NumPy Cheat Sheet for Quick Basic Fundamentals

Welcome to your comprehensive quick reference guide for **NumPy**! This cheat sheet covers the fundamental concepts you need to get started with NumPy for numerical computing and data manipulation, enhanced with extended examples for better understanding and quick revision.

Table of Contents

- 1. Introduction
- 2. Installation and Setup
- 3. Importing NumPy
- 4. NumPy Data Structures
- 5. Creating Arrays
- 6. Array Attributes
- 7. Basic Operations
- 8. Indexing and Slicing
- 9. Reshaping Arrays
- 10. Mathematical Functions
- 11. Broadcasting
- 12. Linear Algebra
- 13. Random Module
- 14. Saving and Loading Arrays
- 15. Advanced Topics
- 16. Best Practices
- 17. Tips for Beginners
- 18. Additional Resources

Introduction

Overview: NumPy (Numerical Python) is a foundational library for numerical computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of high-level mathematical functions to operate on these arrays efficiently.

Key Features:

- N-dimensional array object (ndarray)
- · Broadcasting capabilities
- Vectorized operations for performance
- Comprehensive mathematical functions
- Integration with other libraries (e.g., Pandas, Matplotlib)

Installation and Setup

Overview: Set up NumPy on your local machine to start numerical computations seamlessly.

Using pip

If you have Python installed, you can install NumPy using pip.

```
pip install numpy
```

Using Anaconda

Anaconda distribution comes with NumPy pre-installed.

1. Download Anaconda:

• Visit Anaconda Downloads and download the installer for your operating system.

2. Install Anaconda:

• Follow the installation instructions specific to your OS.

3. Verify Installation:

```
import numpy as np
print(np.__version__)
```

Importing NumPy

```
import numpy as np
```

NumPy Data Structures

Overview: Understand the primary data structure in NumPy: the **ndarray**.

ndarray

A homogeneous n-dimensional array of fixed-size items. All elements in an ndarray must be of the same data type.

```
import numpy as np

# Creating a 1D array
arr1 = np.array([1, 2, 3, 4, 5])
print(arr1) # Output: [1 2 3 4 5]

# Creating a 2D array
arr2 = np.array([[1, 2, 3], [4, 5, 6]])
print(arr2)
```

```
# Output:
# [[1 2 3]
# [4 5 6]]
```

Creating Arrays

Overview: Learn various methods to create NumPy arrays.

From Lists or Tuples

```
import numpy as np

# From a list
list1 = [1, 2, 3, 4]
arr = np.array(list1)
print(arr) # Output: [1 2 3 4]

# From a tuple
tuple1 = (5, 6, 7, 8)
arr = np.array(tuple1)
print(arr) # Output: [5 6 7 8]
```

Using Built-in Functions

• np.zeros(): Create an array filled with zeros.

```
zeros = np.zeros((2, 3))
print(zeros)
# Output:
# [[0. 0. 0.]
# [0. 0. 0.]]
```

np.ones(): Create an array filled with ones.

```
ones = np.ones((3, 2), dtype=int)
print(ones)
# Output:
# [[1 1]
# [1 1]
# [1 1]]
```

np.full(): Create an array filled with a specified value.

```
full = np.full((2, 2), 7)
print(full)
# Output:
# [[7 7]
# [7 7]]
```

• np.eye(): Create an identity matrix.

```
identity = np.eye(3)
print(identity)
# Output:
# [[1. 0. 0.]
# [0. 1. 0.]
# [0. 0. 1.]]
```

• np.arange(): Create an array with a range of values.

```
range_arr = np.arange(0, 10, 2)
print(range_arr) # Output: [0 2 4 6 8]
```

• np.linspace(): Create an array with linearly spaced values.

```
linspace_arr = np.linspace(0, 1, 5)
print(linspace_arr) # Output: [0. 0.25 0.5 0.75 1. ]
```

np.random.rand(): Create an array with random values between 0 and 1.

```
random_arr = np.random.rand(2, 3)
print(random_arr)
# Output: Random values, e.g.,
# [[0.5488135  0.71518937  0.60276338]
# [0.54488318  0.4236548  0.64589411]]
```

From Existing Data

- np.frombuffer()
- np.fromfile()
- np.fromiter()

Array Attributes

Overview: Access essential properties of NumPy arrays.

```
import numpy as np

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

print(arr.ndim)  # Number of dimensions: 2

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

print(arr.size)  # Total number of elements: 6

print(arr.dtype)  # Data type: int64 (varies based on system)

print(arr.itemsize)  # Size of each element in bytes: 8

print(arr.data)  # Buffer containing the actual elements
```

Basic Operations

Overview: Perform fundamental mathematical and logical operations on arrays.

