# **Learn Numpy & Pandas**

# What is Numpy?

NumPy (Numerical Python) is a powerful library in Python for numerical computing. It provides support for working with large, multi-dimensional arrays and matrices, along with mathematical functions to operate on these data structures efficiently.

**Key Features of NumPy:**

* **N-dimensional array:** Supports efficient storage and manipulation of multi-dimensional arrays.
* **Mathematical functions:** Includes operations like linear algebra, Fourier transform and statistical computations
* **Broadcasting:** Enables performing operations on arrays of different shapes seamlessly
* **Performance Optimization:** Written in C, making computations significantly faster than Python’s built-in list operations
* **Integration:** Works well with other scientific computing libraries like Pandas, SciPy, and Matplotlib.

**Why use NumPy?**

**Speed:** Much faster compared to Python lists when dealing with numerical data.

**Memory Efficiency:** Uses less memory than Python lists for storing data.

**Ease of Computation:** Allows vectorized operations, eliminating the need for explicit loops.

**Interoperability:** Compatible with various data formats and scientific computing frameworks.

## **Installation of NumPy**

Installing NumPy is straightforward and can be done using pip or conda, depending on your preference. Here’s a step-by-step guide for Windows, Linux, and macOS:

### **Windows Installation:**

1. Open Command Prompt
2. Check Python Installation

python – version

If Python is not installed, download and install it from python.org

1. Install NumPy using pip

pip install numpy

1. Verify Installation

Python -c “import numpy; print(numpy.\_\_version\_\_)”

### **Linux Installation**

1. Open Terminal
2. Check Python Installation

python3 –version

If Python is missing, install it using

sudo apt install python3

1. Install pip (if not installed)

sudo apt install python3-pip

1. Install Numpy

pip3 install numpy

1. Verify Installation

Python3 -c “import numpy; print(numpy.\_\_version\_\_)”

### **macOS**

1. Open Terminal
2. Check Python Installation

python3 – version

If Python is missing, install it using

brew install python

1. Install pip (if not installed)

python3 -m ensurepip

1. Install Numpy

pip install numpy

1. Verify installation

python3 -c “import numpy; print(numpy.\_\_version\_\_)”

Alternatively, you can install NumPy using conda

conda install numpy

This method is useful if you are using Anaconda or Miniconda

## **Creating and Manipulating Arrays**

### **1D Array (One-Dimensional)**

**A 1D Array is simply a list-like array with single dimension.**

**Creation:**

import numpy as np

# Creating a 1D array

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

print(arr\_id)

**Manipulation of Array:**

**Access Elements:** arr\_id[2]

**Slicing:** arr\_id[1:3]

**Mathematical Operations:** arr\_id \* 2

### **2D Array (Two-Dimensional)**

A 2D Array represents rows and column, like a table or matrix

**Creation:**

arr\_2d = np.array([[1,2,3],[4,5,6]])

print(arr\_2d)

**Manipulation of Array:**

**Access Elements:** arr\_2d[1,2]

**Row slicing:** arr\_2d[0,:]

**Column slicing:** arr\_2d[:,-1]

**Matrix Operations:** np.dot(arr\_2d, np.array([[1], [2], [3]]))

### **3D Array (Three-Dimensional)**

A **3D array** adds depth, like multiple 2D arrays stacked.

**Creation:**

python

arr\_3d = np.array([[[1, 2], [3, 4]], [[5, 6], [7, 8]]])

print(arr\_3d)

**Manipulation:**

* **Access elements:** arr\_3d[1, 0, 1] → Retrieves 6.
* **Slicing:** arr\_3d[:, :, 0] → Extracts all first elements from each depth layer.
* **Reshape:** arr\_3d.reshape(4, 2) → Changes shape while preserving data.

### **4D Array (Four-Dimensional)**

A **4D array** is useful for representing complex datasets like images or time-series data.

**Creation:**

python

arr\_4d = np.random.randint(1, 10, (2, 2, 2, 2)) # Creates a random 4D array

print(arr\_4d)

**Manipulation:**

1. **Access elements:** arr\_4d[1, 0, 1, 0] → Retrieves a specific number.
2. **Transpose:** arr\_4d.transpose(1, 0, 2, 3) → Rearranges dimensions.
3. **Reshape:** arr\_4d.reshape(8, 2) → Flattens some dimensions while keeping structure.

**Summary**

* **1D**: Simple list-like array.
* **2D**: Table/matrix-like format.
* **3D**: Stack of 2D arrays.
* **4D**: Complex structured data, useful for advanced computing.

