

Travaux Pratique 1

L'objectif de ce TP est double :

- Prise en main de la bibliothèque *scikit-learn* de Python, dédiée à l'apprentissage automatique,
- Sensibilisation à l'évaluation des modèles appris en classification supervisée.

Jeu de données

Wine est un jeu de données contenant des données qui sont les résultats des analyses chimiques de vins cultivés de trois différents cultivars dans la même région italienne.

Consignes

1. Charger le dataset *Wine* à partir du package *sklearn.datasets*
2. Afficher les informations suivantes :
 - Le nombre d'échantillons et le nombre d'attributs (*Features*)
 - La liste des attributs du jeu de données, et
 - La variable cible
3. Convertir le dataset en DataFrame
4. Afficher les 5 premières lignes.
5. Afficher les types de données
6. Vérifier s'il y a des données manquantes
7. Afficher une description statistique des attributs suivants : *alcohol*, *malic_acid* et *ash*
8. Afficher un graphique pour visualiser la fréquence de chaque classe.
9. Afficher un graphique pour visualiser la distribution des attributs suivants : *alcohol*, *magnesium*, *color_intensity*
10. Afficher un graphique pour visualiser les rapports entre l'*alcohol* et *color_intensity* et les rapports entre l'*alcohol* et *hue*. Que remarquez-vous ?
11. Qu'affiche cette ligne `df[['alcohol','magnesium','color_intensity']].plot.box()`. Que remarquez-vous ?
12. Entraîner un classifieur KNN (par défaut) à reconnaître les classes des vins. Mesurer les performances de votre classifieur.
13. Modifier le code pour entraîner un classifieur KNN avec différentes valeurs de *k*. Que remarquez-vous ?