Arithmetic Operations

```
import numpy as np

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

print(a + b)  # Output: [5 7 9]
print(a - b)  # Output: [-3 -3 -3]
print(a * b)  # Output: [ 4 10 18]
print(a / b)  # Output: [ 0.25 0.4 0.5 ]
print(a ** 2)  # Output: [1 4 9]
print(a % 2)  # Output: [1 0 1]
```

Comparison Operations

```
import numpy as np

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

print(a > b) # Output: [False False True]
print(a == b) # Output: [False True False]
print(a != b) # Output: [ True False True]
```

Logical Operations

```
import numpy as np

a = np.array([True, False, True])
b = np.array([False, False, True])

print(np.logical_and(a, b)) # Output: [False False True]
print(np.logical_or(a, b)) # Output: [True False True]
print(np.logical_not(a)) # Output: [False True False]
```

Universal Functions (ufuncs)

NumPy provides a suite of vectorized functions that operate element-wise on arrays.

Indexing and Slicing

Overview: Access and modify specific elements or subsets of arrays efficiently.

Basic Indexing

```
import numpy as np
arr = np.array([10, 20, 30, 40, 50])
print(arr[0]) # Output: 10
print(arr[3]) # Output: 40
```

Negative Indexing

```
import numpy as np
arr = np.array([10, 20, 30, 40, 50])
print(arr[-1]) # Output: 50
print(arr[-3]) # Output: 30
```

Slicing

```
import numpy as np

arr = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

print(arr[2:5])  # Output: [2 3 4]
print(arr[:4])  # Output: [0 1 2 3]
print(arr[6:])  # Output: [6 7 8 9]
print(arr[::2])  # Output: [0 2 4 6 8]
print(arr[::-1])  # Output: [9 8 7 6 5 4 3 2 1 0]
```

Multi-dimensional Indexing

Boolean Indexing

```
import numpy as np
arr = np.array([10, 15, 20, 25, 30])

# Select elements greater than 20
filtered = arr[arr > 20]
print(filtered) # Output: [25 30]
```

Fancy Indexing

```
import numpy as np

arr = np.array([100, 200, 300, 400, 500])

# Select elements at indices 0, 2, and 4
indices = [0, 2, 4]
selected = arr[indices]
print(selected) # Output: [100 300 500]
```

Reshaping Arrays

Overview: Change the shape of arrays without changing their data.

reshape()

```
import numpy as np

arr = np.arange(12) # [0, 1, 2, ..., 11]

# Reshape to 3x4
reshaped = arr.reshape((3, 4))
print(reshaped)
# Output:
# [[ 0  1  2  3]
# [ 4  5  6  7]
# [ 8  9 10 11]]
```

Flattening Arrays

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

flat = arr.flatten()
print(flat) # Output: [1 2 3 4 5 6]
```

Transposing Arrays

```
import numpy as np

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

transposed = arr.T
print(transposed)
# Output:
# [[1 4]
# [2 5]
# [3 6]]
```

Resizing Arrays

```
import numpy as np

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

# Resize to 2x2
resized = arr.reshape((2, 2))
print(resized)
# Output:
# [[1 2]
# [3 4]]
```

Mathematical Functions

Overview: Perform mathematical operations on arrays.

Basic Math Operations

```
import numpy as np

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

print(np.add(arr, 10))  # Output: [11 12 13 14 15]
print(np.subtract(arr, 2)) # Output: [-1 0 1 2 3]
print(np.multiply(arr, 3)) # Output: [ 3 6 9 12 15]
print(np.divide(arr, 2)) # Output: [0.5 1. 1.5 2. 2.5]
```

Statistical Functions

```
import numpy as np
arr = np.array([1, 2, 3, 4, 5])
print(np.mean(arr))
                    # Output: 3.0
print(np.median(arr)) # Output: 3.0
print(np.std(arr)) # Output: 1.4142135623730951
print(np.var(arr))
                     # Output: 2.0
print(np.min(arr))
                      # Output: 1
print(np.max(arr))
                     # Output: 5
print(np.argmin(arr)) # Output: 0
print(np.argmax(arr)) # Output: 4
print(np.sum(arr))
                   # Output: 15
print(np.cumsum(arr))  # Output: [ 1 3 6 10 15]
print(np.cumprod(arr)) # Output: [ 1  2  6  24 120]
```

Trigonometric Functions

