

K plus proches voisins (KNN)

fiche d'aide

Arts et Metiers

A quoi sert l’algorithme ?

L’algorithme des k plus proches voisins (KNN) est utilisé à la fois pour la classification et la régression dans l’apprentissage supervisé. Ses applications incluent la classification d’images, la recommandation, la détection d’anomalies, la prédiction de valeurs, le traitement du signal, la cartographie, l’apprentissage semi-supervisé et la compression de données. KNN est particulièrement utile lorsque les données sont réparties de manière locale et que des informations provenant de voisins proches peuvent être utilisées pour prendre des décisions.

Comment fonctionne le KNN ?

L’algorithme des k plus proches voisins fonctionne en mesurant la proximité spatiale d’un nouvel exemple à k voisins les plus proches dans un ensemble de données existant. Il attribue au nouvel exemple une classe majoritaire parmi ses voisins pour la classification, ou prédit une valeur moyenne pour la régression. KNN est simple à comprendre, mais le choix de la distance et du nombre de voisins (k) peut influencer les résultats. Cet algorithme est utilisé pour des tâches de classification, de régression, de recommandation et de détection d’anomalies.

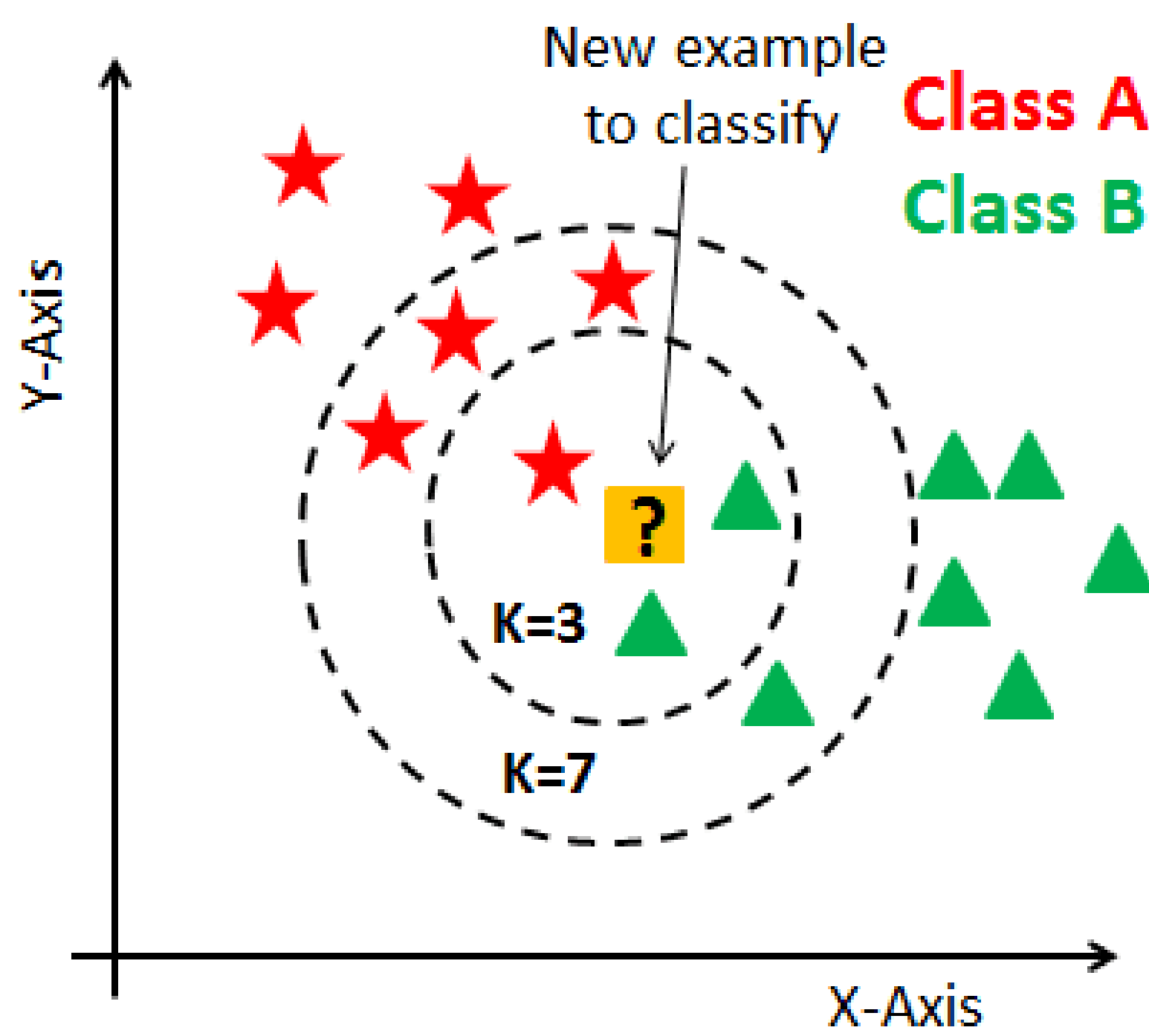


Figure 1: classification d’un element par knn

Avantages et Inconvénients

Avantages

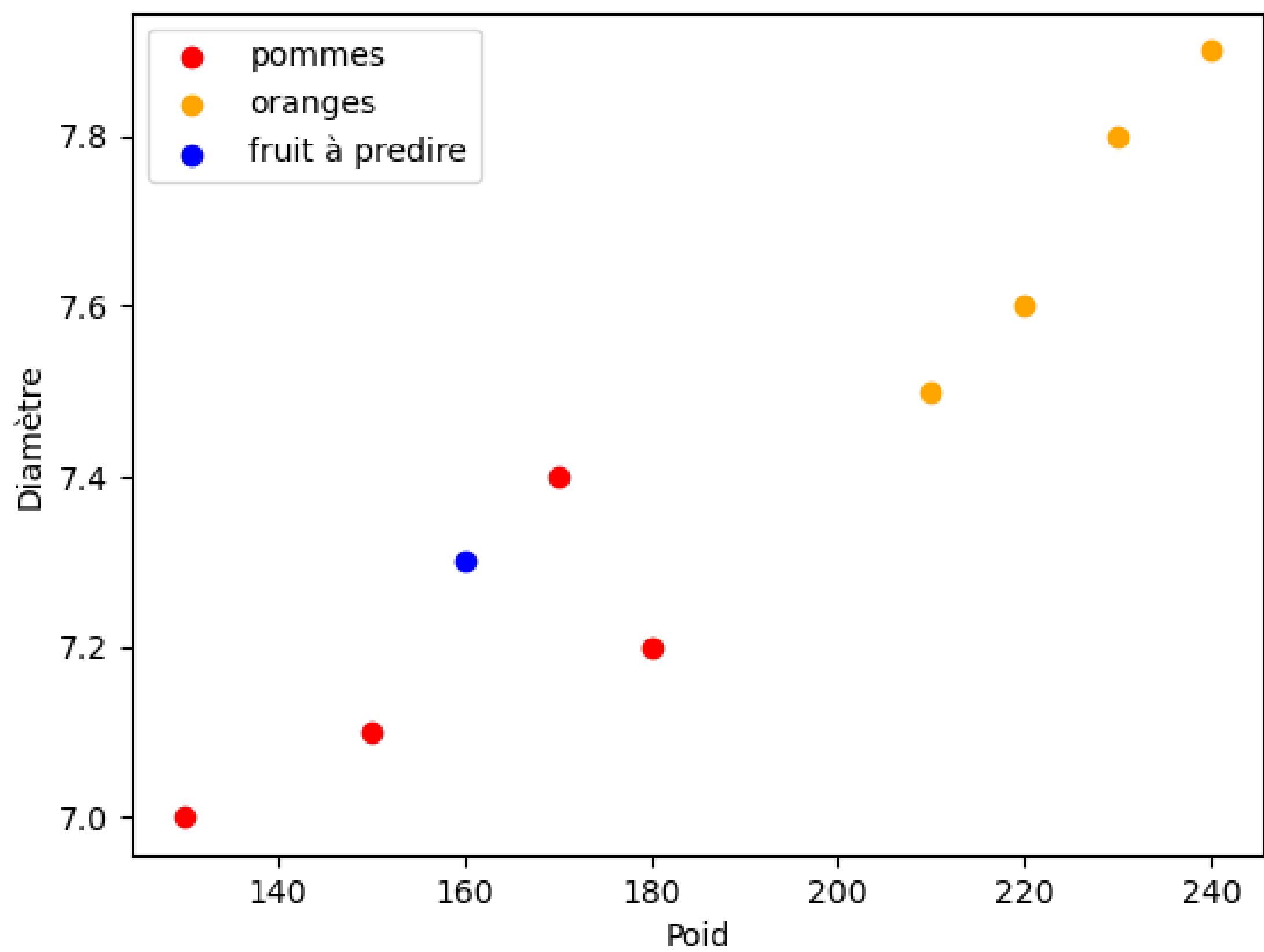
1. Facile à comprendre et à implémenter
2. Une fois entraîné, il ne nécessite pas de réentraînement lors de la modification de données
3. Classification et régression

Inconvénients

1. Sensible aux valeurs aberrantes
2. Demande beaucoup de ressources pour calculer la distance entre les points
3. Besoin de beaucoup de stockage pour les données
4. Choix de k qui influe sur les résultats
5. Ne fonctionne pas bien avec des données de grande dimension

Exemple

On cherche à savoir si un fruit est une pomme ou une orange en fonction de son poids et de son diamètre. On a les données suivantes :



ici, l’algorithme renvoie que le fruit est une pomme.

Pour aller plus loin

1. Scikit-learn.
2. Wikipedia