

Intro to NumPy



SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority



Agenda



What is NumPy



Why use NumPy



NumPy Core Features



Multidimensional Arrays



Broadcasting Mechanism



Mathematical and Statistical Operations



Integration with Other Libraries



▶ What is NumPy

NumPy (Numerical Python) is essential for scientific computing in Python, offering efficient handling of large numerical datasets through its powerful array object.

It introduces powerful multidimensional arrays to Python, offering efficient storage and manipulation of numerical data.

NumPy forms the basis of many higher-level scientific libraries.

Significance:

- Acts as the foundation for Python's data science stack, including libraries like Pandas and SciPy.
- Essential for tasks in data analysis, machine learning, and scientific computing.



Why use NumPy

Performance:

- Uses less memory and operates faster than Python lists due to its efficient storage and execution in C.

Versatility:

- Applicable across various domains with tools for linear algebra, statistics, and more, ensuring broad utility in data science.

Ease of Use:

- Simplifies mathematical operations on arrays with vectorized operations and broadcasting, making code cleaner and more readable.

Community and Support:

- Benefits from a large, active community, rich documentation, and open-source collaboration, ensuring it's continuously improved and well-supported.





Core Features

Multidimensional Arrays:

- Efficient data storage that supports various dimensions (1D, 2D, etc.), with elements of a uniform data type for faster processing.

Vectorized Operations:

- Perform operations on entire arrays without loops for cleaner, faster code. Supports a wide range of mathematical functions directly applied to arrays.

Broadcasting:

- Simplifies operations between arrays of different sizes, automatically adjusting sizes without manual intervention.

Mathematical Functions:

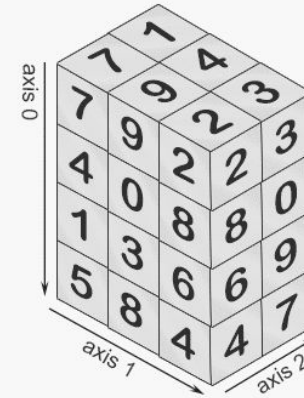
- Offers a vast collection of optimized mathematical functions for linear algebra, statistics, and more.



Multidimensional Arrays

- Efficient storage for large numerical datasets with elements of the same data type.
- Supports 1D, 2D, 3D, and higher-dimensional arrays for versatile data representation.
- Easy data reshaping, slicing, and indexing for effective data manipulation.
- Example:
 - 1D = [1,5,6,9]
 - 2D = [[3.5,4.0,6.5],[0.4,0.9,4.7]]

3D Array



shape : (4, 3, 2)

1D Array



axis 0

shape : (4,)

2D Array



axis 1

shape : (2, 3)

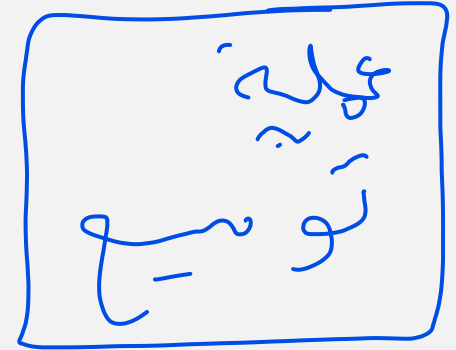
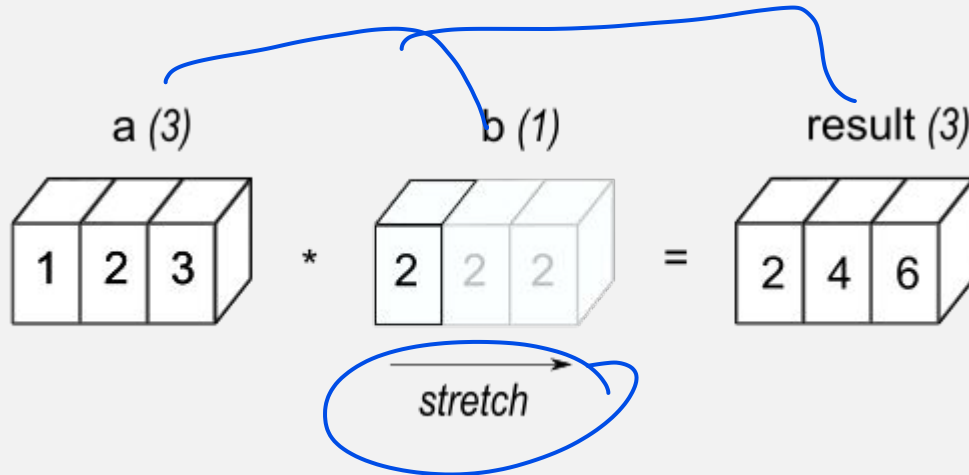


