

Activité débranchée : k-NN

Les k Plus Proches Voisins

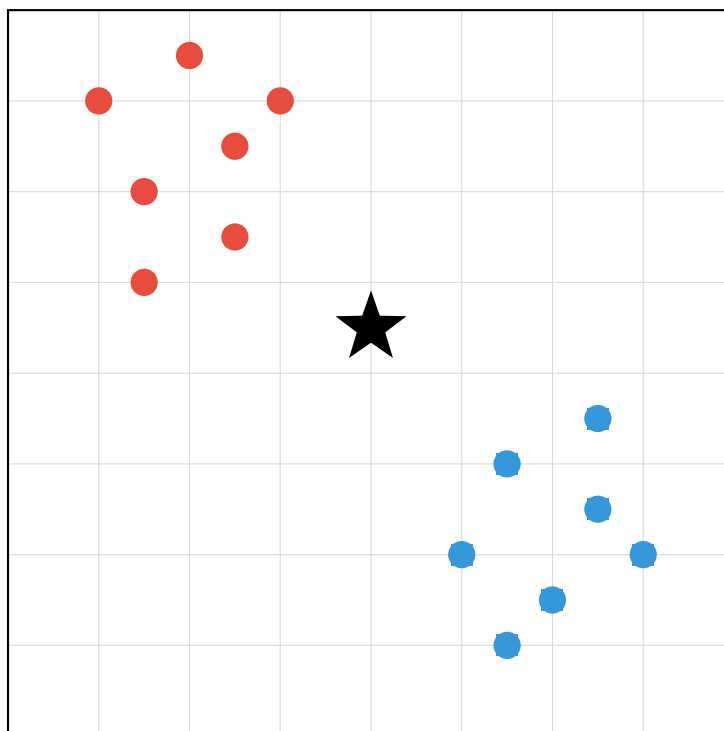
Objectif : Comprendre comment l'algorithme k-NN classe un nouveau point en utilisant la proximité avec les points d'apprentissage.

Principe

- On dispose d'un ensemble de points déjà classifiés (points d'apprentissage)
- Pour classier un nouveau point, on cherche ses k plus proches voisins
- La classe majoritaire parmi ces k voisins devient la classe du nouveau point

Exercice 1 : Classification avec $k = 3$

Voici un ensemble de points sur un plan. Les cercles rouges (●) sont de classe A, les carrés bleus (■) sont de classe B. Le point noir (★) est à classier.



● Classe A ■ Classe B ★ À classier

Questions :

1. Identifiez les 3 points les plus proches du point noir ★. Tracez les distances sur le graphique ci-dessus.

Distances : _____

2. Parmi ces 3 points, combien sont de classe A ? _____ De classe B ? _____
3. À quelle classe appartient donc le point noir avec $k = 3$? _____

Exercice 2 : Influence du paramètre k

Reprenez le même graphique que l'exercice 1.

a) Avec $k = 1$ (1 seul voisin)

- Quel est le point le plus proche ? _____
- Classification du point noir : _____

b) Avec $k = 5$ (5 voisins)

- Listez les 5 points les plus proches : _____

- Nombre de classe A : _____ Nombre de classe B : _____
- Classification du point noir : _____

c) Avec $k = 7$ (tous les voisins de chaque classe)

- Nombre de classe A : _____ Nombre de classe B : _____
- Classification du point noir : _____