## **Difference between NumPy Array and Python List**

NumPy arrays and Python Lists might seem similar at first glance, but they have key differences:

**Speed & Performance**

* NumPy arrays are significantly faster than Python lists because they use contiguous memory allocation and vectorized operations.
* Python lists store references to objects, making them slower when performing numerical computations
* print ("""
* ######################################################################
* #####     Speed Comparison between NumPy vs Python Lists         #####
* ######################################################################
* """)
* import numpy as np
* import time
* # Creating large list and NumPy array
* list\_data = list(range(1000000))
* numpy\_array = np.arange(1000000)
* # Timing list operations
* start = time.time()
* list\_result = [x \* 2 for x in list\_data] # using list comprehension
* end = time.time()
* print(f"List computation time: {end - start:.5f} seconds")
* # Timing NumPy array operations
* start = time.time()
* numpy\_result = numpy\_array \* 2 # vectorized operation
* end = time.time()
* print(f"NumPy computation time: {end - start: .5f} seconds")
* print ("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")

**Memory Efficiency**

* NumPy arrays consume less memory than lists because they store elements as a homogenous type (all elements must be of the same data type)
* Lists, on the other hand, store elements as individual objects, leading to higher memory usage
* print("""
* ######################################################################
* #####  Memory utilization Comparison between NumPy vs Python List ####
* ######################################################################
* """)
* import numpy as np
* import sys
* # Creating a Python list and a NumPy array with the same elements
* python\_list = [i for i in range(1000)]
* numpy\_array = np.arange(1000)
* # Checking memory usage
* python\_list\_size = sys.getsizeof(python\_list) + sum(sys.getsizeof(i) for i in python\_list)
* numpy\_array\_size = numpy\_array.nbytes
* print(f"Memory used by Python list: {python\_list\_size} bytes")
* print(f"Memory used by NumPy array: {numpy\_array\_size} bytes")
* print ("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")

**Functionality & Operations**

* NumPy supports element-wise operations, meaning you can perform arithmetic operations directly on arrays.
* Python lists require explicit looping or list comprehensions to achieve similar operations.

**Type Consistency**

* NumPy arrays enforce a single data type (e.g., all elements must be integers or floats)
* Lists can hold mixed data types, such as integers, strings, or even other lists

**Built-in Methods**

* NumPy provides numerous powerful mathematical and statistical functions
* List rely on Python’s built-in methods (append(), sort(), remove() etc.,) which are more general-purpose

*Bottom Line:* If you're working with numerical data or large datasets, NumPy arrays are the way to go for efficiency and speed. If you need flexibility and mixed data types, Python lists might be better.

## **NumPy Arrays**

### **Creating NumPy arrays**

NumPy provides several ways to create arrays, each suited to different use cases.

#### **np.array**

This is the most basic way to create a NumPy array using Python list or tuples

Import numpy as np

# Creating a NumPy array from a Python list

list\_data = [1,2,3,4]

numpy\_array = np.array(list\_data)

print(numpy\_array)

print(type(numpy\_array))

* np.array() converts a list or tuple into NumPy array
* It automatically determines the data type (dtype) of elements

#### **np.arange**

This function creates arrays with evenly spaced values

# Creating an array with values from 0 to 9

numpy\_array2 = np.arange(10)

print(numpy\_array2)

# Creating an array with a specific step

numpy\_array3 = np.arange(1,10,2)

print(numpy\_array3)

#### **np.linspace**

Generates a specified number of evenly spaced values between a given range.

# Creating an array from 0 to 10 with 5 elements

numpy\_array4 = np.linspace(0,10, 5)

print(numpy\_array4) # Output: [0. 2.5 5. 7.5 10.]

* np.linspace(start, stop, num\_elements) ensures precision when dividing a range into equal parts
* Often used in plotting or mathematical computations

#### **np.zero() and np.ones()**

Create arrays filled with zeros or ones

# Creating a 1D array with five zeros

zeros\_array = np.zeros(5)

print(zeros\_array) # Output: [ 0. 0. 0. 0. 0. ]

# Creating a 2x3 matrix filled with ones

ones\_array = np.ones((2,3))

print (ones\_array)

output:

[[1. 1. 1.]

[1. 1. 1.]]

