19se02it058

Assignment 1

1] Numpy:

- NumPy is a Python library used for working with arrays.
- It also has functions for working in domain of linear algebra, fourier transform, and matrices.
- NumPy stands for Numerical Python.
- In Python we have lists that serve the purpose of arrays, but they are slow to process.
- NumPy aims to provide an array object that is up to 50x faster than traditional Python lists.
- The array object in NumPy is called ndarray, it provides a lot of supporting functions that make working with ndarray very easy.
- Arrays are very frequently used in data science, where speed and resources are very important.
- NumPy arrays are stored at one continuous place in memory unlike lists, so processes can access and manipulate them very efficiently.
- This behaviour is called locality of reference in computer science.
- This is the main reason why NumPy is faster than lists. Also it is optimized to work with latest CPU architectures.
- NumPy is a Python library and is written partially in Python, but most of the parts that require fast computation are written in C or C++.

Advantages of Numpy:

- The core of Numpy is its arrays. One of the main advantages of using Numpy arrays is that they take less memory space and provide better runtime speed when compared with similar data structures in python(lists and tuples).
- Numpy support some specific scientific functions such as linear algebra. They help us in solving linear equations.
-

- Numpy support vectorized operations, like elementwise addition and multiplication, computing Kronecker product, etc. Python lists fail to support these features.
- It is a very good substitute for MATLAB, OCTAVE, etc as it provides similar functionalities and supports with faster development and less mental overhead(as python is easy to write and comprehend)
- NumPy is very good for data analysis.

Disadvantages of Numpy:

- **Using "nan" in Numpy:** "Nan" stands for "not a number". It was designed to address the problem of missing values. NumPy itself supports "nan" but lack of cross-platform support within Python makes it difficult for the user. That's why we may face problems when comparing values within the Python interpreter.
- **Require a contiguous allocation of memory:** Insertion and deletion operations become costly as data is stored in contiguous memory locations as shifting it requires shifting.

2] Scipy:

- SciPy in Python is an open-source library used for solving mathematical, scientific, engineering, and technical problems. It allows users to manipulate the data and visualize the data using a wide range of high-level Python commands. SciPy is built on the Python NumPy extention. SciPy is also pronounced as "Sigh Pi."

Sub-packages of SciPy:

- File input/output – scipy.io
- Special Function – scipy.special
- Linear Algebra Operation – scipy.linalg
- Interpolation – scipy.interpolate
- Optimization and fit – scipy.optimize
- Statistics and random numbers – scipy.stats
- Numerical Integration – scipy.integrate
- Fast Fourier transforms – scipy.fftpack

- Signal Processing – scipy.signal

- Image manipulation – scipy.ndimage

Benefits of Scripy:

- High-level commands and classes for visualizing and manipulating data.
- Powerful and interactive sessions with Python.
- Classes, and web and database routines for parallel programming.
- Easy and fast.
- Open-source.

3] Scikit-learn:

- Scikit-learn is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

4] Theano:

- Theano is a Python library that allows us to evaluate mathematical operations including multi-dimensional arrays so efficiently. It is mostly used in building Deep Learning Projects. It works a way more faster on Graphics Processing Unit (GPU) rather than on CPU. Theano attains high speeds that gives a tough competition to C implementations for problems involving large amounts of data. It can take advantage of GPUs which makes it perform better than C on a CPU by considerable orders of magnitude under some certain circumstances.
  It knows how to take structures and convert them into very efficient code that uses numpy and some native libraries. It is mainly designed to handle the types of computation required for large neural network algorithms used in Deep Learning. That is why, it is a very popular library in the field of Deep Learning.\

- Theano is a sort of hybrid between numpy and sympy, an attempt is made to combine the two into one powerful library. Some advantages of theano are as follows:
- **Stability Optimization:-**
- Theano can find out some unstable expressions and can use more stable means to evaluate them

- **Execution Speed Optimization:** As mentioned earlier, theano can make use of recent GPUs and execute parts of expressions in your CPU or GPU, making it much faster than Python

- **Symbolic Differentiation:** Theano is smart enough to automatically create symbolic graphs for computing gradients

5]TensorFlow:-

- Unlike other numerical libraries intended for use in Deep Learning like Theano, TensorFlow was designed for use both in research and development and in production systems, not least RankBrain in Google search and the fun DeepDream project.
- It can run on single CPU systems, GPUs as well as mobile devices and large scale distributed systems of hundreds of machines.

6]Keras:-

- Keras is a high-level, deep learning API developed by Google for implementing neural networks. It is written in Python and is used to make the implementation of neural networks easy. It also supports multiple backend neural network computation.

7]PyTorch:-

- PyTorch is an optimized Deep Learning tensor library based on Python and Torch and is mainly used for applications using GPUs and CPUs. PyTorch is favored over other Deep Learning frameworks like TensorFlow and Keras since it uses dynamic computation graphs and is completely Pythonic. It allows scientists, developers, and neural network debuggers to run and test portions of the code in real-time. Thus, users don’t have to wait for the entire code to be implemented to check if a part of the code works or not.

8]Pandas:-

- Pandas has been one of the most commonly used tools for Data Science and Machine learning, which is used for data cleaning and analysis.
- Here, Pandas is the best tool for handling this real-world messy data. And pandas is one of the open-source python packages built on top of NumPy.

9]Matplotlib:-

- Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications.