

Université Mohammed Premier
École Nationale des Sciences Appliquées
Filière Génie Informatique

TP : Machine Learning

Partie 5 : K-plus proche voisins

K-plus proche voisins (KNN)

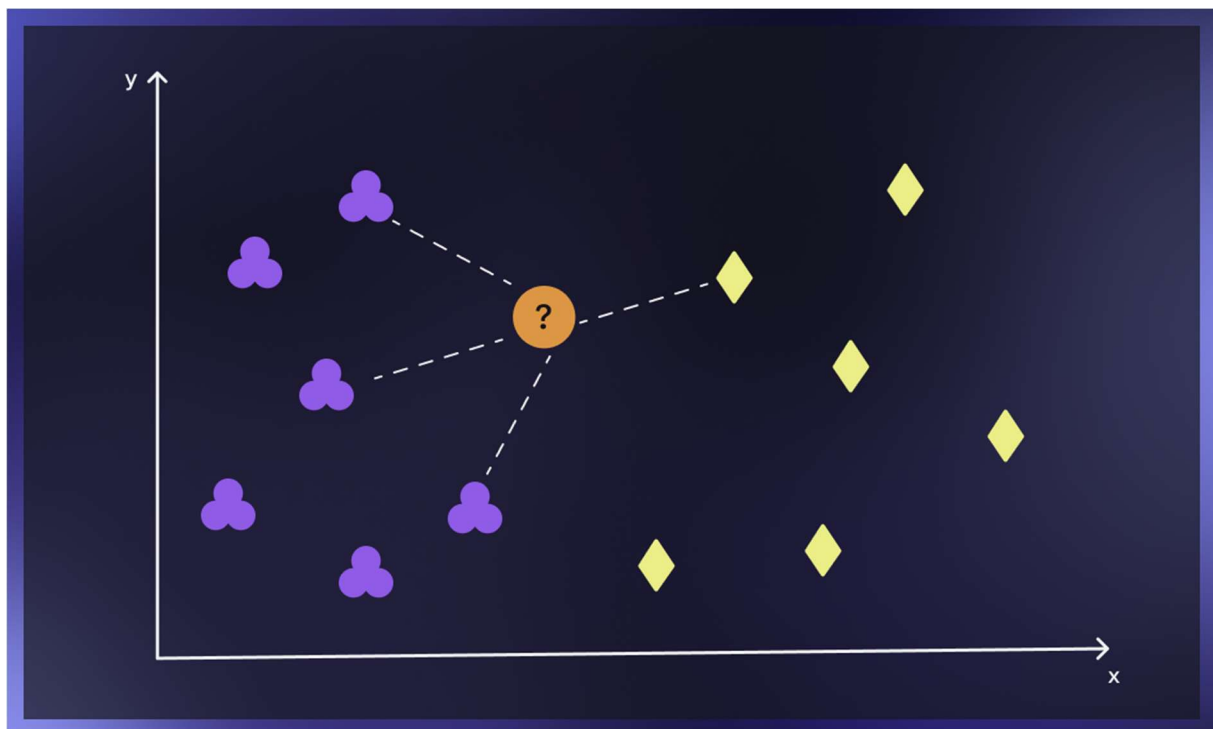
Le classificateur des k-plus proche voisins (kNN) est un algorithme d'apprentissage automatique supervisé non paramétrique. Il est basé sur la distance : il classe les objets en fonction des classes de leurs voisins les plus proches. Le kNN est le plus souvent utilisé pour la classification, mais il peut également être appliqué aux problèmes de régression.

Quelle est la valeur de k dans l'algorithme KNN ?

Le paramètre k dans kNN fait référence au nombre de points étiquetés (voisins) pris en compte pour la classification. La valeur de k indique le nombre de ces points utilisés pour déterminer le résultat. Notre tâche consiste à calculer la distance et à identifier les catégories les plus proches de notre entité inconnue.

Comment fonctionne l'algorithme de classification kNN ?

Le concept principal des kNN est le suivant. Étant donné un point dont nous ne connaissons pas la classe, nous pouvons essayer de comprendre quels sont les points de notre espace de caractéristiques qui sont les plus proches de ce point. Ces points sont les k-voisins les plus proches. Étant donné que des éléments similaires occupent des places similaires dans l'espace des caractéristiques, il est très probable que le point appartienne à la même classe que ses voisins.



Les étapes de l'algorithme KNN sont les suivantes :

- **Étape 1** : Sélection du nombre K de voisins
- **Étape 2** : Calculer la distance euclidienne de K voisins
- **Étape 3** : Prendre les K voisins les plus proches en fonction de la distance euclidienne calculée
- **Étape 4** : Parmi ces k voisins, compter le nombre de points de données dans chaque catégorie.
- **Étape 5** : Affecter les nouveaux points de données à la catégorie pour laquelle le nombre de voisins est le plus élevé.



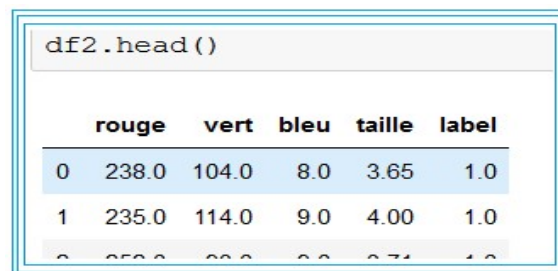
Exercice N° 1

L'objectif de l'exercice est de construire un classificateur K-plus proche voisin (KNN) en utilisant les fonctionnalités du module *sklearn*.

1. Charger L'ensemble de données 'knn_data_1.csv' et afficher le graphique en nuage de points qui montre les différentes classes du Dataset en fonction de 'attribut_1' et 'attribut_2'.
2. Divisez le Dataset en deux 70% pour le train_set et le reste pour le test_set.
3. Normaliser les données numériques
4. Construisez un modèle de K-plus proche voisin qui permet la classification des éléments du Dataset en se basent sur 'attribut_1' et 'attribut_2'.
5. Mesurez la précision du votre modèle.
6. Utilisation la validation croisée afin d'identifier la meilleure valeur pour K
7. Afficher sous forme de graphe, le score en fonction des valeurs K utilisées lors de la validation croisée.
8. Refaite la prédiction et l'évaluation avec la nouvelle valeur K trouvée

Exercice N° 2

L'objectif de ce TP est de construire un modèle de K-plus proche voisin avec le Dataset 'knn_data_2.csv' qui permet la classification des fleurs on se basant sur les couleurs et la taille.



```
df2.head()
```

	rouge	vert	bleu	taille	label
0	238.0	104.0	8.0	3.65	1.0
1	235.0	114.0	9.0	4.00	1.0
2	250.0	90.0	8.0	3.74	1.0

1. Charger L'ensemble de données 'knn_data_2.csv' et afficher le graphique en nuage de points qui montre les différentes classes du Dataset.
2. Divisez le Dataset en deux 80% pour le train_set et le reste pour le test_set.
3. Normaliser les données numériques
4. Construisez un modèle de K-plus proche voisin qui permet la classification des éléments du Dataset en se basent sur quatre attributs.
5. Mesurez la précision du votre modèle.
6. Utilisation la validation croisée afin d'identifier la meilleure valeur pour K
7. Afficher sous forme de graphe, le score en fonction des valeurs K utilisées lors de la validation croisée.
8. Refaite la prédiction et l'évaluation avec la nouvelle valeur K trouvée