```
import numpy as np

angles = np.array([0, np.pi/2, np.pi])

print(np.sin(angles)) # Output: [ 0.000000e+00 1.000000e+00 1.224647e-16]
print(np.cos(angles)) # Output: [ 1.0000000e+00 6.123234e-17 -1.0000000e+00]
print(np.tan(angles)) # Output: [ 0.0000000e+00 1.6331249e+16 0.0000000e+00]
```

Exponential and Logarithmic Functions

Power and Square Root Functions

```
import numpy as np
arr = np.array([1, 4, 9, 16])

print(np.sqrt(arr))  # Output: [1. 2. 3. 4.]
print(np.power(arr, 3)) # Output: [ 1 64 729 4096]
print(np.square(arr))  # Output: [ 1 16 81 256]
```

Rounding Functions

```
import numpy as np

arr = np.array([1.234, 2.345, 3.456, 4.567])

print(np.round(arr, 2))  # Output: [1.23 2.35 3.46 4.57]

print(np.floor(arr))  # Output: [1. 2. 3. 4.]

print(np.ceil(arr))  # Output: [2. 3. 4. 5.]
```

Broadcasting

Overview: Understand how NumPy handles operations on arrays of different shapes.

Basic Broadcasting Rules

- 1. If the arrays do not have the same number of dimensions, prepend the shape of the smaller array with 1s until both shapes have the same length.
- 2. Arrays are compatible in a dimension if they are equal or one of them is 1.
- 3. The arrays can be broadcast together if they are compatible in all dimensions.
- 4. The resulting array shape is the maximum size in each dimension.

Example 1: Scalar and Array

```
import numpy as np
arr = np.array([1, 2, 3])
scalar = 10

result = arr + scalar
print(result) # Output: [11 12 13]
```

Example 2: Array and Array with Different Shapes

Example 3: Higher-Dimensional Arrays

```
import numpy as np

arr1 = np.array([[1], [2], [3]]) # Shape: (3,1)
arr2 = np.array([10, 20, 30]) # Shape: (3,)

result = arr1 + arr2
print(result)
# Output:
# [[11 21 31]
# [12 22 32]
# [13 23 33]]
```

Broadcasting with Functions

Linear Algebra

Overview: Perform linear algebra operations using NumPy.

Matrix Multiplication

```
import numpy as np
A = np.array([[1, 2],
              [3, 4]])
B = np.array([[5, 6],
              [7, 8]])
# Using dot()
result = np.dot(A, B)
print(result)
# Output:
# [[19 22]
# [43 50]]
# Using @ operator (Python 3.5+)
result = A @ B
print(result)
# Output:
# [[19 22]
# [43 50]]
```

Transpose

Determinant

Inverse

Eigenvalues and Eigenvectors

```
# [[ 0.70710678 -0.70710678]
# [ 0.70710678  0.70710678]]
```

Solving Linear Systems

Random Module

Overview: Generate random numbers and perform random operations using NumPy's random module.

Basic Random Number Generation

```
import numpy as np

# Random floats between 0 and 1
rand_arr = np.random.rand(3, 2)
print(rand_arr)

# Random integers between low (inclusive) and high (exclusive)
rand_int = np.random.randint(0, 10, size=(2, 3))
print(rand_int)

# Random numbers from a normal distribution
rand_normal = np.random.randn(4)
print(rand_normal)
```

Seed for Reproducibility

```
import numpy as np

np.random.seed(42)
print(np.random.rand(3)) # Output: [0.37454012 0.95071431 0.73199394]

np.random.seed(42)
print(np.random.rand(3)) # Output: [0.37454012 0.95071431 0.73199394]
```

Shuffling Arrays

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5])
np.random.shuffle(arr)
print(arr) # Output: Shuffled array, e.g., [2 5 1 4 3]
```

Choice and Permutation

```
import numpy as np

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

# Random choice
choice = np.random.choice(arr, size=3, replace=False)
print(choice) # Output: Randomly selected elements, e.g., [20 50 10]

# Permutation
perm = np.random.permutation(arr)
print(perm) # Output: Permuted array, e.g., [30 50 10 20 40]
```

Saving and Loading Arrays

Overview: Persist and retrieve NumPy arrays using various methods.