* np.zeros(shape) creates an array filled with zeros
* np.ones(shape) creates an array filled with ones

#### **np.full()**

Creates an array filled with specific values

# Creating a 3 x 3 array filled with value 7

full\_array = np.full((3,3),7)

print (full\_array)

output:

[[7 7 7]

[7 7 7]

[7 7 7]]

* np.full(shape, value) creates an array with a predefined value
* handy for defining constant matrices

#### **np.eye()**

Generates an identity matrix

# Create a 4 x 4 identity matrix

identity\_matrix = np.eye(4)

print(identity\_matrix)

output:

[[1. 0. 0. 0.]

[0. 1. 0. 0.]

[0. 0. 1. 0.]

[0. 0. 0. 1.]]

* Identity matrices are often used in linear algebra operations

#### **np.random()**

Creates arrays filled with random values

# Creating a 3 x 3 array with random numbers between 0 and 1

random\_array = np.random.rand(3,3)

print (random\_array)

Output:

[[0.342 0.892 0.102]

[0.423 0.015 0.923]

[0.712 0.831 0.531]]

* np.random.rand(shape) generates uniform random numbers between 0 and 1
* np.random.randint(low, high, shape) generates random integers

#### **np.empty()**

Creates an uninitialized array (contains garbage values)

# Creating an empty array

empty\_array = np.empty((2,3))

print(empty\_array)

'''

Output (values may vary):

[[4.66706551e-310 2.14156671e-316 0.00000000e+000]

[0.00000000e+000 5.02034658e+175 5.56417452e-309]]

'''

* np.empty(shape) allocates memory but does not initialize values.
* Faster than np.zeros() and np.ones() when initial values are irrelevant.

### **Understanding array data types**

NumPy arrays are homogenous, meaning all elements must have the same data type. The dtype attribute in NumPy defines the type of elements stored in an array, ensuring efficient memory usage and optimized computations

#### **Checking the Data Type of an Array**

You can check the data type of a NumPy array using .dtype

import numpy as np

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

print(arr.dtype) # Output: int64

The .dtype attribute returns the data type of the array elements

#### **Specifying Data Type while creating an array**

You can explicitly define the data type when creating an array

arr\_float = np.array([1,2,3], dtype=np.float32)

print (arr\_float) # Output: float32

#### **Common NumPy Data Types**

|  |  |
| --- | --- |
| **Data Type** | **Description** |
| int8, int16, int32, int64 | Integer types of different sizes |
| float16, float32, float64 | Floating-point numbers |
| bool\_ | Boolean values (True or False) |
| complex64, complex128 | Complex numbers |
| str\_, unicode\_ | String data types |

#### **Changing Data Type**

You can convert an array to a different data type using .astype()

arr = np.array([1.5,2.7,5.6])

arr\_int = arr.astype(np.int32)

print(arr\_int) # Output: [1,2,3]

print (arr\_int.dtype) # Output: int32

This converts floating point to integers, truncating the decimal part.

#### **Memory Efficiency with NumPy Data Types**

NumPy optimizes memory usage compared to Python lists. For example:

import sys

py\_list = [1, 2, 3, 4, 5]

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

print(sys.getsizeof(py\_list)) # Memory used by Python list

print(np\_array.nbytes) # Memory used by NumPy array

NumPy arrays are **more memory-efficient** than Python lists

#### **Structured Data Types**

NumPy allows defining **structured arrays** with multiple fields.

dt = np.dtype([('name', np.str\_, 16), ('age', np.int32)])

arr = np.array([('Alice', 25), ('Bob', 30)], dtype=dt)

print(arr['name']) # Output: ['Alice' 'Bob']

print(arr['age']) # Output: [25 30]

This is useful for handling tabular data

### **Changing array shape**

NumPy provides several ways to reshape arrays, allowing you to modify their dimensions without changing the underlying data. This is useful for organizing data efficiently in scientific computing and machine learning.

#### **.reshape()**

The .reshape() method allows you to change the shape of any array while maintaining the same number of elements.

import numpy as np

# Creating a 1D Array

arr = np.arange(12) Output: [0,1,2,3,4,5,6,7,8,9,10,11]

# Reshaping into 3 x 4 matrix

reshaped\_arr = arr.reshape(3,4)

print (reshaped\_arr)

Output:

[[ 0 1 2 3]

[ 4 5 6 7]

[ 8 9 10 11]]

* The total number of elements **must remain the same** (12 elements in both cases).
* .reshape(new\_shape) does **not** modify the original array; it returns a new view.

#### **.ravel() and .flatten()**

These methods convert mult-dimensional array into 1D arrays

# Flattening the reshaped array

flat\_arr = reshaped\_arr.ravel()

print(flat\_arr) # Output: [0 1 2 3 4 5 6 7 8 9 10 11]

* .ravel() returns a **view** (changes affect the original array).
* .flatten() returns a **copy** (changes do not affect the original array).

