# NumPy库入门



肖毅

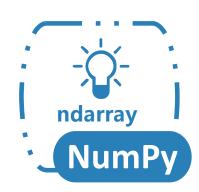


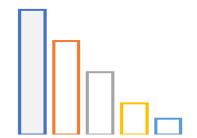
#### Python数据分析与展示

掌握表示、清洗、统计和展示数据的能力





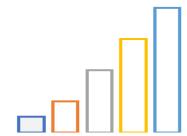


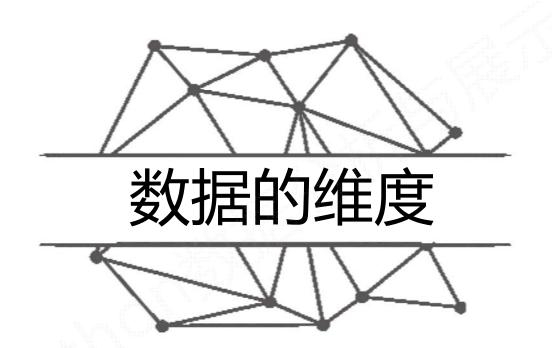












#### 从一个数据到一组数据

3.14



3.1404

3.1413

3.1401 3.1376

3.1398

3.1349

一个数据

表达一个含义

一组数据

表达一个或多个含义

#### 维度:一组数据的组织形式

```
3.1404
3.1413
3.1401 3.1376
3.1398
3.1349
```

3.1413,3.1398,3.1404,3.1401,3.1349,3.1376 或 3.1398, 3.1349, 3.1376 3.1413, 3.1404, 3.1401

一组数据

数据的组织形式

#### 一维数据

一维数据由对等关系的有序或无序数据构成,采用线性方式组织

3.1413, 3.1398, 3.1404, 3.1401, 3.1349, 3.1376

对应列表、数组和集合等概念

#### 列表和数组

#### 一组数据的有序结构

#### 区别

列表:数据类型可以不同

3.1413, 'pi', 3.1404, [3.1401, 3.1349], '3.1376'

数组:数据类型相同

3.1413, 3.1398, 3.1404, 3.1401, 3.1349, 3.1376

#### 二维数据

#### 二维数据由多个一维数据构成,是一维数据的组合形式

				指标得分
排名	学校名称	省市	总分	生源质量 (新生高考成绩得分)
1	清华大学	北京	94.0	100.0
2	北京大学	北京	81.2	96.1
3	浙江大学	浙江	77.8	87.2
4	上海交通大学	上海	77.5	89.4
5	复旦大学	上海	71.1	91.8
6	中国科学技术大学	安徽	65.9	91.9
7	南京大学	江苏	65.3	87.1
8	华中科技大学	湖北	63.0	80.6
9	中山大学	广东	62.7	81.1
10	哈尔滨工业大学	黑龙江	61.6	76.4

表格是典型的二维数据 其中,表头是二维数据的一部分

#### 多维数据

#### 多维数据由一维或二维数据在新维度上扩展形成

				指标得分					指标得分
排名	学校名称	省市	总分	生源质量(新生高考成绩得分) ▼	排名	学校名称	省市	总分	生源质量(新生高考成绩得分) ▼
1	清华大学	北京市	95.9	100.0	1	清华大学	北京	94.0	100.0
2	北京大学	北京市	82.6	98.9	2	北京大学	北京	81.2	96.1
3	浙江大学	浙江省	80	88.8 时间	住层	浙江大学	浙江	77.8	87.2
4	上海交通大学	上海市	78.7	90.6	4	上海交通大学	上海	77.5	89.4
5	复旦大学	上海市	70.9	90.4		复旦大学	上海	71.1	91.8
6	南京大学	江苏省	66.1	90.7	6	中国科学技术大学	安徽	65.9	91.9
7	中国科学技术大学	安徽省	65.5	90.1 2010	7	南京大学	江苏	65.3	87.1
8	哈尔滨工业大学	黑龙江省	63.5	80.9	8	华中科技大学	湖北	63.0	80.6
9	华中科技大学	湖北省	62.9	83,5	9	中山大学	广东	62.7	81.1
10	中山大学	广东省	62.1	81.8	10	哈尔滨工业大学	黑龙江	61.6	76.4

