

Les algorithmes KNN

A l'école des sorciers



Première étape :

Suivre le notebook 02_knn_Algorithme.ipynb

Répondre aux questions posées dans le Notebook sur le formulaire suivant :

https://forms.office.com/Pages/ResponsePage.aspx?id=DQSlkWdsW0yxEjajBLZtrQAAAAAAAAAAAAAN_th3gwdURFFOU1dMOVhJV0dCN1BaR1JCRDdLSzZEWS4u

Deuxième étape :

Travail à réaliser : travail d'approfondissement en individuel

@crédits : « Aidons le choixpeau magique », Prépubac 1ere NSI, éditions Hatier. Sujet adapté et remanié.

A l'entrée à l'école de **Poudlard**, le choixpeau magique répartit les élèves dans les différentes maisons (Gryffondor, Serpentard, Serdaigle et Poufsouffle) en fonction de leur courage, leur loyauté, leur sagesse, et leur malice.

Le choixpeau magique dispose d'un fichier CSV (sic !) dans lequel sont répertoriées les données d'un échantillon d'élèves. Voici les 6 premières lignes de ce fichier :

	Nom	Courage	Loyauté	Sagesse	Malice	Maison
0	Adrian	9	4	7	10	Serpentar
1	Andrew	9	3	4	7	Griffondor
2	Angelina	10	6	5	9	Griffondor
3	Anthony	2	8	8	3	Serdaigle
4	Arthur	10	4	2	5	Griffondor
5	Bellatrix	10	4	9	9	Serpentar

Et voici les élèves que le choixpeau magique souhaite orienter :

	Nom	Courage	Loyauté	Sagesse	Malice
0	Hermione	8	6	6	6
1	Drago	6	6	5	8
2	Cho	7	6	9	6
3	Cédric	7	10	5	6

L'objectif de cet exercice est d'aider le choixpeau à déterminer la maison des nouveaux élèves.

Partie I : Modéliser un élève

On décide de modéliser chaque élève par un dictionnaire avec les données à disposition. Par exemple :

```
adrian = {"nom" : "Adrian", "courage":9, "loyaute":4, "sagesse":7, "malice":10, "maison": "serpentard"}
hermione = {"nom" : "Hermione", "courage":8, "loyaute":6, "sagesse":6, "malice":6}
```

1. Donner la modélisation de l'élève Anthony
2. On décide d'utiliser la distance de Manhattan pour calculer la distance entre 2 élèves, c'est-à-dire :
$$\text{distance}(\text{eleve1}, \text{eleve2}) = |c1-c2| + |l1-l2| + |s1-s2| + |m1-m2|$$
 - 2.a. Avec cette formule, vérifier que la distance entre Hermione et Adrian est bien égale à 8
 - 2.b. Quelle est la distance entre Arthur et Drago ?
 - 2.c. Écrire le code d'une fonction `distance` qui prend deux élèves en paramètre et qui renvoie la distance entre ces 2 élèves. Ne pas oublier de préciser la documentation (`docstring`) et de donner au moins un test (`assert`).

Note : la valeur absolue d'un nombre s'obtient par `abs(nombre)`

Partie II : Charger les données dans un dataframe

Pour la suite du travail, nous allons continuer à utiliser pandas.

Voici le code d'une fonction qui permet de récupérer les données des élèves d'un fichier CSV pour les stocker dans un dataframe pandas.

```
def charger_eleve(fichier) :
    """
    permet de charger une liste d'eleves d'un fichier CSV
    Return : dataframe
    """
    return pd.read_csv(fichier, encoding = 'utf-8')

poudlard = charger_eleve('choixpeauMagique.csv')
```

Répondre aux questions suivantes à l'aide d'une instruction (et sans regarder le CSV) :

1. Quel est le nom du premier élève ?
2. Indiquer le nombre d'élèves par maison
3. Quel est la maison ayant le plus d'élève ?

Partie III : Trouver la maison majoritaire

Écrire le code d'une fonction qui prend en paramètre un dataframe Pandas et qui renvoie la maison la plus représentée dans la liste.

Ne pas oublier de préciser la documentation (`docstring`) et de donner au moins un test (`assert`).

Partie IV : Sept plus proches voisins

1. Compléter la formulation suivante :

Données :

DataFrame : une liste des élèves déjà présent à Poudlard

Nouveau : un nouvel élève qui n'a pas encore de maison

Résultat : Maison qui sera attribuée au nouvel élève selon le principe des 7 plus proches voisins.

Algorithme :

(a compléter)

2. Implémenter cet algorithme en python