Réflexion

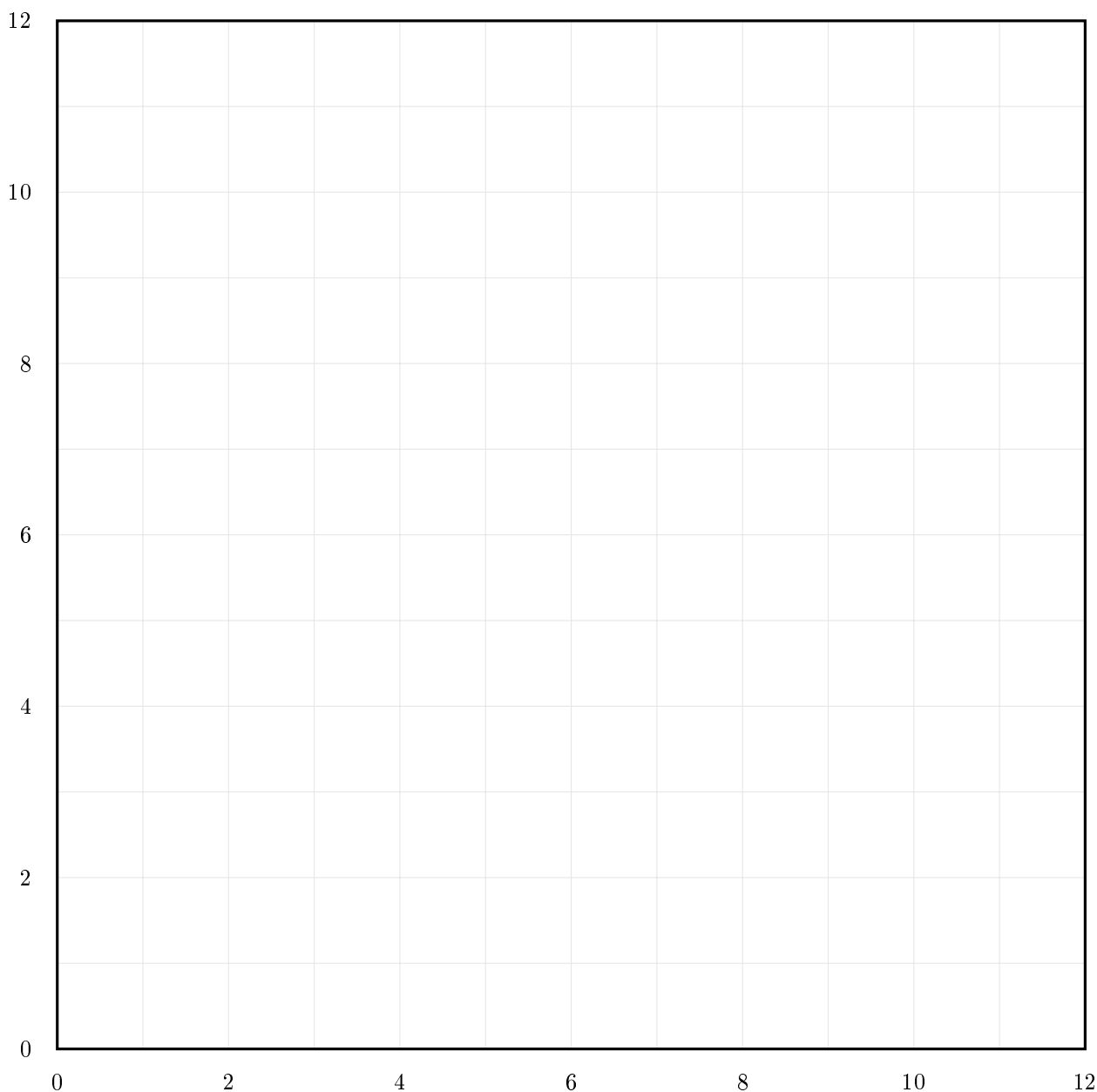
1. La classification change-t-elle selon la valeur de k ? Pourquoi ?

2. Que se passe-t-il si k est trop petit ($k = 1$) ?

3. Que se passe-t-il si k est trop grand ($k = \text{nombre total de points}$) ?

Exercice 3 : À vous de jouer !

Créez votre propre situation de classification avec $k = 3$.



Instructions

- Placez au moins 10 points : 5 cercles rouges (classe A) et 5 carrés bleus (classe B)
- Placez une étoile noire (point à classifier)
- Tracez les 3 distances les plus courtes
- Déterminez la classe du point noir

Classification finale : _____

Pour aller plus loin

Applications réelles de k-NN

- Reconnaissance d'écriture manuscrite
- Systèmes de recommandation (films, musique, produits)
- Diagnostic médical (classification de maladies)
- Détection de spam dans les emails
- Reconnaissance faciale
- Prévion météorologique

Questions de réflexion

1. Avantages de k-NN :

- Simple à comprendre et à implémenter
- Pas besoin d'entraînement complexe
- Fonctionne bien pour des frontières non linéaires

2. Limitations de k-NN :

- Lent avec beaucoup de données (doit calculer toutes les distances)
- Sensible aux données aberrantes
- Nécessite de choisir la bonne valeur de k
- Ne fonctionne pas bien si les classes sont déséquilibrées

3. Comment choisir k ?

- k trop petit ($k = 1$) : sensible au bruit, risque de sur-apprentissage
- k trop grand : perd les détails, risque de sous-apprentissage
- Conseil : tester plusieurs valeurs et choisir celle qui donne les meilleurs résultats
- Souvent, k impair pour éviter les égalités ($k = 3, 5, 7...$)

Défi bonus

Imaginez une situation où k-NN pourrait donner un mauvais résultat. Dessinez cette situation sur une feuille séparée et expliquez pourquoi.

Indices : Pensez aux données bruitées, aux classes déséquilibrées, ou aux points isolés...