

Activité débranchée : k-NN

Classification par micro-régions

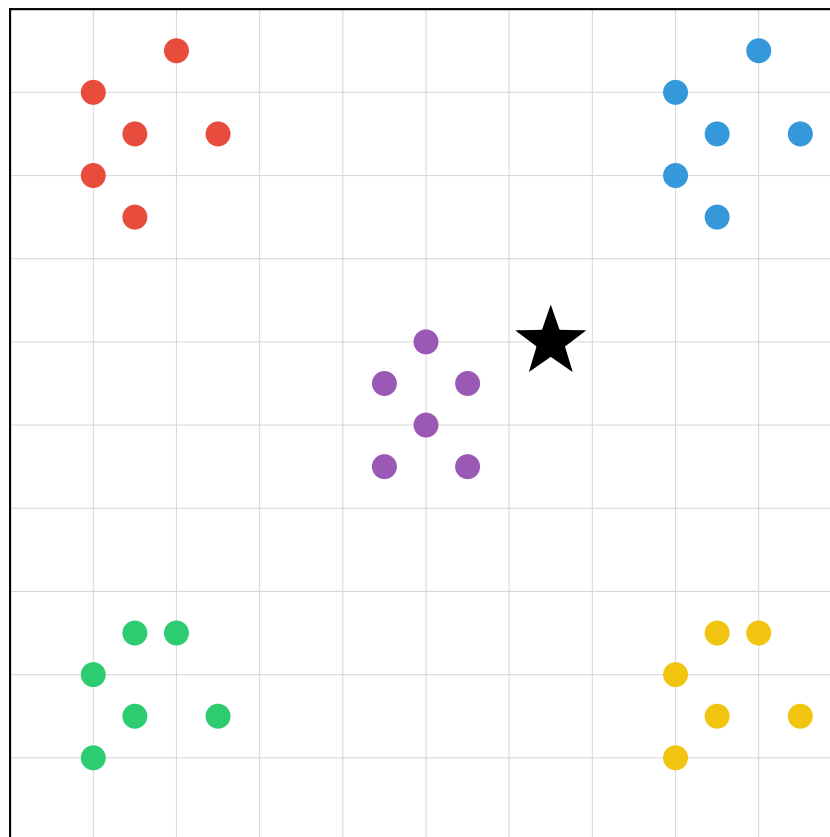
Objectif : Utiliser l'algorithme k-NN pour déterminer à quelle micro-région appartient un nouveau point, en se basant sur ses plus proches voisins.

Principe

- On dispose de points classifiés en plusieurs *zones* (micro-régions)
- Pour classier un nouveau point, on identifie ses k plus proches voisins
- La zone majoritaire parmi ces k voisins devient la classification du point
- **Différence avec 2 classes :** Avec plusieurs zones, les votes sont plus complexes !

Exercice 1 : Classification avec 5 zones

Voici des points répartis en 5 zones de couleurs différentes. Le point noir (★) est à classier.



● Zone 1 ● Zone 2 ● Zone 3 ● Zone 4 ● Zone 5

Questions avec $k = 3$:

1. Identifiez les 3 points les plus proches du point noir ★. Tracez les distances.

Les 3 plus proches : Zone _____, Zone _____, Zone _____

2. Comptez les votes pour chaque zone parmi ces 3 voisins :

Zone 1 : _____ Zone 2 : _____ Zone 3 : _____

Zone 4 : _____ Zone 5 : _____

3. Quelle est la zone majoritaire? _____

4. Y a-t-il égalité entre plusieurs zones? _____

Exercice 2 : Influence de k avec plusieurs classes

Reprenez le même graphique que l'exercice 1.

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

— Listez les 5 points les plus proches :

Zone _____, Zone _____, Zone _____,

Zone _____, Zone _____

— Votes par zone :

Zone 1 : _____ Zone 2 : _____ Zone 3 : _____

Zone 4 : _____ Zone 5 : _____

— Classification du point noir : Zone _____

b) Avec $k = 7$ (7 voisins)

— Votes par zone :

