Numpy

This library provides the ndarray object for efficient storage and manipulation of dense data arrays in Python.

- There are different libraries in python like numpy, pandas, seaborn and matplot lib.
- Numpy stands for num(numerical) py(python).

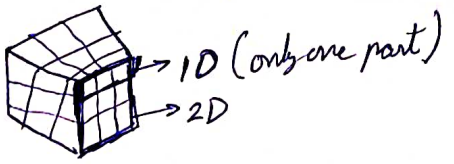
Numpy is a numerical python package (or) a library. It does.

Linear Algebra
Trignometric
Logarithmic
Arithmetic
Matrix
Fourier Transform
Exponential Functions
Error Functions (erf)

- import numpy as np
  import(whenever we need to call a library and need the help, you call this function)
  as (alias for a shorter name)
- Python has other libraries like math, statistics, why do we need Numpy?

→ Data Scientist has to work with data in different directions.
→ It can be a data in 2D space or a 3D space.
→ we need to study the location of the data, physical characteristics of the data.
→ Numpy will work with data in different dimensions.

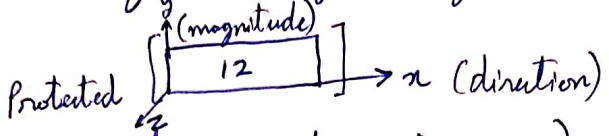
Take the example of a rubics cube:-



→ '.' (dot) is accession operator
(To give access to the data)

→ np.array() ← Function (we are creating an array, asking numpy to create an array)

array is a function of numpy.



1D (meaning of dimension here is row).

→ Python can convert complex into array

Example :-

Inp:- np.array(12.6767+7j) [Instead of writing 7 python tells to write 7.0 for more precision and more accuracy

Out:- array(12.6767+7.j)

How did the point (.) appear in the output :-

- Both numpy and pandas has a default data type. Default means a special kind of data which is preferred by the libraries. That data type is float 64.

→ Array has some protection and direction
The only data protection we have is in the tuple.

→ Benefits of working with arrays :-

- Makes you work faster
- Data has direction and protection
- It will not waste memory in computer or server.
- Processing time is very less (in microseconds)
- It is helping the data scientist to keep pace with the ML.

- Input :- [56+8j]
output :- [(56+8j)] (why output is covered with tuple)
Because python indicates that this is one complex number and not two seperate numbers.

• Tuple is the most secured data type

• Example:-
Inp:- abc = (1, 2, 3)
np.array(abc)
out:- array([1, 2, 3]) (why the list appears?)

(It is converted because list is a mutable data type. That is why behavioural change has happened.)

→ Python is a language of behaviours. It is behaving very delicately with the data. With different sets of data, it will change the behaviour.

→ Python automatically secures complex numbers.

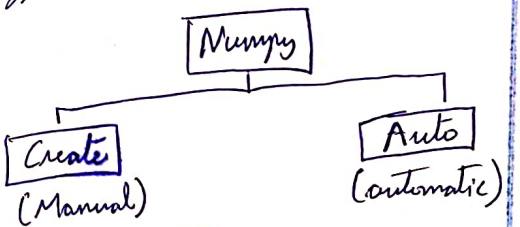
→ Numpy array works with mutable data because we need to work with the data.

• Example:-
Inp:- np.array(None)
Out:- array(None, dtype=object) → In python none means uncertainty, we are uncertain about the result.

• In python, there are two types of dictionaries → unordered
→ Ordered

• Dictionary data type is also known as mapping data type, relational data type, compound data type

• In Numpy we create(1) and auto generated(2)



• There are automatic and manual methods in python.

• np.linspace(0,10,3) (divide the line into 3 parts)
output:- array([0., 5., 10.])

- Numpy has some automatic methods of data generation
- static array :- It is a one dimensional array and this one dimensional array the data is fixed
- np.arange (0,50,-2)
  It will not accept the negative step. It will give an error.
  # negative stepsize not allowed ; empty array as output
  input :- np.arrange (0,50, -2)
  Output :- ([], dtype = int 32)
- static data means the data is fixed
- All the 7 data types can be converted into arrays but not a string because it contains alphabets.
- a.ndim (ndim means number of dimensions)
- Python numpy arrays are more compact compared to lists
- Numpy array can store n-dimensions.
- Time required for computations in an array is very less compared to list.

Stacking operations like hstack, vstack, column, dstack. (in python notebook)
Splitting operations like uniform, non uniform splitting.
Iteration in numpy arrays.
Indexing And slicing

- Array split is interested in equal distribution of data points.
- Horizontal split is related to columns.
- In vertical split, it arranges the data row-wise.
- Horizontal and vertical splitting doesn't allow negative splitting.
- Dividing into fractional parts is not allowed in numpy.
- axis = 0 represents row and axis = 1 means a column.
- GCS (Garbage Collection System)

• What are the most important features of Pandas, and why did we learn?
→ Pandas is the only data analytics library in the world which works with numbers, strings, all text formats, all file formats etc, it works with gbq, sql, webpages, html etc.
→ Data Alignment
→ Pandas works with all kind of data in analytics, has the ability to involve numerical libraries like numpy, for scientific operations scipy, for machine learning libraries it will include scikit library and so on.

• Pandas can convert everything into series except set. set is an unordered collection of unique elements.
• None is uncertainty in python.

• Pandas is used for data manipulation, analysis and cleaning. It is well suited for different kinds of data, such as:-
• Tabular data with heterogeneously-typed columns.
• Ordered and unordered time series data.
• Arbitrary matrix data with row and column labels, as for example the data in sql table or excel spreadsheet.
• Unlabeled data
• Any other form of observational or statistical data sets.

Key Features
• Easy handling of missing data.
• Size mutability
• Flexible reshaping and pivoting of data sets.
• Hierarchical labelling of axes.
• Time series functionality.

Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the python programming language.
• Pandas data structure is divided into two types:-
→ Series and Data Frame

Series:-
- One dimensional labelled array
- Supports multiple data types.
- Data in the array can be of any type (int, float, strings, objects).
- It supports even when the data within the array is homogeneous.
- Pandas Series object are amphibian in character, exhibiting both ndarray and dictionary.

DataFrame:-
- Two dimensional labelled array
- Supports multiple datatypes
- Input can be a series.
- Input can be another Data Frame.

Parameters:-
Data:- This is the value you want your series to posses.
Index:- This is the index related to the value you use for the series.
dtype:- This specifies the type of values in the series.
Copy:- This copies the data which was input.

matplotl
seabor

- why plt.show
  without this, it will show object of the graph.
- Seaborn is more flexible
- In seaborn we have scatterplot.
- Subplot:- gives you the functionality of drawing two different graphs on a certain plot.
- Correlation:- only the numerical values correlation is shown.