#### 高维数据

高维数据仅利用最基本的二元关系展示数据间的复杂结构

## 数据维度的Python表示

#### 数据维度是数据的组织形式

一维数据:列表和集合类型 [3.1398, 3.1349, 3.1376] 有序

{3.1398, 3.1349, 3.1376} 无序

二维数据:列表类型

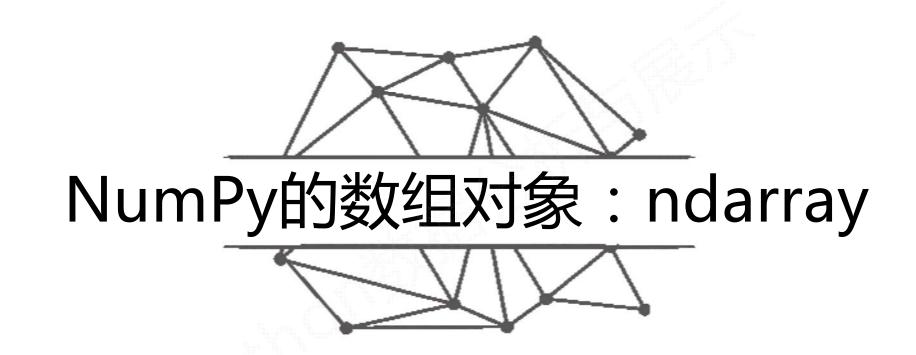
[ [3.1398, 3.1349, 3.1376],

[3.1413, 3.1404, 3.1401]]

多维数据:列表类型

## 数据维度的Python表示

#### 数据维度是数据的组织形式



#### NumPy

NumPy是一个开源的Python科学计算基础库,包含:

- 一个强大的N维数组对象 ndarray
- 广播功能函数
- 整合C/C++/Fortran代码的工具
- 线性代数、傅里叶变换、随机数生成等功能

NumPy是SciPy、Pandas等数据处理或科学计算库的基础

## NumPy的引用

import numpy as np

引入模块的别名

尽管别名可以省略或更改,建议使用上述约定的别名

#### N维数组对象: ndarray

Python已有列表类型,为什么需要一个数组对象(类型)?

例:计算 A2+B3,其中,A和B是一维数组

```
def pySum():
    a = [0, 1, 2, 3, 4]
    b = [9, 8, 7, 6, 5]
    c = []

    for i in range(len(a)):
        c.append(a[i]**2 + b[i]**3)

    return c

print(pySum())

import num

def npSum

a = np

b = np

c = a<sup>2</sup>

return

print(pySum())
```

```
import numpy as np

def npSum():
    a = np.array([0, 1, 2, 3, 4])
    b = np.array([9, 8, 7, 6, 5])

    c = a**2 + b**3

    return c

print(npSum())
```

### N维数组对象:ndarray

Python已有列表类型,为什么需要一个数组对象(类型)?

- 数组对象可以去掉元素间运算所需的循环,使一维向量更像单个数据
- 设置专门的数组对象,经过优化,可以提升这类应用的运算速度

观察:科学计算中,一个维度所有数据的类型往往相同

• 数组对象采用相同的数据类型,有助于节省运算和存储空间

### N维数组对象: ndarray

ndarray是一个多维数组对象,由两部分构成:

- 实际的数据
- 描述这些数据的元数据(数据维度、数据类型等)

ndarray数组一般要求所有元素类型相同(同质),数组下标从0开始

## ndarray实例

#### IPython提示符

ndarray在程序中的别名是:array

轴(axis):保存数据的维度;秩(rank):轴的数量

# ndarray对象的属性

属性	说明
.ndim	秩,即轴的数量或维度的数量
.shape	ndarray对象的尺度,对于矩阵,n行m列
.size	ndarray对象元素的个数,相当于.shape中n*m的值
.dtype	ndarray对象的元素类型
.itemsize	ndarray对象中每个元素的大小,以字节为单位

## ndarray实例



# ndarray的元素类型(1)

数据类型	说明
bool	布尔类型,True或False
intc	与C语言中的int类型一致,一般是int32或int64
intp	用于索引的整数,与C语言中ssize_t一致,int32或int64
int8	字节长度的整数,取值:[-128,127]
int16	16位长度的整数,取值:[-32768, 32767]
int32	32位长度的整数,取值:[-2 <sup>31</sup> , 2 <sup>31</sup> -1]
int64	64位长度的整数,取值:[-2 <sup>63</sup> , 2 <sup>63</sup> -1]