Broadcasting Mechanism

Enables operations on arrays of different shapes by temporarily expanding the smaller array.

Increases efficiency and simplifies code by eliminating the need for manual array resizing.

Example: `np.array([1, 2, 3]) + 2` adds 2 to each element, illustrating scalar to array operation without duplicating data.



العمليات الحسابية التي تقوم بها NumPy

Mathematical and Statistical Operations

NumPy provides a wide range of mathematical functions that operate directly on arrays for linear algebra, statistics, and more.

Linear Algebra Examples:

- `linalg.solve` for solving linear equations
- `dot` for performing dot products

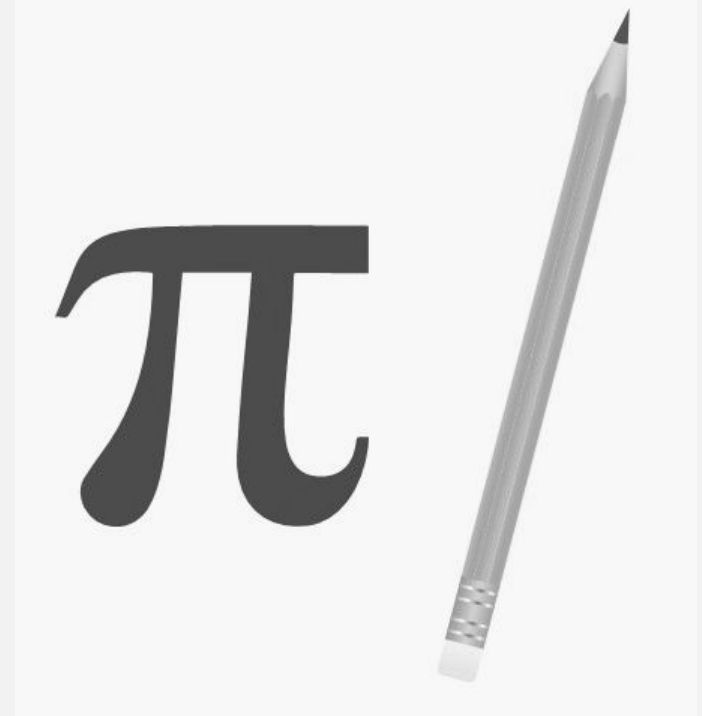
Statistics Examples:

- `mean` for calculating average
- `median` for calculating median
- `std` for calculating standard deviation

Calculus:

- `diff` for differentiation
- `cumsum` for cumulative sums

These operations are highly optimized and execute much faster than if implemented in pure Python, especially for large datasets.



Integration with Other Libraries

NumPy arrays serve as the foundational data structure for other libraries, enabling seamless data manipulation and analysis across the ecosystem.

Key Integrations:

- **Pandas**: Utilizes NumPy for handling large datasets and performing operations on time series and tabular data.
- **SciPy**: Builds on NumPy's array for advanced mathematical functions, optimization, and signal processing.
- **Scikit-learn**: Leverages NumPy arrays for implementing machine learning algorithms on large datasets.
- **Matplotlib**: Uses NumPy arrays to plot and visualize data in various formats.

to go !!



Thank you



SDAIA

الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority