

K-NN

K NEAREST NEIGHBORS

K PLUS PROCHES VOISINS

GÉNÉRALITÉS

- Apprendre par analogie

Recherchant d'un ou des cas similaires déjà résolus

- Classifier ou estimer

“Dis moi qui sont tes amis, et je te dirais qui tu es”

- Pas de construction de modèle

C'est l'échantillon d'apprentissage, associé à une fonction de distance et d'une fonction de choix de la classe en fonction des classes des voisins les plus proches, qui constitue le modèle

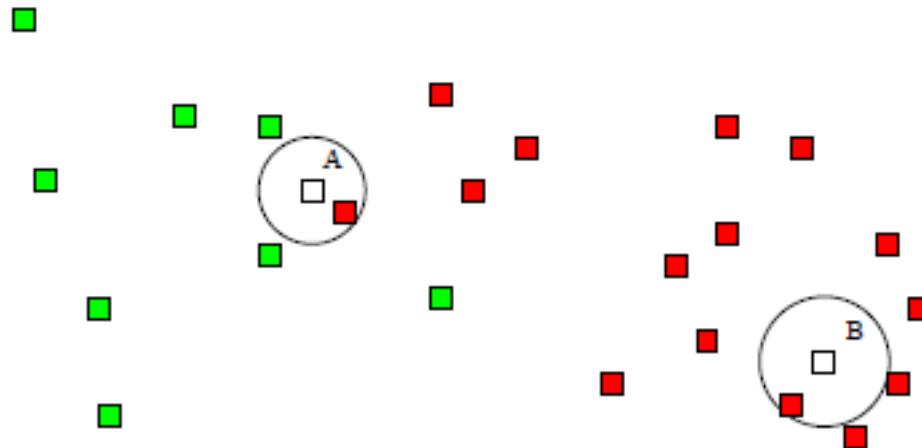
PRINCIPE DES K PLUS PROCHES VOISINS

Principe :

- ✓ Regarder la classe des k exemples les plus proches ($k = 1, 3, \dots$)
- ✓ Affecter la classe majoritaire au nouvel exemple

Exemple :

- ✓ Deux classes : verte et rouge
- ✓ $k = 1$



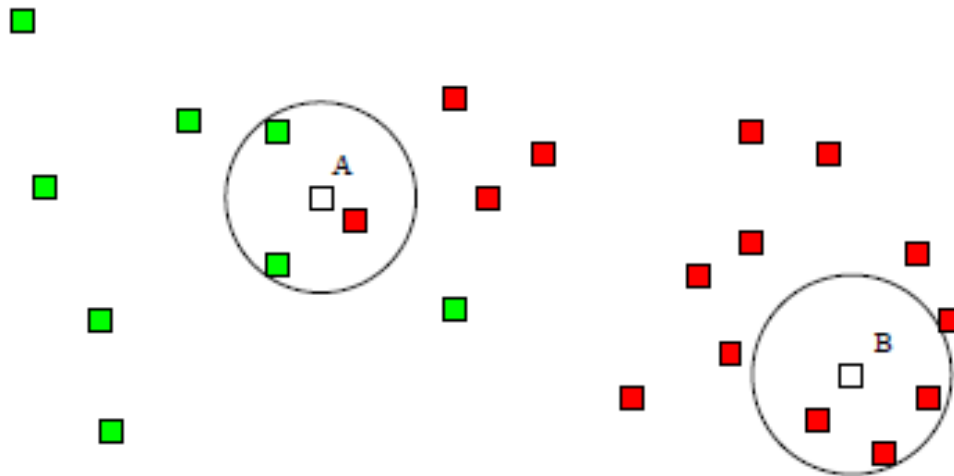
PRINCIPE DES K PLUS PROCHES VOISINS

Principe :

- ✓ Regarder la classe des k exemples les plus proches ($k = 1, 3, \dots$)
- ✓ Affecter la classe majoritaire au nouvel exemple

Exemple :

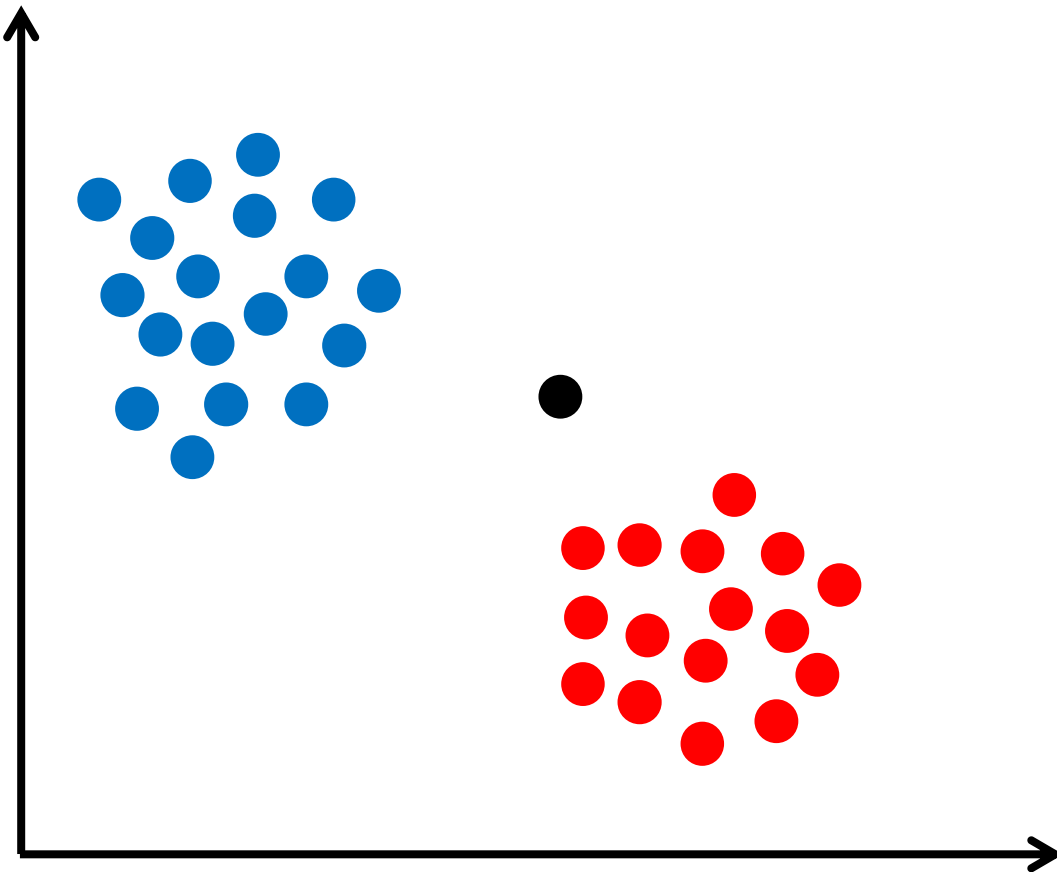
- ✓ Deux classes : verte et rouge
- ✓ $k = 3$