# ndarray的元素类型(2)

数据类型	说明
uint8	8位无符号整数,取值:[0, 255]
uint16	16位无符号整数,取值:[0,65535]
uint32	32位 <del>无符号</del> 整数,取值:[0,2 <sup>32</sup> -1]
uint64	32位无符号整数,取值:[0, 2 <sup>64</sup> -1] /(符号)尾数*10 <sup>指数</sup>
float16	16位半精度浮点数:1位符号位,5位指数,10位尾数
float32	32位半精度浮点数:1位符号位,8位指数,23位尾数
float64	64位半精度浮点数:1位符号位,11位指数,52位尾数

## ndarray的元素类型(3)

数据类型	说明
complex64	复数类型,实部和虚部都是32位浮点数
complex128	复数类型,实部和虚部都是64位浮点数

实部(.real) + j虚部(.imag)

### ndarray的元素类型

ndarray为什么要支持这么多种元素类型?

对比: Python语法仅支持整数、浮点数和复数3种类型

- 科学计算涉及数据较多,对存储和性能都有较高要求
- 对元素类型精细定义,有助于NumPy合理使用存储空间并优化性能
- 对元素类型精细定义,有助于程序员对程序规模有合理评估

# 非同质的ndarray对象

In [62]: x = np.array([ [0, 1, 2, 3, 4],

```
ndarray数组可以由非同质对象构成
                 [9, 8, 7, 6] ])
In [63]: x.shape
Out[63]: (2,)
In [64]: x.dtype
Out[64]: dtype('0')
In [65]: x
                                     非同质ndarray元素为对象类型
Out[65]: array([[0, 1, 2, 3, 4], [9, 8, 7, 6]],
dtype=object)
In [66]: x.itemsize
Out[66]: 4
In [67]: x.size
                   非同质ndarray对象无法有效发挥NumPy优势,尽量避免使用
Out[67]: 2
```



- 从Python中的列表、元组等类型创建ndarray数组
- 使用NumPy中函数创建ndarray数组,如:arange, ones, zeros等
- 从字节流 (raw bytes)中创建ndarray数组
- 从文件中读取特定格式,创建ndarray数组

(1)从Python中的列表、元组等类型创建ndarray数组

x = np.array(list/tuple)

x = np.array(list/tuple, dtype=np.float32)

当np.array()不指定dtype时, NumPy将根据数据情况关联一个dtype类型

(1)从Python中的列表、元组等类型创建ndarray数组

(2)使用NumPy中函数创建ndarray数组,如:arange,ones,zeros等

函数	说明		
np.arange(n)	类似range()函数,返回ndarray类型,元素从0到n-1		
np.ones(shape)	根据shape生成一个全1数组,shape是元组类型		
np.zeros(shape)	根据shape生成一个全0数组,shape是元组类型		
np.full(shape,val)	根据shape生成一个数组,每个元素值都是val		
np.eye(n)	创建一个正方的n*n单位矩阵,对角线为1,其余为0		

```
In [73]: np.arange(10)
Out[73]: array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
In [74]: np.ones((3,6))
Out[74]:
array([[ 1., 1., 1., 1., 1.],
      [ 1., 1., 1., 1., 1.],
      [1., 1., 1., 1., 1., 1.]])
In [75]: np.zeros((3,6), dtype=np.int32)
Out[75]:
array([[0, 0, 0, 0, 0, 0],
      [0, 0, 0, 0, 0, 0],
      [0, 0, 0, 0, 0, 0]])
In [76]: np.eye(5)
Out[76]:
array([[ 1., 0., 0., 0., 0.],
      [0., 1., 0., 0., 0.],
      [0., 0., 1., 0., 0.],
      [0., 0., 0., 1., 0.],
      [0., 0., 0., 0., 1.]])
```

(2)使用NumPy中函数创建ndarray数组,如:arange, ones, zeros等

函数	说明
np.ones_like(a)	根据数组a的形状生成一个全1数组
np.zeros_like(a)	根据数组a的形状生成一个全0数组
np.full_like(a,val)	根据数组a的形状生成一个数组,每个元素值都是val

(3)使用NumPy中其他函数创建ndarray数组

函数	说明
np.linspace()	根据起止数据等间距地填充数据,形成数组
np.concatenate()	将两个或多个数组合并成一个新的数组



# ndarray数组的变换

对于创建后的ndarray数组,可以对其进行维度变换和元素类型变换

a = np.ones((2,3,4), dtype=np.int32)

# ndarray数组的维度变换

方法	说明
.reshape(shape)	不改变数组元素,返回一个shape形状的数组,原数组不变
.resize(shape)	与.reshape()功能一致,但修改原数组
.swapaxes(ax1,ax2)	将数组n个维度中两个维度进行调换
.flatten()	对数组进行降维,返回折叠后的一维数组,原数组不变

#### ndarray数组的维度变换

a = np.ones((2,3,4), dtype=np.int32)

#### ndarray数组的维度变换

a = np.ones((2,3,4), dtype=np.int32)

### ndarray数组的类型变换

new\_a = a.astype(new\_type)

```
In [121]: b = a.astype(np.float)
In [119]: a = np.ones((2,3,4), dtype=np.int)
In [120]: a
                                                   In [122]: b
Out[120]:
                                                   Out[122]:
array([[1, 1, 1, 1],
                                                   array([[[ 1., 1., 1., 1.],
       [1, 1, 1, 1],
                                                           [ 1., 1., 1., 1.].
       [1, 1, 1, 1]],
                                                           [1., 1., 1., 1.]
      [[1, 1, 1, 1],
                                                          [[ 1., 1., 1., 1.],
      [1, 1, 1, 1],
                                                          [ 1., 1., 1., 1.],
       [1, 1, 1, 1]]])
                                                           [1., 1., 1., 1.]
```

astype()方法一定会创建新的数组(原始数据的一个拷贝),即使两个类型一致

#### ndarray数组向列表的转换

ls = a.tolist()



索引:获取数组中特定位置元素的过程

切片:获取数组元素子集的过程

一维数组的索引和切片:与Python的列表类似

```
In [131]: a = np.array([9, 8, 7, 6, 5])

In [132]: a[2]
Out[132]: 7

起始编号: 终止编号(不含): 步长,3元素冒号分割

In [133]: a[1:4:2]
Out[133]: array([8, 6])

编号0开始从左递增,或-1开始从右递减
```

#### 多维数组的索引:

```
In [148]: a[1, 2, 3]
Out[148]: 23
In [149]: a[0, 1, 2]
Out[149]: 6
In [150]: a[-1, -2, -3]
Out[150]: 17
```

每个维度一个索引值,逗号分割

#### 多维数组的切片:

```
In [158]: a[:, 1, -3]
Out[158]: array([ 5, 17])
                              选取一个维度用:
In [159]: a[:, 1:3, :]
Out[159]:
                             每个维度切片方法
array([[[ 4, 5, 6, 7],
      [8, 9, 10, 11]],
                              与一维数组相同
     [[16, 17, 18, 19],
      [20, 21, 22, 23]]])
In [160]: a[:, :, ::2]
                             每个维度可以使用
Out[160]:
array([[[ 0, 2],
                               步长跳跃切片
      [4, 6],
      [ 8, 10]],
     [[12, 14],
      [16, 18],
      [20, 22]]])
```



#### 数组与标量之间的运算

#### 数组与标量之间的运算作用于数组的每一个元素

#### 实例

## NumPy一元函数

#### 对ndarray中的数据执行元素级运算的函数

函数	说明
np.abs(x) np.fabs(x)	计算数组各元素的绝对值
np.sqrt(x)	计算数组各元素的平方根
np.square(x)	计算数组各元素的平方
<pre>np.log(x) np.log10(x) np.log2(x)</pre>	计算数组各元素的自然对数、10底对数和2底对数
<pre>np.ceil(x) np.floor(x)</pre>	计算数组各元素的ceiling值 或 floor值

# NumPy一元函数

函数	说明
np.rint(x)	计算数组各元素的四舍五入值
np.modf(x)	将数组各元素的小数和整数部分以两个独立数组形式返回
<pre>np.cos(x) np.cosh(x) np.sin(x) np.sinh(x) np.tan(x) np.tanh(x)</pre>	计算数组各元素的普通型和双曲型三角函数
np.exp(x)	计算数组各元素的指数值
np.sign(x)	计算数组各元素的符号值 , 1(+),0,-1(-)

