

Python Lab Assignment 9

AIM: File Handling in Python

THEORY:

- Files: Files are named locations on disk to store related information. They are used to permanently store data in a non-volatile memory i.e. hard disk.
- Types of Files in Python:
 - 1) Text File: Text file store the data in the form of characters. They are used to store characters or strings.
Eg. abc.txt.
 - 2) Binary File: Binary file store the data in the form of bytes. They ~~can~~ can be used to store text, image, audio ~~and~~ and video.
- Different File Modes in Python:
 - 1) ~~not~~ Write mode (w): This mode is used to add data to a file. If data is already present in the file, it would be deleted and new ~~not~~ data will be stored.
 - 2) Read mode (r): This mode is used to read data from a file. It is used for existing files only.

3) Append mode (a): This mode is used to append data to a file. It either adds data to an existing file or creates a new file and add data to the file.

4) wt mode: This mode is used to read and write data in a file. It will override existing data.

5) rt mode: This mode is used to read and write data in a file. The previous data will not be ~~erased~~.

6) at mode: This mode is used to append and read data from the file. It doesn't override existing data.

7) x mode: This mode is used to open a file in exclusive creation mode for write ~~operator~~ operator.

- Pickling: Pickling is the process by which ~~by object~~ objects are converted ~~to~~ to byte stream. It is about serializing the object structure in ~~by~~ by
- Unpickling: Unpickling is the process of retrieving original Python objects from the stored string representation i.e. the pickle file.

CONCLUSION: We learnt about different file modes in Python and we also learnt about pickling and unpickling files.

Python Lab
Assignment 11

AIM : To study and implement CRUD operations using pymysql Python package and MySQL Database Connectivity.

THEORY:

Steps to install/connect MySQL Database with Python.

- 1) Install MySQL connector module
Use the pip command to install MySQL connector Python.
- 2) Import MySQL connector module
Import using a import mysql.connector statement so you can use this modules to communicate with the MySQL database.
- 3) Use the connect() method
Use the connect() method of the MySQL Connector class with the required arguments to connect MySQL.
- 4) Use the cursor() method
Use the cursor() method of a MySQL Connection object to create a cursor object to perform various SQL operations
- 5) Use the execute() method
The execute() method runs the SQL query & returns the result.
- 6) Execute / Extract result using fetchall()
Use cursor.fetchmany() or fetchone() or fetchall() to reach query result.

7) Close cursor & connection objects
Use cursor.close() & connection.close() method to close open connection to end the connection.

- fetchall(): It fetches all the rows of a query result. It returns all the rows as a list of tuples. An empty list is returned if there is no record to fetch.
- execute(): It helps us to execute the query and return records according to the query.
Syntax : cursor.execute(query, args=None).
- executemany(): It helps us to INSERT or REPLACE multiple records at once.

CONCLUSION: We learnt how to connect database application like MySQL to Python and executed many commands in it with different queries.

Nikhil Marchhoni
S13
48

Python lab
Assignment 12

AIM: To study different types of plots using Numpy & Matplotlib.

THEORY:

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multiplatform data visualization library built on Numpy arrays & designed to work with the broader ~~Scipy~~ SciPy stack. One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. It consists of several plots like line, bar, scatter, histogram.

- subplots() : It is a Matplotlib function that is used to display multiple plots in one figure. It takes various arguments such as a number of rows, columns, etc.
- Line Plots : A line plot is used to see the relationship between the x & y - axis. The plot() function in the Matplotlib library's pyplot module is used to create a 2D line histogram plot of the co-ordinate x & y. plot() will take various ~~as~~ arguments like plot x, y are co-ordinates of the horizontal & vertical axis, they are optional arguments. scale x, scale y used to autoscale x & y axis.