#### **.transpose()**

Swaps rows and column in a matrix

transposed\_arr = reshaped\_arr.transpose()

print(transposed\_arr)

Output:

[[ 0 4 8]

[ 1 5 9]

[ 2 6 10]

[ 3 7 11]]

* .transpose() flips the shape ((3,4) → (4,3)).

#### **.resize()**

Unlike .reshape(), .resize() modifies the original array

arr.resize(4,3)

print(arr)

Output:

[[ 0 1 2]

[ 3 4 5]

[ 6 7 8]

[ 9 10 11]]

* .resize(new\_shape) **changes the array in place**.
* If the new shape is larger, NumPy fills extra spaces with 0.

#### Using -1 for Automatic Reshaping

NumPy can infer one dimension automatically

When you specify -1 as a dimension, NumPy automatically calculates that dimension based on the total number of elements

It ensures that the total number of elements remains the same after reshaping

arr = np.arange(12)

# Reshaping into 2D with 6 columns

auto\_reshaped = arr.reshape(-1,6)

print(auto\_reshaped)

Output:

[[ 0 1 2 3 4 5]

[ 6 7 8 9 10 11]]

Here, NumPy automatically determines that **2 rows** are needed (12 / 6 = 2).

arr = np.arange(20) # 20 elements

reshaped\_array = arr.reshape(4, -1) # NumPy figures out the missing dimension

print(reshaped\_array.shape) # Output: (4, 5)

* Since arr has **20 elements**, and we specify 4 rows, NumPy infers 5 columns.

### **Indexing and slicing arrays**

NumPy provides powerful tools for indexing and slicing arrays, allowing efficient manipulation of data.

#### **Indexing in NumPy**

Indexing refers to accessing elements of an array using their position

##### **1D Array Indexing**

import numpy as np

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

print (arr[0]) # Output: 1

print (arr[-1]) #Output: 5

##### **2D Array Indexing**

import numpy as np

arr2d = np.array([[1,2,3],[4,5,6],[7,8,9]])

print (arr2d[0,1]) # Element at row index 0 and column index 1: Output: 2

print (arr2d[2,2]) # Element at row index 2 and column index 2:Output: 9

##### **Boolean Indexing**

arr =np.array([10,20,30,40,50,60])

print(arr[arr>25]) # Returns elements greater than 25: [30,40,50,60]

### **Slicing in NumPy**

Slicing allows extracting portions of an array

#### **1D Array Slicing**

arr = np.array([10,20,30,40,50,60,70,80])

print(arr[1:4]) # Elements from index to 3 Output: [20,30,40]

print(arr[:3]) # First three elements Output: [10,20,30]

print (arr[::2]) # Every second element Output: [10,30,50,70]

#### **2D Array Slicing**

arr2d = np.array([[1,2,3], [4,5,6], [7,8,9])

print(arr2d[:2, :2]) # First two rows and first two columns Output:[ [1,2],[4,5],[7,8]]

print(arr2d[:,1]): # All rows second column Output: [[2 ],[5],[6]]

print(arr2d[1:,:]) # All rows from first row and all columns Output: [[4,5,6],[7,8,9]]

#### **Step Slicing**

arr = np.array([10,20,30,40,50,60])

print(arr[1:5:2]) # Elements from index 1 to 4 with step 2: Output: [20,40]

#### **Modifying Slices**

Since NumPy Slices are views, modifying them affects the original array.

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

slice\_arr = arr[1:4]

slice\_arr[:] = 99 # Modifies original array

print(arr) #Output: [1,99,99,99,5]

To avoid modifying the original array, use .copy()

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

slice\_arr = arr[1:4].copy()

slice\_arr = 99

print(arr) # Output: [1,2,3,4,5]

#### **Fancy Indexing**

Fancy Indexing allows selecting multiple elements at once.

arr = np.array([10,20,30,40,50])

indices = [0,2,4]

print(arr[indices]) # Output: [10,30,50]

arr2d = np.array([[1,2,3],[4,5,6],[7,8,9]])

rows = [0,2]

cols = [1,2]

print(arr2d[rows, cols]) # Output: [2,9]

### **Copying vs View (memory optimization)**

In NumPy, copying and viewing refer to different was of handling array data:

#### **Copy**

A copy creates a new array with duplicated data

Changes made to the copy do not affect the original array

Copies are memory-intensive but necessary when independent modifications are required

You can explicitly create a copy using ndarray.copy()

Example:

import numpy as np

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

x = arr.copy()

arr[0] = 100

print(x) # Output: [1,2,3,4,5]

print(arr) # Output: [100,2,3,4,5]

#### **View**

A view provides a different perspective on the same data buffer.

