

Alternatives to Numpy and Pandas

Assignment ii

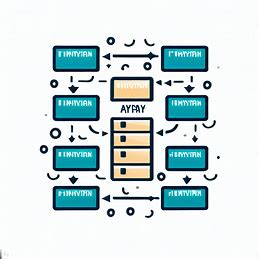
Daniel Pereira | Computing for Data Science | 23rd September, 2023.

# What is Numpy

Numpy is a python library that is used for converting arrays into matrices and vectors for matrix and vector operations. It can perform a wide range of mathematical operations and is used widely in the field of data science, artificial intelligence, etc.

# Alternatives to Numpy

## Tiny Array



Tiny Array is a python library similar to Numpy but used for handling smaller data. On computing smaller data, Tiny Array is 3 to 7 times faster than Numpy. However, it cannot handle large data. Another drawback of tiny array is that it is not used in most data science libraries like pandas, TensorFlow, etc.

An example of tiny array is: -

import tinyarray as ta

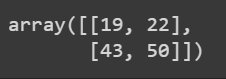
a = ta.array([[1, 2], [3, 4]])

b = ta.array([[5, 6], [7, 8]])

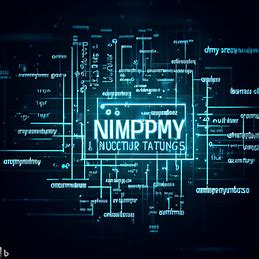
c = ta.dot(a,b)

c

Output:



## Tiny Numpy:



Tiny NumPy is a python library that is similar to Numpy but has limited functionality. One can calculate matrices and vectors using Tiny Numpy. However, like tiny array, it is faster than NumPy but not suitable for large arrays.

An example of Tiny Numpy:

from tinynumpy import tinynumpy as tnp

a = tnp.array([1.2, 3.4, 5.6])

b = tnp.multiply(a,a)

b

Output:



## SciPy:



SciPy is a python library that is built on NumPy and can be integrated with other data science libraries like pandas and matplotlib. It can perform many mathematical operations that include modules for linear algebra, optimization, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers, etc.

An example of SciPy:

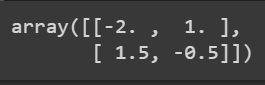
import numpy as np

from scipy import linalg

A = np.array([[1,2],[3,4]])

linalg.inv(A)

Output:



## Sympy:



SymPy is a python library for symbolic mathematics. That is unlike other libraries that converts the equations into numeric arrays before solving, SymPy solves equations as it is.

An example of SymPy:

import sympy

sympy.init\_printing()

x, y = sympy.symbols('x y')

eq = sympy.Eq(x + y, 2)

sol = sympy.solve(eq, x)

sol

Output:



# What is Pandas?

Pandas is a popular library used widely in data science. Pandas is used for converting data into a data frame from different formats that include csv, Json and excel before performing operations on it. Pandas can perform a variety of operations that include data sorting, merging, slicing, searching, etc.

# Alternatives to Pandas:

## Pandapy:



PandaPy is a lightweight alternative to pandas. It is suitable for handling data frames up to 50,000 rows. It is more memory efficient than pandas for smaller datasets.

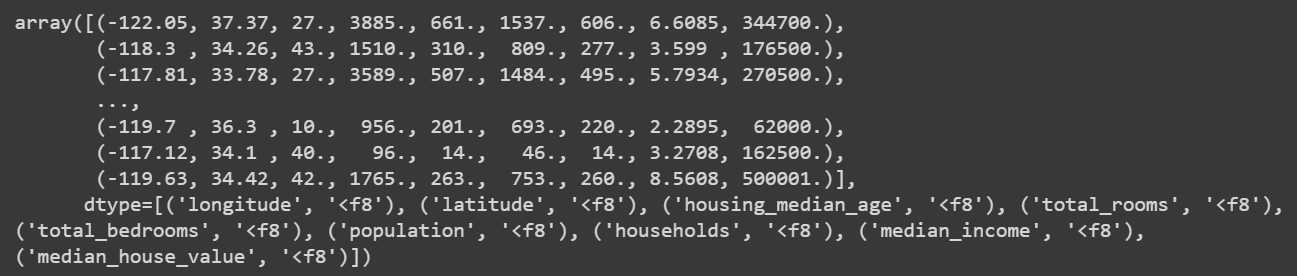
An example of PandaPy:

import pandapy as pp

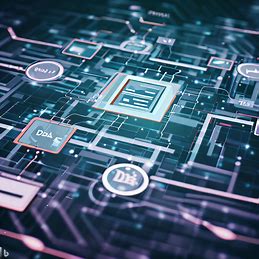
df = pp.read("/content/sample\_data/california\_housing\_test.csv")

df

Output:



## Datatables:



Data Tables is a python library used for handling large datasets. It can handle up to 100 GB of data and is suitable for deep learning that uses large amounts of data.

Example:

import datatables as dt

df = dt.fread("/content/sample\_data/california\_housing\_test.csv")

df

# Conclusion

Though there may be many alternatives to NumPy and pandas, most of those packages are either built on them or have performance not up to NumPy and Pandas.