## NumPy一元函数实例

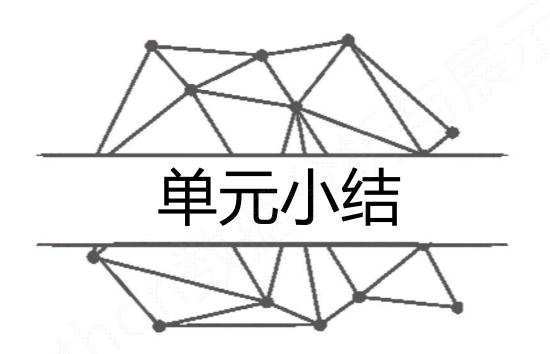
```
In [176]: a = np.arange(24).reshape((2,3,4))
                                                         In [180]: np.modf(a)
                                                         Out[180]:
In [177]: np.square(a)
                                                         (array([[[ 0. , 0. , 0.41421356, 0.73205081],
Out[177]:
                                                                 [ 0. , 0.23606798, 0.44948974, 0.64575131],
array([[ 0, 1, 4, 9],
                                                                 [0.82842712, 0., 0.16227766, 0.31662479]],
      [ 16, 25, 36, 49],
      [ 64, 81, 100, 121]],
                                                                [[0.46410162, 0.60555128, 0.74165739, 0.87298335],
                                                                [ 0. , 0.12310563, 0.24264069, 0.35889894],
      [[144, 169, 196, 225],
                                                                 [ 0.47213595, 0.58257569, 0.69041576, 0.79583152]]]),
      [256, 289, 324, 361],
                                                          array([[[ 0., 1., 1., 1.],
      [400, 441, 484, 529]]], dtype=int32)
                                                                [2., 2., 2., 2.],
                                                                 [2., 3., 3., 3.]],
In [178]: a = np.sqrt(a)
                                                                [[ 3., 3., 3., 3.],
In [179]: a
                                                                [4., 4., 4., 4.]
Out[179]:
                                                                 [4., 4., 4., 4.]]))
array([[[ 0. , 1. , 1.41421356, 1.73205081],
      [ 2. , 2.23606798, 2.44948974, 2.64575131],
      [ 2.82842712, 3. , 3.16227766, 3.31662479]],
      [[3.46410162, 3.60555128, 3.74165739, 3.87298335],
      [4. , 4.12310563, 4.24264069, 4.35889894],
      [ 4.47213595, 4.58257569, 4.69041576, 4.79583152]]])
```

# NumPy二元函数

函数	说明
+ - * / **	两个数组各元素进行对应运算
<pre>np.maximum(x,y) np.fmax() np.minimum(x,y) np.fmin()</pre>	元素级的最大值/最小值计算
np.mod(x,y)	元素级的模运算
np.copysign(x,y)	将数组y中各元素值的符号赋值给数组x对应元素
> < >= <= !=	算术比较,产生布尔型数组

## NumPy二元函数实例

```
In [203]: a = np.arange(24).reshape((2,3,4))
In [204]: b = np.sqrt(a)
In [205]: a
Out[205]:
array([[0, 1, 2, 3],
       [4, 5, 6, 7],
       [8, 9, 10, 11]],
      [[12, 13, 14, 15],
       [16, 17, 18, 19],
       [20, 21, 22, 23]]])
In [206]: b
Out[206]:
       [ 0. , 1. , 1.41421356, 1.73205081], [ 2. , 2.23606798, 2.44948974, 2.64575131],
array([[[ 0.
       [ 2.82842712, 3. , 3.16227766, 3.31662479]],
      [[3.46410162, 3.60555128, 3.74165739, 3.87298335],
       [4. , 4.12310563, 4.24264069, 4.35889894],
       [ 4.47213595, 4.58257569, 4.69041576, 4.79583152]]])
```



# NumPy库入门

数据的维度:一维、多维、高维

#### ndarray类型属性、创建和变换

np.arange(n)

np.ones(shape)

np.zeros(shape)

np.full(shape,val)

np.eye(n)

np.ones\_like(a)

np.zeros\_like(a)

np.full\_like(a,val)

数组的索引

和切片

数组的运算

一元函数

二元函数

.ndim

.reshape(shape)

.shape

.resize(shape)

.size

.swapaxes(ax1,ax2)

.dtype

.flatten()

.itemsize