Changes made to a view reflect in the original array.

Views are efficient as they avoid unnecessary memory duplication

You can create a view using ndarray.view()

Example

import numpy as np

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

x = arr.view()

arr[0] = 5555

print (arr) # Output: [5555,2,3,4,5]

print(x) # Output: [5555,2,3,4,5]

### **Broadcasting in NumPy**

Broadcasting in NumPy refers to the ability of NumPy to perform element-wise operations on arrays of different shapes without the need for explicit looping. This feature allows for efficient computation and memory usage, making operations faster compared to traditional looping approaches.

#### **How Broadcasting works (Step-by-Step Explanation)**

1. **Understanding the Rules:** Broadcasting follows specific rules to match arrays of different shapes:

* The arrays are aligned starting from the rightmost dimension
* If dimensions differ, the smaller array is expanded by repeating its values.
* A dimension with size 1 can be stretched to match the corresponding dimension of the larger array

1. **Compatible Shapes:** Two arrays can be broadcast together only if:

* They have the same shape
* The shape of one of the arrays is 1 in a dimension that differs

1. **Applying Broadcasting:** Instead of manually reshaping and repeating arrays NumPy automatically extends the smaller array to match the larger array during operations.

Example:

import numpy as np

# Define a 2D array (3x3)

A = np.array([

[1,2,3],

[4,5,6],

[7,8,9]

])

# Define a 1D Array (1x3)

B = np.array([10,20,30])

# Perform element-wise addition

result = A + B

print (result)

#### **Advantages of Broadcasting**

Efficient – No explicit looping needed, making operations faster

Memory Optimized: Avoids unnecessary copying of data.

Intuitive Syntax: Makes mathematical operations simpler and more readable

#### **Examples on Variety of Broadcasting**

1. **Broadcasting with Scalar Addition**

A scalar (single value) can be broadcast across a NumPy array.

python

import numpy as np

# Define a 2D array

A = np.array([[1, 2, 3],

[4, 5, 6]])

# Scalar addition

result = A + 10

print(result)

**Output:**

[[11 12 13]

[14 15 16]]

**Explanation:** The scalar 10 is broadcast to every element in the array.

1. **Broadcasting with 1D and 2D Arrays**

python

B = np.array([[10], [20]]) # Shape (2,1)

C = np.array([1, 2, 3]) # Shape (1,3)

result = B + C # Broadcasting applied

print(result)

**Output:**

[[11 12 13]

[21 22 23]]

**Explanation:** B (2×1) expands to match C (1×3), resulting in a (2×3) matrix.

1. **Broadcasting with Element-wise Multiplication**

python

X = np.array([[1, 2, 3], [4, 5, 6]]) # Shape (2,3)

Y = np.array([10, 20, 30]) # Shape (3,)

result = X \* Y # Broadcasting applied

print(result)

**Output:**

[[10 40 90]

[40 100 180]]

**Explanation:** Y is expanded vertically to match X.

1. **Broadcasting Across Higher Dimensions**

python

D = np.ones((3, 1, 2)) # Shape (3,1,2)

E = np.array([5, 10]) # Shape (2,)

result = D \* E # Broadcasting applied

print(result)

**Output Shape:** (3,1,2)

[[[ 5 10]]

[[ 5 10]]

[[ 5 10]]]

**Explanation:** E expands to match the second dimension of D.

### **Aggregation Functions in NumPy**

Aggregation functions in NumPy allow you to perform statistical operations on array efficiently. These functions help summarize data by computing values such as sum, mean, max, min, standard deviation, variance and more

1. **Sum (np.sum)**

Computes the sum of all elements in an array

import numpy as np

A = np.array([

[1,2,3],

[4,5,6]

])

# Compute sum of all elements

total = np.sum(A)

# Compute sum along rows (axis = 1)

row\_sum = np.sum(A, axis=1)

# Compute sum along columns (axis = 0)

col\_sum = np.sum(A, axis=0)

print (f"Array A contains:\n {A}")

print (f"Total Sum: {total}")

print (f"Row Sum: {row\_sum}")

print (f"Col Sum: {col\_sum}")

