

Python for Data Science

Module 1 : Numpy

Introduction to NumPy:-

Welcome to this module on Python for Data Science.

In this module

You will learn about the two most important and popular Python libraries for handling data: NumPy and Pandas. Let's hear from our SME, Behzad Ahmadi, who will explain the contents of the module in the following video.

In this module, you will learn about the basics of NumPy, which is the fundamental package for scientific computing in Python. NumPy consists of a robust data structure called multidimensional arrays. Pandas is another powerful Python library that provides a fast and easy-to-use data analysis platform.

Important Note:

To enhance the learning outcome, you are expected to pause the videos and code along with the instructor. You will be provided with a structured and blank IPython notebook for coding. It is a must for you to answer certain in-segment questions, as they serve the purpose of practice, and the final notebook will have practice tasks for you to solve.

You will understand the advantages of using NumPy. You will also learn how to:

- Create NumPy arrays,
- Convert lists and tuples to NumPy arrays,
- Inspect the structure and content of arrays, and
- Subset, slice, index and iterate through arrays.

Before we get into the technicalities of a NumPy array, explore its useful functions and understand how it is implemented in Python, it is crucial to understand **why NumPy is an important library for working with data.**

NumPy, an acronym for the term 'Numerical Python', is a library in Python which is used extensively for efficient mathematical computing. This library allows users to store large amounts of data using less memory and perform extensive operations efficiently. It provides optimised and simpler functionalities to perform aforementioned operations using homogenous, one-dimensional and multidimensional arrays (You will learn more about this later.).

Now, before delving deep into the concept of NumPy arrays, it is important to note that Python lists can very well perform all the actions that NumPy arrays perform; it is simply the fact that NumPy arrays are faster and more convenient than lists when it comes to extensive computations, which make them

extremely useful, especially when you are working with large amounts of data.

Basics of NumPy:-

NumPy, which stands for 'Numerical Python', is a library meant for scientific calculations. The basic data structure of NumPy is an array. A NumPy array is a collection of values stored together, similar to a list.

In the following video, you will learn about the difference between lists and NumPy arrays.

This video mentioned two different advantages that NumPy arrays have over lists. These include:

1. Ability to operate on individual elements in the array without using loops or list comprehension
2. Speed of execution

The demonstration in the video above did not cover the aspect of speed, so for now, you can assume that a NumPy array is faster than a list. Later in this session, you will be able to take a look at a detailed demonstration to compare the speed of NumPy arrays.

You can download the IPython notebook used in the lecture from the link given below. As mentioned in the introduction, you are expected to code along with the instructor in the notebook.

Module 2 : Pandas