Zone 1 : _____ Zone 2 : _____ Zone 3 : _____

Zone 4 : _____ Zone 5 : _____

— Classification du point noir : Zone _____

Réflexion : Plusieurs classes

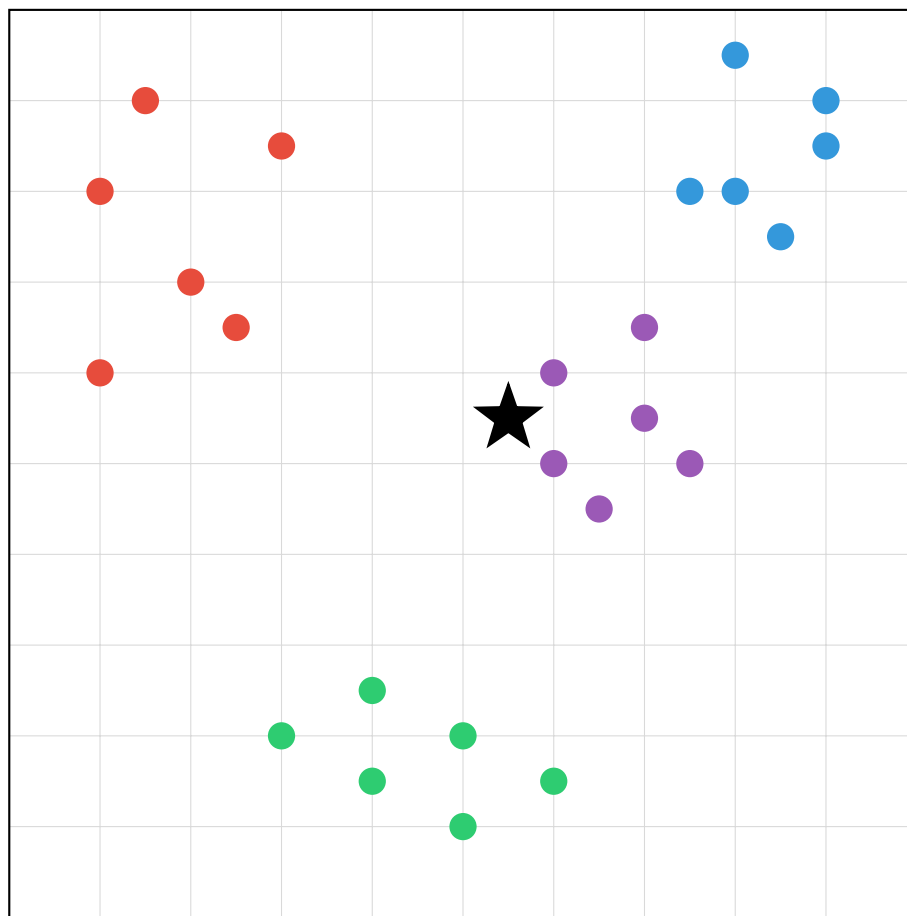
1. Qu'est-ce qui change par rapport à 2 classes seulement ?

2. Peut-il y avoir des égalités entre zones ? Donnez un exemple.

3. Comment résoudre une égalité ?

Exercice 3 : Cas complexe avec 4 zones

Voici une nouvelle situation avec 4 zones différentes. Classifiez le point \star avec $k = 5$.



● Zone A ● Zone B ● Zone C ● Zone D

Travail à faire :