1. **Mean (np.mean)**
   1. import numpy as np
   2. A = np.array([
   3. [1,2,3],
   4. [4,5,6]
   5. ])
   6. # Compute sum of all elements
   7. total\_mean = np.mean(A)
   8. # Compute sum along rows (axis = 1)
   9. row\_mean = np.mean(A, axis=1)
   10. # Compute sum along columns (axis = 0)
   11. col\_mean= np.mean(A, axis=0)
   12. print (f"Array A contains:\n {A}")
   13. print (f"Total Mean: {total\_mean}")
   14. print (f"Row Mean: {row\_mean}")
   15. print (f"Col Mean: {col\_mean}")
2. Maximum (np.max) and Minimum (np.min)
   1. import numpy as np
   2. A = np.array([
   3. [1,2,3],
   4. [4,5,6]
   5. ])
   6. # Compute max & min of all elements
   7. total\_max = np.max(A)
   8. total\_min = np.min(A)
   9. # Compute max & min along rows (axis = 1)
   10. row\_max = np.max(A, axis=1)
   11. row\_min = np.min(A, axis=1)
   12. # Compute max & min along columns (axis = 0)
   13. col\_max= np.max(A, axis=0)
   14. col\_min= np.min(A, axis=0)
   15. print (f"Array A contains:\n {A}")
   16. print (f"Total Max: {total\_max}")
   17. print (f"Total Min: {total\_min}")
   18. print (f"Row Max: {row\_max}")
   19. print (f"Row Min: {row\_min}")
   20. print (f"Col Max: {col\_max}")
   21. print (f"Col Min: {col\_min}")
3. **Standard Deviation (np.std) and Variance (np.var)**
   1. import numpy as np
   2. A = np.array([
   3. [1,2,3],
   4. [4,5,6]
   5. ])
   6. # Compute Std & Var of all elements
   7. total\_std = np.std(A)
   8. total\_var = np.var(A)
   9. # Compute std & var along rows (axis = 1)
   10. row\_std = np.std(A, axis=1)
   11. row\_var = np.var(A, axis=1)
   12. # Compute std & var along columns (axis = 0)
   13. col\_std= np.std(A, axis=0)
   14. col\_var= np.var(A, axis=0)
   15. print (f"Array A contains:\n {A}")
   16. print (f"Total STD: {total\_std}")
   17. print (f"Total VAR: {total\_var}")
   18. print (f"Row STD: {row\_std}")
   19. print (f"Row VAR: {row\_var}")
   20. print (f"Col STD: {col\_std}")
   21. print (f"Col VAR: {col\_var}")
4. **Median (np.median)**
   1. import numpy as np
   2. A = np.array([
   3. [1,2,3],
   4. [4,5,6]
   5. ])
   6. # Compute median of all elements
   7. total\_median = np.median(A)
   8. # Compute median along rows (axis = 1)
   9. row\_median = np.median(A, axis=1)
   10. # Compute median along columns (axis = 0)
   11. col\_median= np.median(A, axis=0)
   12. print (f"Array A contains:\n {A}")
   13. print (f"Total Median: {total\_median}")
   14. print (f"Row Median: {row\_median}")
   15. print (f"Col Median: {col\_median}")
5. **Cumulative Sum (np.cumsum)**
   1. A = np.array([1, 2, 3, 4, 5])
   2. # Compute Cumulative Sum
   3. total\_cumsum = np.cumsum(A)
   4. print (f"Values in Array A: {A}")
   5. print (f"Cumulative Sum of A: {total\_cumsum}")
   6. B = np.array([
   7. [1,2,3],
   8. [4,5,6]
   9. ])
   10. # Compute Cumulative Sum
   11. total\_cumsum = np.cumsum(B)
   12. # Row wise Cumulative Sum
   13. row\_cumsum = np.cumsum(B, axis=1)
   14. # Col wise Cumulative Sum
   15. col\_cumsum = np.cumsum(B, axis=0)
   16. print (f"Values in Array B: \n{B}")
   17. print (f"Cumulative Sum of B: \n{total\_cumsum}")
   18. print (f"Cumulative Row Sum of B: \n{row\_cumsum}")
   19. print (f"Cumulative Col Sum of B: \n{col\_cumsum}")
6. **Cumulative Product (np.cumprod)**
   1. X = np.array([1, 2, 3, 4, 5])
   2. # Compute Cumulative Sum
   3. total\_cumprod = np.cumprod(A)
   4. print (f"Values in Array X: {X}")
   5. print (f"Cumulative Product of X: {total\_cumprod}")
   6. P = np.array([
   7. [1,2,3],
   8. [4,5,6]
   9. ])
   10. # Compute Cumulative Product
   11. total\_cumprod = np.cumprod(P)
   12. # Row wise Cumulative Product
   13. row\_cumprod = np.cumprod(P, axis=1)
   14. # Col wise Cumulative Product
   15. col\_cumprod = np.cumprod(P, axis=0)
   16. print (f"Values in Array P: \n{P}")
   17. print (f"Cumulative Product of P: \n{total\_cumprod}")
   18. print (f"Cumulative Row Product of P: \n{row\_cumprod}")
   19. print (f"Cumulative Col Product of P: \n{col\_cumprod}")

