Introduction to NumPy: Takeaways

by Dataquest Labs, Inc. - All rights reserved © 2024

Syntax

SELECTING ROWS, COLUMNS, AND ELEMENTS FROM AN NDARRAY

- Convert a list of lists into an ndarray:

```
import numpy as np
f = open("nyc_taxis.csv", "r")
taxi_list = list(csv.reader(f))
taxi = np.array(taxi_list)
```

- Selecting a row from an ndarray:

```
second_row = taxi[1]
```

- Selecting multiple rows from an ndarray:

```
all_but_first_row = taxi[1:]
```

- Selecting a specific element from an ndarray:

```
fifth_row_second_column = taxi[4, 1]
```

SLICING VALUES FROM AN NDARRAY

- Selecting a single column:

```
second_column = taxi[:, 1]
```

- Selecting multiple columns:

```
second_third_columns = taxi[:, 1:3]
cols = [1, 3, 5]
second_fourth_sixth_columns = taxi[:, cols]
```

- Selecting a 2D slice:

```
twod_slice = taxi[1:4, :3]
```

VECTOR MATH

- `vector_a + vector_b` : addition
- `vector_a - vector_b` : subtraction
- `vector_a * vector_b` : multiplication (this is unrelated to the vector multiplication used in linear algebra)
- `vector_a / vector_b` : division

CALCULATING STATISTICS FOR 1D NDARRAYS

- `ndarray.min()` to calculate the minimum value

- `ndarray.max()` to calculate the maximum value
- `ndarray.mean()` to calculate the mean average value
- `ndarray.sum()` to calculate the sum of the values

CALCULATING STATISTICS FOR 2D NDARRAYS

- Max value for an entire 2D ndarray:

```
taxi.max()
```

- Max value for each row in a 2D ndarray (returns a 1D ndarray):

```
taxi.max(axis=1)
```

- Max value for each column in a 2D ndarray (returns a 1D ndarray):

```
taxi.max(axis=0)
```

Concepts

- Python is considered a high-level language because it abstracts away the need for manual memory allocation and explicit CPU operation instructions.
- Low-level languages like C offer more control and can lead to performance improvements, but they typically require more time and effort from programmers.
- The NumPy library bridges this gap by allowing us to write Python code while leveraging the performance benefits of C.
- One key feature of NumPy that enhances performance is **vectorization**, which uses **Single Instruction Multiple Data (SIMD)** to process data more efficiently.
- In NumPy, a list is referred to as a 1D ndarray, while a list of lists is called a 2D ndarray.
- Ndarrays use indices for both rows and columns, providing a convenient method for selecting and slicing values within the array.
- NumPy **broadcasting** allows for element-wise operations between compatible arrays with different shapes.

Resources

- Arithmetic functions from the NumPy documentation.
- NumPy ndarray documentation
- NumPy broadcasting

Takeaways by Dataquest Labs, Inc. - All rights reserved © 2024