

Exercice knn : Diabète

On dispose d'un fichier [diabete.csv](#) comportant 8 colonnes de diverses mesures (insuline, âge, ...). Le séparateur de données est la virgule.

La 9ème colonne est constituée de 0 et de 1:

- 1 signifie que la personne suit un traitement pour des problèmes liés au diabète,
- 0 qu'elle n'en suit pas.

Il s'agit de nos 2 classes pour ce problème.

1. Un nouveau patient est caractérisé par les mesures suivantes (il s'agit, dans l'ordre, des données correspondant aux 8 premières colonnes):

nouveauPatient = [1, 89, 67, 24, 80, 23, 0.6, 65]

En réutilisant (et modifiant) le code écrit dans l'exercice des iris, déterminer si ce nouveau patient devra suivre un traitement. (testez diverses valeurs de k dans votre application du principe k-NN)

2. On aimeraient avoir une idée de la **fiabilité de la méthode**.

Pour cela, on découpe les données en 2 listes:

- une liste [entraînement](#) qui contiendra une certaine proportion des données du fichier. Par exemple 90% des données.
- une liste [test](#) qui contiendra le reste des données.

Comme on connaît la classe des individus de TEST, on peut vérifier la proportion de bonnes réponses par la méthode sur les éléments de TEST en utilisant ENTRAINEMENT comme jeu de données.

Écrire une telle fonction et faites des essais.