

01-09-22 Assignment

Q] What is numpy? Why we need this.

→ It is a library for the Python programming language, used for the working with arrays.

- It has function which are working in domain of linear algebra, fourier transform, & matrices.
- Numpy was created in 2005 by Travis Oliphant.
- Also it is open source project & you can use it freely.
- It stands for Numerical Python.
- Its programming languages are Python, C.
- It provide an array object that is up to 50x faster than traditional Python lists.
- It provides a lot of supporting functions.

Why we need

- It is one of the most commonly used packages for Scientific computing in Python.
- It provides a multidimensional array objects.
- It provides better runtime & space complexity.
- We can do comprehensive mathematical functions, random number generators, ~~linear~~ linear algebra routines, etc.
- Its code core is well-optimized C code.

Q] What is Pd & explain.

→ Pd is also known as "Pandas" (pandas)

- It is a popular Python-based data analysis toolkit.
- It present parsing multiple file formats to converting an entire data table into Numpy matrix array.

- In python alias are an alternate name for referring to the same thing. 05
- Now pandas package can be referred to as pd insted of pandas.

3] How we read CSV files?

→ We have some steps.

1] Import the CSV library, import csv.

2] Open the CSV file

```
file = open('Salary_Data.csv')  
type(file)
```

3] Use the csv.reader object to read CSV file.

```
CSVreader = csv.reader(file)
```

4] Extract the field names

```
header = []  
header = next(CSVreader)  
header
```

5] Extract the rows/records.

```
rows = []  
for row in CSVreader: row.append(row)  
rows  
[[ '1', '39343.00'],  
 ...  
 ]]
```

6] Close the file

```
file.close()
```


4] What is `iloc()` function?

- • It is one of the functions defined in pandas module that helps us to select a specific row or column from the data set.
- Using this we can easily retrieve any particular value from a row or column.
 - It means Purely integer-location based indexing for selection by position.
 - It is primarily integer position based (from 0 to length-1) may also be used with boolean array.

5] What is `loc` & how it differs from `iloc`.

- • `loc` is used to select rows & columns by names/labels.
- `iloc` is used to select rows & columns by integer index / position. (zero based index position)
 - `loc` is the label-based.

6] What / Differentiate `fit()` & `predict()`.

→ fit :-

- It is implemented by every estimator & it accepts an input for sample data & for supervised models it also accepts an argument for labels.
- This is usually responsible for numerous operations.
- OR, This method takes the training data as arguments, which can be one array in case of unsupervised learning.

predict:-

- This function enables us to predict labels of data values on the basis of trained model.
- It accepts only single argument which is usually data to be tested.

Syntax:-

model.predict(data)

7] What is meant by train-test split

- • It is used to estimate the performance of machine learning algorithm that are applicable for prediction-based algorithm/application.
- This method is fast & easy procedure to perform, such that we compare our machine learning model's result to machine result.
- It is useful for classification & can be used for any supervised learning algorithm.