

TD1 – Classification non supervisée

L'objectif de ce TD est d'implémenter en *Python* l'algorithme des centres mobiles étudié en cours. Des expérimentations seront ensuite menées pour évaluer la performance du modèle sur des jeux d'essai.

1. Présentation des éléments fournis

Un fichier Python (`centres_mobiles.py`) est fourni, contenant des fonctions à compléter afin de rendre le programme fonctionnel.

Afin de tester les fonctions développées, le fichier de données `iris.txt` est fourni. Il s'agit d'un célèbre jeu de données constitué de 150 exemplaires correspondant à des plantes de type « Iris ». Ces exemplaires sont caractérisés par 4 variables prédictives (longueur des sépales, largeur de sépales, longueur des pétales, largeur des pétales). Chaque exemplaire appartient à l'une des trois familles d'Iris : Setosa (code 0), Versicolour (code 1), Virginica (code 2).

Les données sont organisées avec les exemplaires en ligne (150) lignes, et les variables en colonne (4 variables prédictive + la variable cible correspondant au code famille).

2. Ecriture du programme

Compléter le fichier fourni afin rendre le programme fonctionnel. Il est conseillé de suivre l'ordre des fonctions présentes dans le fichier, et de les tester à chaque étape.

3. Expérimentations

Une fois le programme fonctionnel, réaliser des expérimentations afin d'évaluer la performance du modèle. Par exemple :

- Influence du nombre de classes k
- Influence de la stratégie d'initialisation des centres
- Application sur d'autres jeux de données (par exemple à partir de <https://archive.ics.uci.edu/>).