Using np.save() and np.load()

```
import numpy as np

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

# Save array to a binary file in .npy format
np.save('my_array.npy', arr)

# Load array from the file
loaded_arr = np.load('my_array.npy')
print(loaded_arr) # Output: [1 2 3 4 5]
```

Using np.savez() and np.load()

```
import numpy as np
```

```
arr1 = np.array([1, 2, 3])
arr2 = np.array([4, 5, 6])

# Save multiple arrays into a single file
np.savez('arrays.npz', first=arr1, second=arr2)

# Load arrays from the file
data = np.load('arrays.npz')
print(data['first']) # Output: [1 2 3]
print(data['second']) # Output: [4 5 6]
```

Using np.savetxt() and np.loadtxt()

Advanced Topics

Overview: Explore more sophisticated features and optimizations in NumPy.

Structured Arrays

```
# Accessing fields
print(arr['name']) # Output: ['Alice' 'Bob']
print(arr['age']) # Output: [25 30]
```

Memory Layout

Vectorization and Performance

```
import numpy as np
import time
# Vectorized operations
arr = np.arange(1000000)
start = time.time()
vec_result = arr * 2 + 1
end = time.time()
print(f"Vectorized time: {end - start} seconds")
# Loop-based operations
arr = np.arange(1000000)
result = np.empty like(arr)
start = time.time()
for i in range(arr.size):
    result[i] = arr[i] * 2 + 1
end = time.time()
print(f"Loop-based time: {end - start} seconds")
```

Broadcasting with Multi-dimensional Arrays

```
import numpy as np

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

```
# Broadcasting addition
C = A + B
print(C)
# Output:
# [[2. 3. 4. 5.]
# [2. 3. 4. 5.]
# [2. 3. 4. 5.]]
```

Masked Arrays

```
import numpy as np
import numpy.ma as ma

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

masked_arr = ma.array(arr, mask=mask)
print(masked_arr)
# Output: [1 -- 3 -- 5]
```

Fancy Indexing and Advanced Selection

```
import numpy as np

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

# Select specific rows and columns
rows = [0, 2]
cols = [1, 3]
selected = arr[rows, cols]
print(selected) # Output: [1 13]

# Using boolean masks for advanced selection
mask = (arr % 2 == 0) & (arr > 5)
print(arr[mask])
# Output: [ 6  8 10 12 14]
```

ufuncs with Multiple Arguments

```
import numpy as np

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

# Add
print(np.add(a, b)) # Output: [5 7 9]
```

```
# Multiply
print(np.multiply(a, b)) # Output: [ 4 10 18]

# Power
print(np.power(a, b)) # Output: [ 1 32 729]
```

Best Practices

Overview: Adopt best practices to write efficient, readable, and maintainable NumPy code.

Use Vectorized Operations

• Avoid Loops: Utilize NumPy's vectorized operations for better performance.

```
import numpy as np

# Inefficient loop-based multiplication
arr = np.arange(1000)
result = np.empty_like(arr)
for i in range(len(arr)):
    result[i] = arr[i] * 2

# Efficient vectorized multiplication
result = arr * 2
```

Leverage Broadcasting

• **Simplify Operations:** Use broadcasting to perform operations on arrays of different shapes without explicit replication.

```
import numpy as np

matrix = np.ones((3, 3))
vector = np.array([1, 2, 3])
result = matrix + vector # Broadcast vector to each row
print(result)
```

Understand Memory Layout

• **Performance Optimization:** Choose the right memory layout (C-order vs. Fortran-order) based on the operations you perform.

```
import numpy as np
```

```
arr_c = np.ones((1000, 1000), order='C')
arr_f = np.ones((1000, 1000), order='F')

print(arr_c.flags['C_CONTIGUOUS']) # True
print(arr_f.flags['F_CONTIGUOUS']) # True
```

Utilize Built-in Functions

• Maximize Efficiency: Use NumPy's comprehensive set of built-in functions for common operations.

```
import numpy as np
arr = np.random.rand(1000)

# Compute mean using built-in function
mean = np.mean(arr)

# Avoid manual computation
# mean = sum(arr) / len(arr)
```

Manage Data Types

• **Memory Efficiency:** Choose appropriate data types to optimize memory usage.

