# **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]