

2. Target Skills and Learning Outcomes

The skills that should be acquired while taking this course are summarized in the following table

Skills	Learning outcomes
<i>Numpy basics</i>	<i>How to load dataset on numpy array How to make basic operations on data by a numpy array</i>
<i>Numpy advanced</i>	<i>How to aggregate and combine multiple arrays</i>
<i>Scipy overview</i>	<i>How to make some linear algebra computations</i>
<i>Matplotlib and seaborn</i>	<i>How to draw and customize main types of plots: line, scatter, histogram...</i>
<i>Practical session : solve linear regression from scratch</i>	<i>How to use only numpy and scipy to solve linear regression in different way</i>
<i>Pandas basics</i>	How to create dataframe from a dataset and make some basic operations
<i>Pandas advanced</i>	How to use the groupby on dataframe How to concatenate multiple dataframe How to manipulate time series data
<i>Practical session: Exploratory data analysis (EDA)</i>	How to analyze and extract valuable information of a real-world dataset

session1

Session milestones:

The following steps must be covered in the following order

- Explain what for we use NumPy: the advantages of NumPy array over pure python data structure
- Create a 1D-Array and 2D-Array from common NumPy functions
- Manipulate different the standard data types of NumPy
- See the most basics attribute of array
- See the most basics universal functions of NumPy
- Transform arrays: Boolean indexing of arrays, slice arrays, reshape arrays

Homework

Exercises on NumPy

2. Session 2: Intermediate NumPy and SciPy overview

In this session, students will go further in understanding of NumPy. The students must understand the rules of broadcasting and its traps. Students must what is an axis and how to sort and aggregate an array on different axis.

On the other hand, student will see an overview of important packages of SciPy library. The student will discover user-friendly and efficient numerical routines, such as routines for numerical library contains integration, interpolation, optimization, linear algebra, and statistics.

Pre-Requisites

- Session 1

Session milestones

The following steps must be covered in the following order

- Rules of broadcasting and its traps.
- Axis of rows and axis of columns of 2D array
- Aggregate data depending on different axis
- Aggregate data depending on different axis
- How to read a simple dataset with Numpy
- How to write to simple dataset with Numpy

Homework

Compute some basic statistics on a dataset.

Skills

NumPy advanced

3. Session 3: SciPy overview

In this session, student will see an overview of important packages of SciPy library. The student will discover user-friendly and efficient numerical routines, such as routines for numerical library contains integration, interpolation, optimization, linear algebra, and statistics.

Pre-Requisites

- Session 1

Session milestones

The following steps must be covered in the following order:

- Usage of linear algebra SciPy package
- Basic usage of interpolation SciPy package
- Basic usage of optimization SciPy package

Homework

Exercises with Scipy

4. Session 4: Matplotlib and seaborn

The students will plot data with matplotlib by looking at the theory and following along with practical examples. To address some limits of Matplotlib, student will see Seaborn which provides a high-level interface for drawing attractive and informative statistical graphics.

Pre-Requisites

- NumPy

Session milestones

- Grasp the key concepts in the design of matplotlib
- Understand subplots
- Visualize arrays with matplotlib
- Draw simple line Plots
- Draw scatter Plots
- Histograms, Binnings, and Density
- Customizing Plot Legends
- Seaborn Versus Matplotlib

6. Session 6: Pandas Overview

The goal of this session is to introduce the library Pandas which is a newer package built on top of NumPy and provides an efficient implementation of a DataFrame and Series. These two data structures of Pandas provide various intuitive helper functions which is extremely useful for extracting valuable information.

Pre-Requisites

Matplotlib, numpy

Session milestones

- What a Pandas Series is and how to create one
- What a Pandas DataFrame is and how to create one
- How to read and write data to and from files by a Pandas Dataframe
- How to access, modify, add, sort, filter, and delete data
- How to handle missing values
- How to quickly visualize data
- How to compute some basic statistics

Homework

Some basic transformations of a dataset will be achieved.

7. Session 7 : Advanced Pandas

In this session, students will go further with pandas by learning the group-by operation. Student will also learn how to concatenate different dataframe following different axis. Finally, students will also learn how to handle time series data.

Pre-Requisites

Pandas beginner

Session milestones

- How to use Pandas Group By operations on real-world data
- How the split-apply-combine chain of operations works
- How to concatenate for combining DataFrames across rows or columns
- How to work with time-series data
-

Homework

Exercises on Pandas

Skills

Advanced transformation on dataset with pandas