### **Boolean Masking and Filtering in NumPy**

Boolean masking and filtering in NumPy allows you to select, modify or filter elements in an array based on a specific conditions. This technique is widely used in data analysis, machine learning and scientific computing.

#### **Understanding Boolean Masks**

A Boolean mask is a NumPy array containing True/False values that correspond to each element in another array. The True values indicate which elements should be selected.

Example:

A = np.array([1,5,10,25,190,50, 71,20])

A\_bool = A > 50

print (f"Elements for A : {A}")

print (f"Masking elements of A that are > 50 {A\_bool}")

#### **Filtering Elements Using Boolean Mask**

You can use the Boolean mask to filter elements from the array

Example:

A = np.array([1,5,10,25,190,50, 71,20])

A\_bool = A > 50

print (f"Elements for A : {A}")

print (f"Filtering elements of A that are > 50 : {A[A % 2 != 0]}")

### **Sorting and Filtering Arrays in NumPy**

NumPy provides the numpy.sort() function to sort elements in an array.

**Example:**

**Basic Sorting**

import numpy as np

# Create an array

arr = np.array([4,61,2,51,5])

# Sorting the array

sorted\_arr = np.sort(arr)

print(sorted\_arr)

The np.sort() functions returns a new sorted array without modifying the original one.

**Sorting a Multi-Dimensional Array**

arr\_2D = np.array([[3,2,1], [6,5,4]])

# Sorting along rows (default axis=1)

sorted\_rows = np.sort(arr\_2D, axis = 1)

print (sorted\_rows)

# Sortnig along columns

sorted\_columns = np.sort(arr\_2D, axis = 0)

print(sorted\_columns)

axis = 1 sorts each row individually, and axis = 0 sorts each column.

### **Searching for elements in Arrays using NumPy**

Searching involves finding specific values or indices.

#### **Finding an Element’s Index**

arr = np.array([10, 20, 30, 40, 50])

# Find index of element 30

index = np.where(arr == 30)

print(index) # Output: (array([2]),)

The np.where() function returns the index where the element is found.

#### **Searching with Conditions**

arr = np.array([10, 25, 30, 45, 50])

# Find all elements greater than 25

indices = np.where(arr > 25)

print(arr[indices]) # Output: [30 45 50]

You can use conditions to filter elements.

#### **Checking for an Element's Existence**

arr = np.array([5, 10, 15, 20])

# Check if 15 is in the array

exists = np.isin(arr, 15)

print(exists) # Output: [False False True False]

np.isin() checks for the presence of a value.

### **Vectorized Computations & Efficiency Improvements using NumPy**

Vectorized computations in NumPy allow operations to be performed on entire arrays without explicit loops, making them significantly faster and more efficient. This is because NumPy leverages optimized C-based implementations under the hood.

#### **Why Vectorization Matters?**

**Improves Performance:** Eliminates slow Python loops

**Simplifies Code:** Makes it cleaner and more readable

**Handles Large Data Efficiently:** Ideal for big datasets in Machine Learning and Scientific Computing.

##### **Example-1: Element-wise operations without loops**

import numpy as np

# Create an array

arr = np.array([2,4,6,8,10])

# Vectorized addition

result = arr + 2

print (result) # Output: [4,6,8,10,12]

##### **Example-2: Vectorized Matrix Multiplication**

import numpy as np

# Define two matrices

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

B = np.array([[5,6],[7,8]])

# Matrix multiplication using np.dot()

result = np.dot(A, B)

print(result)

##### **Example-3: Logical Operations on Array**

arr = np.array([10,20,30])

# Check which elements are greater than 15

result = arr > 15

print (result) # Output: [False, True, True]

##### **Example-4: Applying Custom Functions with np.vectorize()**

If you have a custom function, np.vectorize() helps apply it efficiently

import numpy as np

# Define a Custom Function

def custom\_fun(x):

return x\*\*2 + 2\*x + 1

# Create and Array

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

# Apply function using np.vectorize()

vectorized\_func = np.vectorize(custom\_func)

result = vectorized\_fun(arr)

print(result) # Output: [4, 9, 16, 25]

**Summary**

* **Vectorization eliminates loops**, making computations faster.
* **NumPy operations are optimized** for performance.
* **Matrix operations, logical conditions, and custom functions** can be vectorized.

