矩阵、线性代数、傅里叶变换等的解决方法。
- NumPy包含：
  - N维矩阵对象
  - 线性代数运算功能
  - 傅里叶变换
  - Fortran代码集成的工具
  - C++代码集成的工具

Scipy模块

- SciPy 是基于Numpy构建的一个集成了多种数学算法和方便的函数的Python模块。SciPy的子模块需要单独import。
  - **constants** 物理和数学常数
  - **fftpack**快速傅立叶变换程序
  - **integrate** 积分和常微分方程求解器
  - **interpolate**拟合和平滑曲线
  - **linalg**线性代数
  - **optimize**最优路径选择
  - **signal**信号处理
  - **sparse**稀疏矩阵和以及相关程序
  - **special**特殊函数
  - **stats**统计上的函数和分布

数学数组

- Python中用列表(list)保存一组值，可以用来当作数组使用，不过由于列表的元素可以是任何对象，因此列表中所保存的是对象的指针。这样为了保存一个简单的[1,2,3]，需要有3个指针和三个整数对象。对于数值运算来说这种结构显然比较浪费内存和CPU计算时间。
- 此外Python还提供了一个array模块，可以直接保存数值，但是不支持多维，也没有各种运算函数，因此也不适合做数值运算。
- NumPy的诞生弥补了这些不足，NumPy提供了两种基本的对象：ndarray（N-dimensional array object）和ufunc（universal function object）。ndarray(下文统一称之为数组)是存储单一数据类型的多维数组，而ufunc则是能够对数组进行处理的函数。

数学数组创建

- `>>> import numpy as np`
- `>>> a = np.array([1, 2, 3, 4], dtype='int32')`
- `>>> a = np.array([[3,4,5],[3,6,7]])`
- `np.arange(0,1,0.1) np.zeros(2,3) np.ones(5)`
- `np.linspace(0, 1, 12)  np.logspace(0, 2, 20)`

- 函数式创建：

```
def func(i, j):
    return (i+1) * (j+1)
a = np.fromfunction(func, (9,9))
```

np.array 与 list 的区别

- a=range(5)
- a + 1
- a * 2
- a + a
- a > 3

```
np.array(a)
```

- b=np.arange(5)
- b + 1
- b * 2
- b + b
- b > 3
- b / 2

```
list(b)
```

一维数组取样

```
>>> a = np.arange(10)
>>> a[5] # 用整数作为下标可以获取数组中的某个元素
>>> a[3:5] # 用范围作为下标获取数组的一个切片，包括a[3]不包括a[5]
>>> a[:5] # 省略开始下标，表示从a[0]开始
>>> a[:-1] # 下标可以使用负数，表示从数组后往前数
array([0, 1, 2, 3, 4, 5, 6, 7, 8])
>>> a[2:4] = 100,101 # 下标还可以用来修改元素的值
>>> a[1:-1:2]    # 范围中的第三个参数表示步长，2表示隔一个元素取一个元素
```

二维数组取样

```
>>> a = np.arange(10).reshape(2,-1)
>>> a
array([[0, 1, 2, 3, 4],
       [5, 6, 7, 8, 9]])
>>> a[1,1]      #单个元素
6
>>> a[1]        #整行
array([5, 6, 7, 8, 9])
>>> a[:,2]      #整列
array([2, 7])
>>> a[0][::2]     #抽取某行特定元素
array([0, 2, 4])
```

条件取样

- a = np.arange(10).reshape(-1,2)

```
>>> a[a[:,1]>3]
array([[4, 5],
       [6, 7],
       [8, 9]])
>>> a[a[:,1]%3==0]
array([[2, 3],
       [8, 9]])
```

- a[(a[:,1]>3)*(a[:,1]%3==0)]

数组排序

- argsort函数返回数组值从小到大的索引
  - x = np.array([3,1,2])
  - np.argsort(x)
  - x[np.argsort(x)] # 排序后的数组

  - x=np.array([[0,3],[4,2]])
  - np.argsort(x, axis=1) # 排序每行
  - a[a[:,1].argsort()] # 按第二列排序

数学数组方法

- >>> a = np.arange(6).reshape(2,3)
- >>> a.shape　　(2, 3)
- >>> a.dtype　　dtype('int32')
- a.sum()　a.min()　a.max()　a.mean()
- a.reshape(3,2) <=>　a.T
- a.ravel()　#展开数组
- a.repeat(2,axis=0)　#复制元素

数组合并

```
>>> a = np.array([1, 2, 3])
>>> b = np.array([2, 3, 4])
>>> np.r_[a,b]
>>> np.hstack((a,b))
array([1, 2, 3, 2, 3, 4])
>>> np.vstack((a,b))
array([[1, 2, 3],
       [2, 3, 4]])
>>> np.c_[a,b]
array([[1, 2],
       [2, 3],
       [3, 4]])
```

