

数据处理常用库: numpy

1. 什么是 NumPy?

NumPy (Numerical Python) 是 Python 的一个开源数值计算库，提供了多维数组对象和一些用于处理数组的函数。它是进行科学计算和数据分析的核心工具之一。

2. NumPy 数组 (ndarray)

NumPy 的核心是多维数组对象 ndarray，用于存储相同类型的数据。它可以是一维、二维、三维甚至更高维度的数组。

2.1 创建 NumPy 数组

```
import numpy as np

# 从 Python 列表创建数组

arr1 = np.array([1, 2, 3, 4, 5])

# 从列表的列表创建二维数组

arr2 = np.array([[1, 2, 3], [4, 5, 6]])

# 基于范围创建数组

arr3 = np.arange(10)

# 使用 numpy 提供的函数创建数组

zero_arr = np.zeros((3, 3)) # 创建一个全为 0 的 3x3 数组
```

```
ones_arr = np.ones((2, 2)) # 创建一个全为 1 的 2x2 数组
```

2.2 数组属性和操作

```
arr = np.array([1, 2, 3, 4, 5])

print("Shape:", arr.shape) # 数组形状

print("Size:", arr.size) # 数组元素个数

print("Max:", arr.max()) # 最大值

print("Min:", arr.min()) # 最小值

print("Sum:", arr.sum()) # 元素求和


# 索引和切片操作

print(arr[0]) # 索引取值

print(arr[1:4]) # 切片取值
```

3.数组运算

NumPy 支持对数组进行逐元素的运算，以及数组之间的运算。

```
a = np.array([1, 2, 3])

b = np.array([4, 5, 6])
```

```
c = a + b # 数组相加
```

```
d = a * 2 # 数组元素乘以常数
```

```
e = np.dot(a, b) # 数组的点积
```

4. 常用函数

4.1 np.where

np.where 函数用于根据条件从数组中选择元素。

```
arr = np.array([1, 2, 3, 4, 5])
```

```
condition = arr > 2
```

```
result = np.where(condition, arr, 0) # 大于 2 的保留, 小于等于 2 的变为 0
```

4.2 np.concatenate

np.concatenate 函数用于连接数组。

```
arr1 = np.array([1, 2, 3])
```

```
arr2 = np.array([4, 5, 6])
```

```
concatenated_arr = np.concatenate((arr1, arr2)) # 连接两个数组
```

5. 数组的形状操作

```
arr = np.array([[1, 2, 3], [4, 5, 6]])

print("Shape:", arr.shape) # 输出: (2, 3)


reshaped_arr = arr.reshape(3, 2) # 重新调整形状

flattened_arr = arr.flatten() # 展平成一维数组
```

6. 广播 (Broadcasting)

广播是 NumPy 中一种处理不同形状数组之间运算的方式，使得不同形状的数组能够进行逐元素运算。广播的基本原理是将低维度的数组沿缺失维度方向复制，直到两数组有相同大小

```
arr = np.array([[1, 2, 3], [4, 5, 6]])

scalar = 10


result = arr + scalar # 数组与标量的运算
```