

NumPy Python

Léo

June 2024

1 Introduction

NumPy est considéré comme un des packages / bibliothèque fondamentaux pour tout ce qui est calcul scientifique en python et science des données.

La bibliothèque Pandas est en partie basée sur la bibliothèque NumPy.

2 Comment NumPy fonctionne ?

NumPy est une bibliothèque de tableaux multidimensionnelle. Cela signifie que l'on peut utiliser NumPy pour stocker toutes sortes de données dans des tableaux 1D, 2D, 3D, etc...

Pourquoi utiliser NumPy plutôt que des listes ?

La différence principale vient de la vitesse. Les listes sont très lentes alors que NumPy est très rapide. Une des raisons à cela est que NumPy utilise des types fixes (Fixed Type).

Pour cela, imaginons une matrice 3x4 et concentrons nous sur le 5 : Les ordina-

3	1	2	4
5	7	1	2
4	1	0	1

teurs ne voient pas le 5 que comme un nombre binaire : $5 = 0000\ 0101$ (8 bits constituant 1 octet). NumPy va convertir cet octet en type int 32 : $00000000\ 00000000\ 00000000\ 00000101$ (4 octets).

Quant aux listes, elles vont représenter cet octet en utilisant un type int intégré pour python composé de quatre éléments : la Taille (Size), le nombre de référence (Reference Count), le type de l'objet (Object Type) et la valeur de l'objet (Object Value) : $00000000\ \dots\ (x6)\ 00000101$. On voit donc assez rapidement que les listes nécessitent beaucoup plus d'espaces que NumPy. De plus, nous n'avons

pas besoin de vérifier le type à chaque fois lorsque nous parcourons chaque élément d'un tableau NumPy.

Un autre facteur de cette rapidité est que NumPy utilise une mémoire contigüe = block de mémoire l'un à la suite de l'autre.

Bénéfices :

- Utilisation de Traitement vectoriel SIMD
- Utilisation plus optimisée du cache

Que ce soit avec les listes ou NumPy, on peut faire des insertions, suppressions, concaténations, ect. Seulement, NumPy permet de faire beaucoup plus.

2.1 Applications de NumPy

La première application la plus connue de NumPy est une sorte de remplacement de Matlab.

NumPy est aussi utile pour du Plotting (Matplotlib), pour du back-end (Pandas, Connect 4, Digital Photography) mais aussi extrêmement important pour du Machine Learning (tensors liés à NumPy).

3 Les bases

Voici différents fichiers de code expliquant les bases de NumPy (Tableau, Shape, size, Changement d'éléments, Initialisation, Maths - Algèbre linéaire et Stats, Réorganisation de tableaux - reshape, vstack, hstack, Chargement de fichiers data) :

https://github.com/1leomllr/library_basics/tree/main/NumPy%20Basics/src

```
>> a[ 2:4 , 0:2 ]  
>> a[ [0,1,2,3] , [1,2,3,4] ]  
>> a[ [0,4,5] , 3: ]
```

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30

Figure 1: Problème d'indexation