1. Python
2. Numpy

NumPy的部分功能如下（高效处理大数组的数据。）：

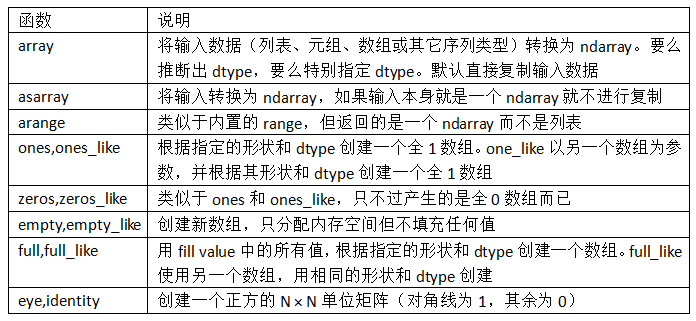
* ndarray，一个具有矢量算术运算和复杂广播能力的快速且节省空间的多维数组。
* 用于对整组数据进行快速运算的标准数学函数（无需编写循环）。
* 用于读写磁盘数据的工具以及用于操作内存映射文件的工具。
* 线性代数、随机数生成以及傅里叶变换功能。
* 用于集成由C、C++、Fortran等语言编写的代码的A C API。

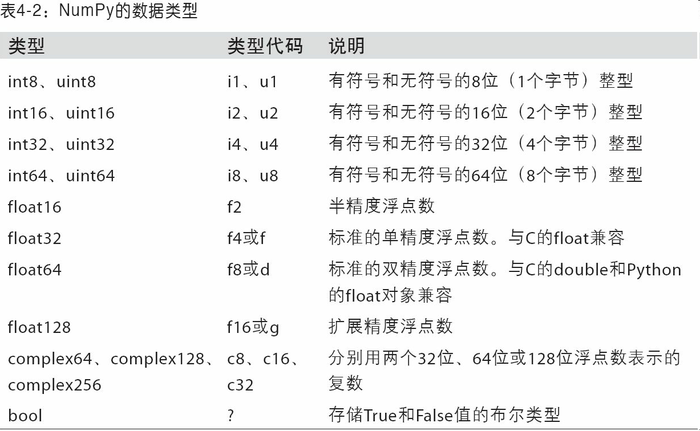
1. 创建ndarray
2. data1 = [6, 7.5, 8, 0, 1]

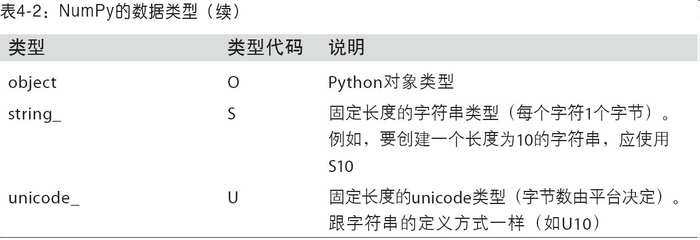
arr1 = np.array(data1)

print(arr1)

1. np.zeros(10)、np.empty((2, 3, 2))
2. np.arange(15)







1. ndarray数据类型

获取类型 ndarray.dtype

数组类型转换ndarray.astype(np.期望的数据类型)

1. Numpy数组运算
2. 数组广播

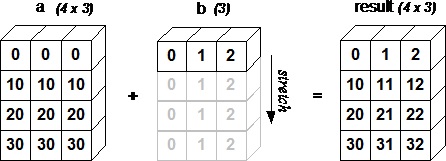
不同维度数组不能进行元素到元素操作，但可以广播计算。

如果满足以下规则，可以进行广播：

* ndim较小的数组会在前面追加一个长度为 1 的维度。
* 输出数组的每个维度的大小是输入数组该维度大小的最大值。
* 如果输入在每个维度中的大小与输出大小匹配，或其值正好为 1，则可以在计算中使用该输入。
* 如果输入的某个维度大小为 1，则该维度中的第一个数据元素将用于该维度的所有计算。

如果上述规则产生有效结果，并且满足以下条件之一，那么数组被称为可广播的。

* 数组拥有相同形状。
* 数组拥有相同的维数，每个维度拥有相同长度，或者长度为 1。
* 数组拥有极少的维度，可以在其前面追加长度为 1 的维度，使上述条件成立。



1. 基本索引和切片

同matlab，数据不会被复制，视图上的任何修改都会直接反映到源数组上。

复制数组使用ndarray.copy()

1. 切片索引

同matlab

1. 布尔值索引
2. data=np.random.randn(n,m)

condition=[true or false,...,true or false] (n\*1)

data[condition]

1. 1
2. 花式索引（Fancy indexing）

arr = np.arange(32).reshape((8, 4))

print(arr)

print(arr[[1, 5, 7, 2], [0, 3, 1, 2]])

1. 数组转置和轴对换 .T、.transpose()、.swapaxes()

利用转置计算矩阵内积：

arr = np.random.randn(6, 3)

print (arr)

print (np.dot(arr.T, arr))

换轴：

arr = np.arange(16).reshape((2, 2, 4))

print(arr)

print('==========分割线==============')

print(arr.transpose((1, 0, 2))) //第一个轴被换成了第二个，第二个轴被换成了第一个，最后一个轴不变。

print(arr)

print('==========分割线==============')

print(arr.swapaxes(1, 2))

1. 11
2. Pandas

[1. 数据结构简介：DataFrame和Series](https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/-1?lang=" \l "Chapter1" \t "https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/_self)

1. Series

[2. 数据索引index](https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/-1?lang=" \l "Chapter2" \t "https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/_self)

[3. 利用pandas查询数据](https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/-1?lang=" \l "Chapter3" \t "https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/_self)

[4. 利用pandas的DataFrames进行统计分析](https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/-1?lang=" \l "Chapter4" \t "https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/_self)

[5. 利用pandas实现SQL操作](https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/-1?lang=" \l "Chapter5" \t "https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/_self)

[6. 利用pandas进行缺失值的处理](https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/-1?lang=" \l "Chapter6" \t "https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/_self)

[7. 利用pandas实现Excel的数据透视表功能](https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/-1?lang=" \l "Chapter7" \t "https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/_self)

[8. 多层索引的使用](https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/-1?lang=" \l "Chapter8" \t "https://tianchi.aliyun.com/mas-notebook/preview/429/4d9ebae0-9166-11e8-8d9d-7348116604e7.ipynb/_self)