

ENSAE TD noté, mercredi 5 novembre 2024

Toutes les questions valent 2 points.

Au pays des schtroumphs, le schtroumph statisticien a décidé d'organiser le recensement des schtroumphs. La première tout s'est déroulé à la perfection et a abouti au résultat suivant :

	schtroumph	schtroumph	index
0	8	7	schtroumph
1	18	18	schtroumph
2	6	8	schtroumph

L'année suivante, le statisticien a délégué le recensement à son apprenti.

	schtroumph	schtroumph	index
0	18	18	schtroumph
1	7	9	schtroumph
2	8	6	schtroumph

Au début, il s'est demandé comment la population avait tant changé. Puis il s'est rappelé que la langue schtroumph est une langue parlée. Tous les mots s'écrivent de la même façon. Le schtroumph statisticien va devoir réconcilier les données.

Question 1 : Implémenter une fonction qui calcule la distance entre deux matrices

```
1 def distance(table1, table2):
2     # ...
3     return ...
4
5 assert distance(np.array([[0, 1], [0, 1]]), np.array([[0, 1], [0, 1]])) == 0
```

Question 2 : Implémenter une fonction qui retourne les permutations des n premiers entiers

Question 3 : Implémenter une fonction qui permute les colonnes d'une matrice.

```
1 def permute_colonne(table, permutation):
2     # ...
3     return ...
```

Question 4 : Faire de même pour les lignes

```
1 def permute_ligne(table, permutation):
2     # ...
3     return ...
```

Faire de même pour les lignes.

Question 5 : Ecrire une fonction qui retourne les deux permutations ligne/colonne qui minimise la distance entre les deux matrices, en déduire la case qui a changé.

Question 6 : Quel est le coût de cette fonction ?

Question 7 : C'est beaucoup trop long. On propose que calculer chaque permutation séparément. On cherche donc la meilleure permutation qui minimise la distribution de la somme par ligne et par colonne entre les deux matrices. Ecrire une fonction qui implémente ce raisonnement.

Question 8 : Mais c'est encore trop coûteux. On cherche la matrice M qui minimise $AM = B$ où A et B sont les sommes sur les colonnes où lignes des matrices de statistiques observées sur deux années.

Question 9 : Comment utiliser cette fonction pour implémenter une version plus rapide de la fonction à la question 5.

Question 10 : La troisième année, une colonne est coupée en deux : une catégorie est divisée en deux sous-catégorie.

Que proposez-vous pour y remédier ?