

Activité 3 : Classification KNN

On souhaite effectuer de la classification en utilisant la méthode des K-plus proches voisins.

Dans cette activité, on se basera sur la base "IRIS" du package `sklearn.datasets`.

1- Charger les données.

a. Créer une matrice qui ne contient que les valeurs des variables (séparer les valeurs des attributs de leurs classes d'appartenances).

b. Créer un vecteur qui ne contient que les étiquettes des classes.

2- Afficher les noms des variables et en déduire leur nombre.

3- Afficher les classes. En déduire le type de classification auquel on s'intéresse.

4- Normaliser les valeurs des variables pour qu'elles suivent une loi normale de moyenne 0 et de variance 1.

5- Ecrire une fonction qui permet de faire la classification avec l'algorithme des k-NN et afficher le taux d'erreur obtenu sur l'ensemble de test.

6- Diviser l'ensemble de données en un ensemble d'apprentissage qui contient 60% des observations et un ensemble de test (40%).

7- Effectuer la classification des données en considérant $k=7$.

8- Utiliser la validation croisée pour déterminer la valeur optimale de k puis refaire la classification en utilisant cette valeur.

9- Afficher la matrice de confusion et la précision du modèle sur les données de test