### **Handling missing values using NumPy**

#### **Understanding NaN and inf in NumPy**

* NaN (Not a Number): Represents undefined or missing values, often encountered in datasets with incomplete records
* Inf (infinity): Represents infinite values, usually resulting from division by zero or mathematical operations with very large numbers

#### **Detecting NaN and inf in an array**

NumPy provides np.isnan() and np.isinf() to identify these values in an array.

**Example:**

import numpy as np

# Creating an array with NaN and inf values

arr = np.array([1,2,np.nan,4, np.inf, 6])

# Detect NaN values

nan\_mask = np.isnan(arr)

print(“NaN mask:”, nan\_mask)

# Detect inf values

Inf\_mask = np.isinf(arr)

print(“Inf mask:”, inf\_mask)

# Detect both NaN and inf values

nan\_inf\_mask = np.isnan(arr) | np.isinf(arr)

print("NaN or Inf mask:", nan\_inf\_mask)

**Output:**

NaN mask: [False False True False False False]

Inf mask: [False False False False True False]

NaN or Inf mask: [False False True False True False]

#### **Replacing NaN and inf values**

You can replace missing values with meaningful values like zeros, mean, or median using np.nan\_to\_num()

clean\_arr = np.nan\_to\_num(arr, nan=0, posinf=10, neginf=-10)

print("Replaced array:", clean\_arr)

Replaced array: [ 1. 2. 0. 4. 10. 6.]

#### **Removing NaN and inf values**

Filtering out invalid values can be helpful when you want a clean dataset.

**Example: Removing NaN and inf**

filtered\_arr = arr[~nan\_inf\_mask] # Keep only valid numbers

print("Filtered array:", filtered\_arr)

**Output:**

Filtered array: [1. 2. 4. 6.]

#### **Handling NaN values in calculations**

Many NumPy functions have special versions that ignore NaN values.

**For example:**

* np.nanmean(), np.nanmedian(), np.nanmax() → Compute statistics **ignoring NaN**.
* np.nansum() → Sum **excluding NaN**.

**Example: Computing mean while ignoring NaN**

mean\_val = np.nanmean(arr)

print("Mean ignoring NaN:", mean\_val)

#### **NumPy Structured Array**

NumPy structured arrays are a powerful feature that allow you to create arrays with named fields, similar to columns in a table or attributes in a struct. They are useful when you want to store heterogeneous data types in a single array, making NumPy function like a lightweight database.

**Step-1: Import NumPy**

import numpy as np

**Step-2: Define a Structured Data Type**

Structured arrays allow you to define custom data types with multiple fields. Here is an example of defining a structured data type for a student record.

student\_dtype = np.dtype([

(‘name’, ‘U20’), # Unicode string with max length 20

(‘age’, ‘i4’), # 4-byte integer

(‘grade’, ‘f4’) # 4-byte integer

])

**Step-3: Create a Structured Array**

Now, lets create a NumPy array with this structure data type

students = np.array([

(‘Alice’, 20, 85.5),

(‘Bob’, 22, 90.2),

(‘Charlie’, 21, 78.8)

], dtype=student\_dtype)

**Step-4: Accessing Data**

You can access individual elements or specific fields easily

print(students[0]) # Access first student’s record

print(students[‘name’]) # Get all names

print(students[‘age’][1]) # Get the age of second student (Bob)

**Step-5: Modifying Values**

Structure arrays allow modifications:

students[1][‘grade’] = 95.0 # Update Bob’s grade

print(students)

**Example-2: Employee Records**

Let’s create a structured array to store employee information (Name, ID,Salary)

employee\_dtype = np.dtype([

(‘name’, ‘u20’),

(‘id’, ‘i4’),

(‘salary’, ‘f8’)

])

employees = np.array([

(‘John’, 101, 50000.0),

(‘Luke’, 102, 60000.0),

(‘Mark’, 103, 70000.0)

], dtype=employee\_dtype)

**Access employee Salaries**

print(employees['salary'])

Sort by Salary

sorted\_employees = np.sort(employees, order=’salary’)

print(sorted\_employees)

##### **Why use Structured Arrays?**

* Efficient storage of heterogeneous data types
* Easy field-based access
* Supports sorting and filtering based on specific fields