```
import numpy as np
arr = np.array([1, 2, 3], dtype=np.int8)
print(arr.dtype) # Output: int8
```

Handle Exceptions Gracefully

• **Robust Code:** Anticipate and handle potential errors in array operations.

```
import numpy as np

try:
    arr = np.array([1, 2, 3])
    result = arr / 0
except FloatingPointError:
    print("Division by zero encountered!")
```

Document Your Code

• Clarity: Use comments and docstrings to explain complex operations and transformations.

```
import numpy as np

def normalize(arr):
    """
    Normalize the array to have a mean of 0 and standard deviation of 1.

Parameters:
    arr (numpy.ndarray): Input array.

Returns:
    numpy.ndarray: Normalized array.
    """
    return (arr - np.mean(arr)) / np.std(arr)
```

Profile and Optimize

Performance Tuning: Use profiling tools to identify and optimize bottlenecks.

```
import numpy as np
import time

arr = np.random.rand(1000000)

start = time.time()
result = np.sqrt(arr)
end = time.time()
print(f"Vectorized sqrt time: {end - start} seconds")

# Compare with loop-based approach if needed
```

Tips for Beginners

Overview: Enhance your learning experience with these essential tips for new NumPy users.

Start with Tutorials

- Official NumPy Tutorials: NumPy User Guide
- Interactive Platforms: DataCamp and Kaggle offer hands-on NumPy courses.

Practice with Real Datasets

- Kaggle Datasets: Explore and analyze datasets from Kaggle.
- UCI Machine Learning Repository: Access datasets at UCI Repository.

Learn by Projects

• **Mini Projects:** Start with small projects like numerical simulations, data analysis, or simple machine learning tasks.

 Capstone Projects: Gradually move to comprehensive projects integrating multiple NumPy functionalities.

Utilize Keyboard Shortcuts in IDEs

- **Boost Productivity:** Learn shortcuts in IDEs like Jupyter Notebook, VSCode, or PyCharm for efficient coding.
 - **Jupyter Notebook:** Press Tab for autocomplete, Shift + Tab for tooltips.
 - VSCode: Use Ctrl + Space for IntelliSense.

Explore NumPy Extensions

 Additional Functionality: Libraries like SciPy for advanced scientific computing or Numba for Just-In-Time (JIT) compilation.

```
import numba
from numba import jit

@jit(nopython=True)
def fast_sum(arr):
    total = 0
    for i in arr:
        total += i
    return total

arr = np.arange(1000000)
print(fast_sum(arr))
```

Participate in Communities

- Forums and Q&A: Engage with communities on Stack Overflow and Reddit.
- **GitHub Repositories:** Explore and contribute to open-source NumPy projects.

Keep Learning and Updating

- Stay Current: NumPy is actively developed. Keep an eye on updates and new features.
- **Advanced Topics:** As you progress, delve into advanced topics like memory optimization, custom dtypes, and integration with other libraries.

Additional Resources

Overview: Access a curated list of resources to deepen your understanding of NumPy.

Official Documentation

- NumPy Documentation: https://numpy.org/doc/stable/
- NumPy Tutorials: https://numpy.org/doc/stable/user/quickstart.html

• Real Python: NumPy Tutorial

• DataCamp: Introduction to NumPy

• Kaggle: NumPy Exercises

Books

• "Python for Data Analysis" by Wes McKinney: Comprehensive guide by the creator of Pandas, with extensive NumPy coverage.

• "Numerical Python" by Robert Johansson: Practical introduction to scientific computing with Python.

Online Communities

• Stack Overflow: NumPy Questions

• Reddit: r/numpy

• **GitHub:** NumPy Repository

Videos and Webinars

• YouTube Tutorials: Search for "NumPy Tutorial" for video guides.

• Official NumPy YouTube Channel: NumPy YouTube

Cheat Sheets

DataCamp NumPy Cheat Sheet: Download PDF

• Cheatography NumPy Cheat Sheet: Download PDF

Tools and Extensions

• Jupyter Notebook: Interactive environment for experimenting with NumPy.

```
pip install jupyter
jupyter notebook
```

• **Numba:** JIT compiler to accelerate NumPy operations.

```
pip install numba
```

Happy Coding!

Leverage this cheat sheet to streamline your NumPy experience. Practice regularly, explore new features, and engage with the community to become proficient in using NumPy for your numerical computing and data analysis projects.