Histogram: The most common graph for displaying frequency distribution is a histogram. We create a bin of ranges, then distribute the whole range of the value into series of intervals & count the values which will fall in the given interval.
`plt.hist()` function with arguments with arguments like `data, bins, color, etc.`

Bar Plot: Mainly barplot is used to show the between the numeric & the categorical values. In a bar plot, we have one axis representing a particular category of the column and another axis representing the values or count of the specific category.
`plt.bar(x, height, width, bottom, align)`

Piechart: A piechart is used to show the percentage of the whole. Hence it is used when we want to compare the individual categories with the whole. It takes the different parameters like `r, labels, autopct`.

CONCLUSION: We learnt how to use Matplotlib to plot ^{any} different types of graphs and charts. Its numerical mathematics extension Numpy helps us with an object-oriented API for embedding plots into applications.

Python Lab
Assignment 13

AIM: Basic operations using pandas like series, data frames, indexing, filtering, combining & merging data frames.

THEORY:

Pandas is an open-source library that is made for working with relational or labeled data both easily and intuitively. It provides various data structures and operators for manipulating numerical data & time series. This library is built top of the Numpy library. After installation of pandas, we import the library into our program. It generally provides two data structures for manipulating data:

1) Series: Pandas series is one-dimensional labelled array capable of holding data of any type. The axis labels are collectively called indexes. Pandas series is nothing but a column in an excel sheet. Labels need not be unique but must be a hashable type. The object supports both integer & label-based indexing & provides a host of methods for performing operations involving the index.

A pandas series will be created by loading the datasets from existing ~~any~~ storage, can be SQL database, CSV file, or Excel file. Pandas series can be created from the list, dictionary & from a scalar value.

2) Database: Pandas DataFrame is a two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled rows (rows & columns). A DataFrame is a 2 dimensional data structure i.e. data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principle components, the data, rows and columns. In the real world, a Pandas DataFrame will be accepted by loading the datasets from ~~an~~ existing storage can be SQL, CSV file or Excel file. It can be created from the lists, dictionaries & from a list of dictionaries, etc.

CONCLUSION: Pandas is an open source Python package that is most widely used for data science / data analysis and machine learning tasks. It offers data structures & operations for manipulating ~~numerical~~ numerical tables and time series.

Python Lab
Assignment 14

AIM: To study Scipy library & linear algebra using Scipy.

THEORY:

Scipy is a scientific computation library that ~~uses~~ Numpy underneath. Scipy stands for Scientific Python. It provides more utility functions for optimization, stats and signal processing. It is open source so it is free to use. It provides many built-in scientific constants. The library supports integration, gradient optimization, special functions, ordinary differential equation solvers, parallel programming tools & many more.

Scipy has a number of sub-packages for the various scientific computing domains. It provides very fast linear algebra capabilities. Linear algebra routine accepts two-dimensional array object and output is also given as a 2-D array. A linear algebra problem can be solved by:

→ linalg.solve(): It is used to solve the linear equation $x + 2y = 3$ for the unknown. Its values of x, y .

$$x + 3y + 10z = 10$$

$$2x + 2y + 7z = 13$$

$$5x + 3y + 8z = 30$$

The above equations can be solved as:

$$a = np.array ([[1,3,10], [3,12,7], [5,8,2]])$$

$$b = np.array ([10], [18], [30])$$

$$x = \text{linAlg}.solve(a, b)$$

→ linAlg.det(): The determinant of the square matrix is found by using the linAlg.det() function. The determinant A is denoted as $|A|$. It accepts a matrix and returns a scalar value.

→ linAlg.eig(): We use this command to find the value of a matrix, they take arguments of the input matrix, ~~and~~ along with its eigen vector.

$$\lambda, v = \text{linAlg}.eig(a)$$

$\lambda \rightarrow$ eigen values

$v \rightarrow$ eigen vectors

→ linAlg.inv(), linAlg.pinv(): It calculates the inverse & the pseudoinverse of a matrix, it takes arguments of the input matrix, condition & rank check prints are optional arguments.

→ linAlg.svd(): The singular-value decomposition (SVD) is a matrix decomposition method for reducing a matrix to its constituent parts to subsequent matrix calculations simpler.

CONCLUSION: SciPy is used for computing scientific and technical computations. We can perform scientific computations with the help of the ~~SciPy~~ library in Python.