随堂练习

- 生成矩阵：
  - 0~10 区间内均匀取10个值　[0,… 10]

```
[[0,0,0,0,0],
 [1,1,1,1,1],
 [2,2,2,2,2],
 [3,3,3,3,3],
 [4,4,4,4,4]]
```

```
[[0,1,2,3,4],
 [0,1,2,3,4],
 [0,1,2,3,4],
 [0,1,2,3,4],
 [0,1,2,3,4]]
```

数据保存

```
numpy.savetxt(fname, X, fmt='%.18e',
delimiter=' ', newline='\n', header='',
footer='', comments='# ')
```

```
>>> x = y = z = np.arange(0.0,5.0,0.5)
>>> np.savetxt('test.out', x, delimiter=',')
# X is an array
>>> np.savetxt('test.out', (x,y,z))
# x,y,z equal sized 1D arrays
>>> np.savetxt('test.out', x, fmt='%6.4f')
# use exponential notation
```

数据读取

```
numpy.loadtxt(fname, dtype=<type 'float'>,
comments='#', delimiter=None,
converters=None, skiprows=0, usecols=None,
unpack=False, ndmin=0)[source]
```

```
data = np.loadtxt('test.out', dtype = float)
data = np.loadtxt('test.out', usecols=[1])
```

和math函数比较

```
import time, math
import numpy as np
n = 1e+6

x = xrange(int(n))
start = time.clock()
for i in x:
    tmp = math.sin(i/n)
print "math.sin:", time.clock() - start

x = np.array(x)/n
start = time.clock()
np.sin(x)
print "numpy.sin:", time.clock() - start
```

星表读写

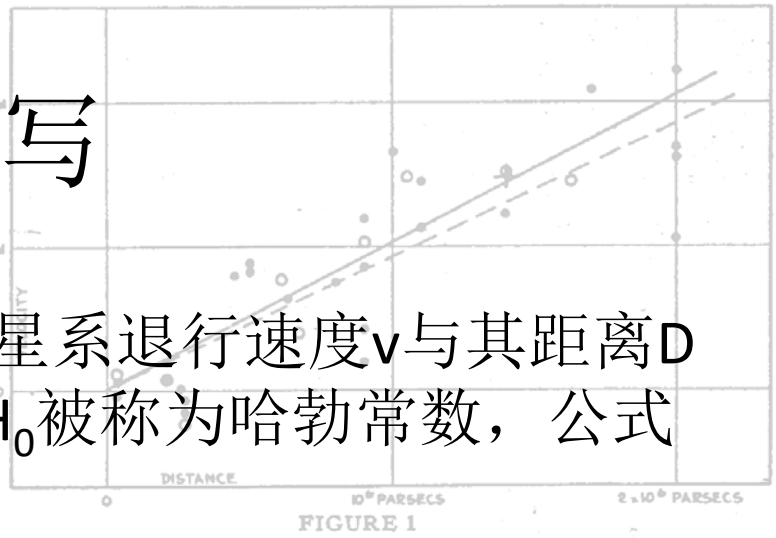

FIGURE 1

- 1929年美国天文学家哈勃发现星系退行速度v与其距离D成正比，即"哈勃定律"。比值H_0被称为哈勃常数，公式记为 $v = H_0D$ 。
- 1992年美国天文学家R.B.Tully等人系统测量了288个邻近星系的距离来确定银河系附近的物质分布。观测结果收录在CDS天文星表数据库中。可以利用其中的视向速度和距离来估计邻域的哈勃常数。
- 因为是期刊星表，对应的目录是J/ApJS/80/479 ：
- http://202.112.85.96/python/
  - http://cdsarc.u-strasbg.fr/viz-bin/Cat?J/ApJS/80/479
  - http://vizier.china-vo.org/ftp/cats/J_ApJS/80/479/

星表内容(table1.dat)

```
11-1    Virgo       130 92 26 15.6 1016
11-0+1  U  8036       1  1    16.8  893
11+2                 18  2    12.9  750
11+2    N  4713           1    10.9
11+2    N  4808           1    14.8
……
```

Bytes	Format	Units	Label	Explanations
1- 7	A7	---	Group	Group identification
9- 18	A10	---	Gal	Galaxy identification
20- 22	I3	---	Memb	Group members
24- 25	I2	---	Ngal	Number of total galaxies
27- 28	I2	---	Ngal2	Number of galaxies with distance
30- 33	F4.1	Mpc	Dist	Distance
35- 38	I4	km/s	Vel	Velocity
40	A1	---	note	Note

练习

- 读取星表
- 统计有红移测量数据的星系
- 将相应数据保存到文件当中
- 从文件中读取数据计算哈勃常数