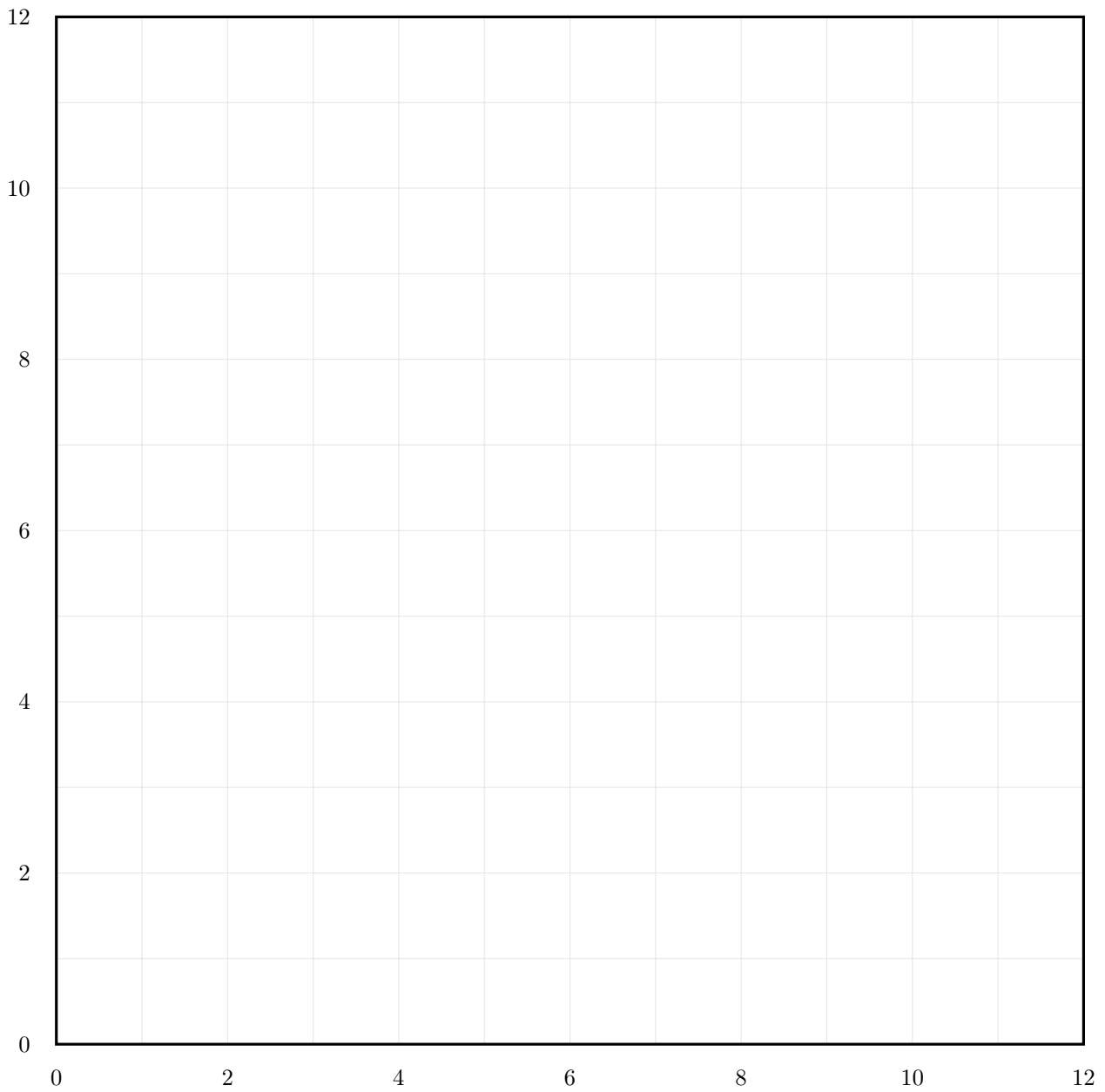
1. Identifiez les 5 plus proches voisins et tracez les distances sur le graphique
2. Complétez le tableau de votes :

Zone A	Zone B	Zone C	Zone D

3. Classification finale : Zone _____
4. Cette classification vous semble-t-elle cohérente visuellement ? Pourquoi ?

Exercice 4 : Création libre

Créez votre propre situation avec au moins 3 zones différentes.



Instructions

- Choisissez 3 à 5 zones (utilisez des couleurs différentes)
- Placez au moins 5 points par zone
- Placez une étoile noire (point à classifier)
- Utilisez $k = 5$ et déterminez la zone du point noir

Vos zones :

Zone 1 : _____ (couleur : _____)

Zone 2 : _____ (couleur : _____)

Zone 3 : _____ (couleur : _____)

Classification avec $k = 5$: Zone _____

Pour aller plus loin

Applications avec plusieurs classes

L'algorithme k-NN avec plusieurs classes est utilisé pour :

- **Reconnaissance de chiffres** : Classifier 0, 1, 2, ..., 9 (10 classes)
- **Classification de fleurs** : Différentes espèces (iris, rose, tulipe...)
- **Zonage géographique** : Micro-régions, quartiers, zones climatiques
- **Diagnostic médical** : Plusieurs types de maladies possibles
- **Reconnaissance vocale** : Identifier différents phonèmes ou mots

Défis avec plusieurs classes

1. **Égalités plus fréquentes** : Avec 2 classes, égalité rare. Avec 5 classes, beaucoup plus probable !
2. **Choix de k important** : Si k est trop petit, risque de ne pas capturer la diversité. Si trop grand, perd la précision locale.
3. **Classes déséquilibrées** : Si une zone a beaucoup plus de points qu'une autre, elle sera sur-représentée.
4. **Frontières complexes** : Avec plusieurs zones, les frontières peuvent être très irrégulières.

Question finale

Pourquoi est-il préférable d'utiliser un k impair lorsqu'on a un nombre pair de classes ?
