

Matrices terme-document

April 2, 2024

- division d'une matrice par un scalaire :

```
>>>
>>> a
array([[ 3,  9, 15],
       [21, 24, 27],
       [ 3, 12,  3]])
>>> a / 3
array([[1.,  3.,  5.],
       [7.,  8.,  9.],
       [1.,  4.,  1.]])
>>>
```

- division d'une matrice par un vecteur :

```
>>>
>>> a
array([[ 3,  9, 15],
       [21, 24, 27],
       [ 3, 12,  3]])
>>> a / np.array([1,3,5])
array([[ 3. ,  3. ,  3. ],
       [21. ,  8. ,  5.4],
       [ 3. ,  4. ,  0.6]])
...
```

- application des fonctions par axe (axis=0 : par colonnes, axis=1 : par lignes) :

```
>>> a
array([[ 3,  9, 15],
       [21, 24, 27],
       [ 3, 12,  3]])
>>> np.amax(a,axis=0)
array([21, 24, 27])
>>>
```


- transformer en entier, string, ... : astype(int/str/float/bool...)

- sélectionner des valeurs dans une matrice selon des critères :

- $a > 0$. (boolean)
- $a[a > 0]$ (valeurs)
- $\text{np.argwhere}(a > 0)$ (indices)

- trouver les éléments identiques dans deux matrices :
 $\text{np.argwhere}(a == b)$

Exercices.

1. Créez un vecteur a avec les entiers de 0 à 9.
2. Créez une matrice 3×3 avec la valeur True partout.
3. Récupérez tous les nombres impairs du vecteur a .
4. Remplacez tous les nombres impairs dans a par -1.
5. Transformez une matrice de dimensions variables en un vecteur de valeurs, avec la fonction reshape et sans la fonction flatten
6. Transformez un vecteur de n dimensions en une matrice $2 \times n$. 
7. Si a et b sont des vecteurs de même nombre de dimensions, donnez les éléments qu'ils ont en commun, et leurs indices.
8. Récupérez tous les éléments de a entre 5 et 10.
9. Echangez les deux premières colonnes dans une matrice.
10. Dans la matrice de co-occurrences du corpus reuters, supprimez les colonnes qui correspondent aux mots avec 0 occurrences dans le corpus.
11. Donnez la liste des 50 mots avec le tf-idf le plus élevé dans le corpus reuters. Affichez le score tf-idf aussi.
12. Représentez le vocabulaire du corpus acl.txt par une matrice terme-contexte : les lignes et les colonnes correspondent aux mêmes mots (wordlist.txt), et les cases aux valeurs de co-occurrences dans une fenêtre de 5 mots.
13. Complétez le programme de manière à demander deux mots à l'utilisateur, et retourner la similarité cosinus entre les deux mots.