PRINCIPE DES K PLUS PROCHES VOISINS

objectifs :

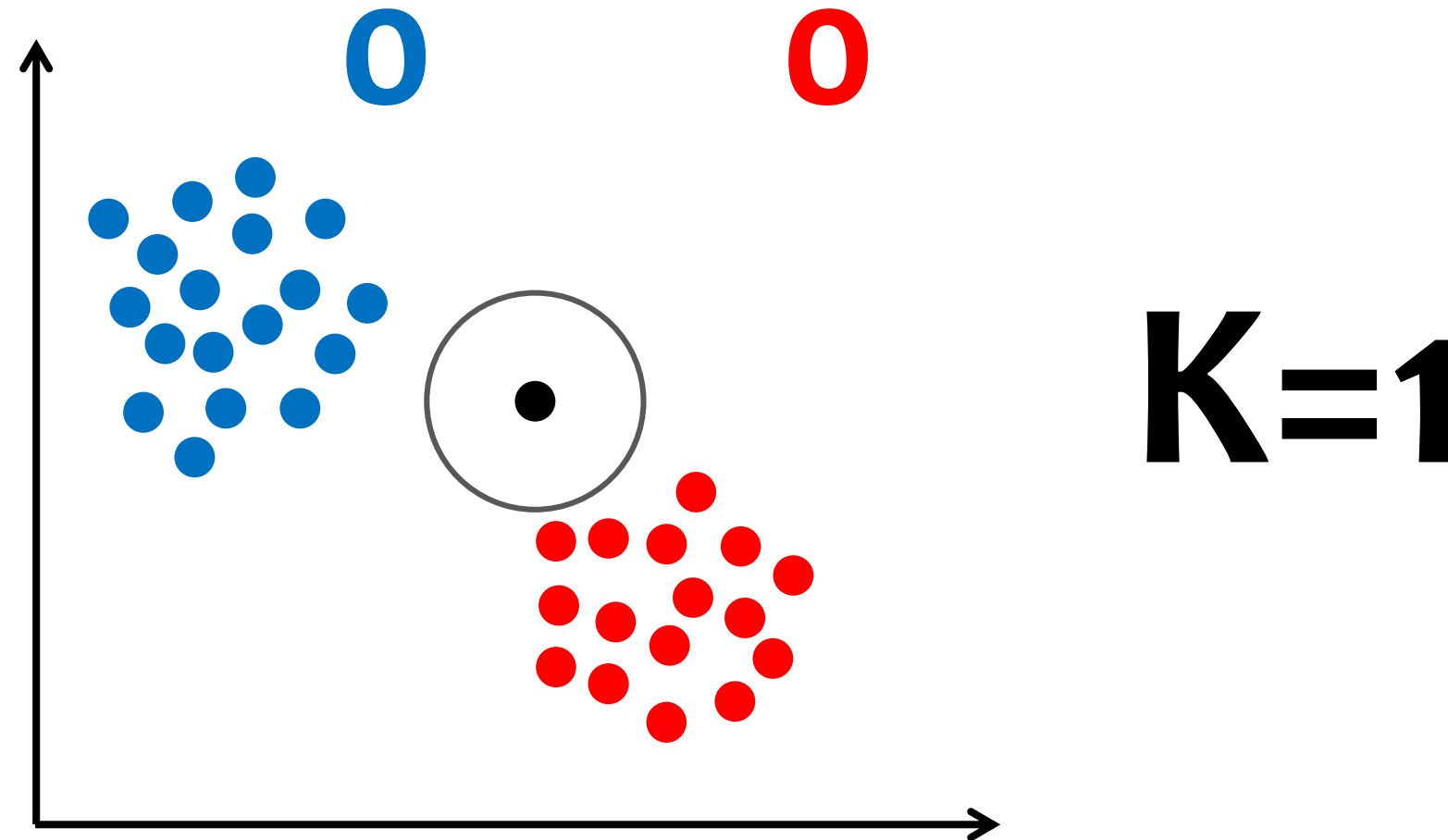
✓ Prédire la classe d'appartenance du point noir ?



PRINCIPE DES K PLUS PROCHES VOISINS

objectifs :

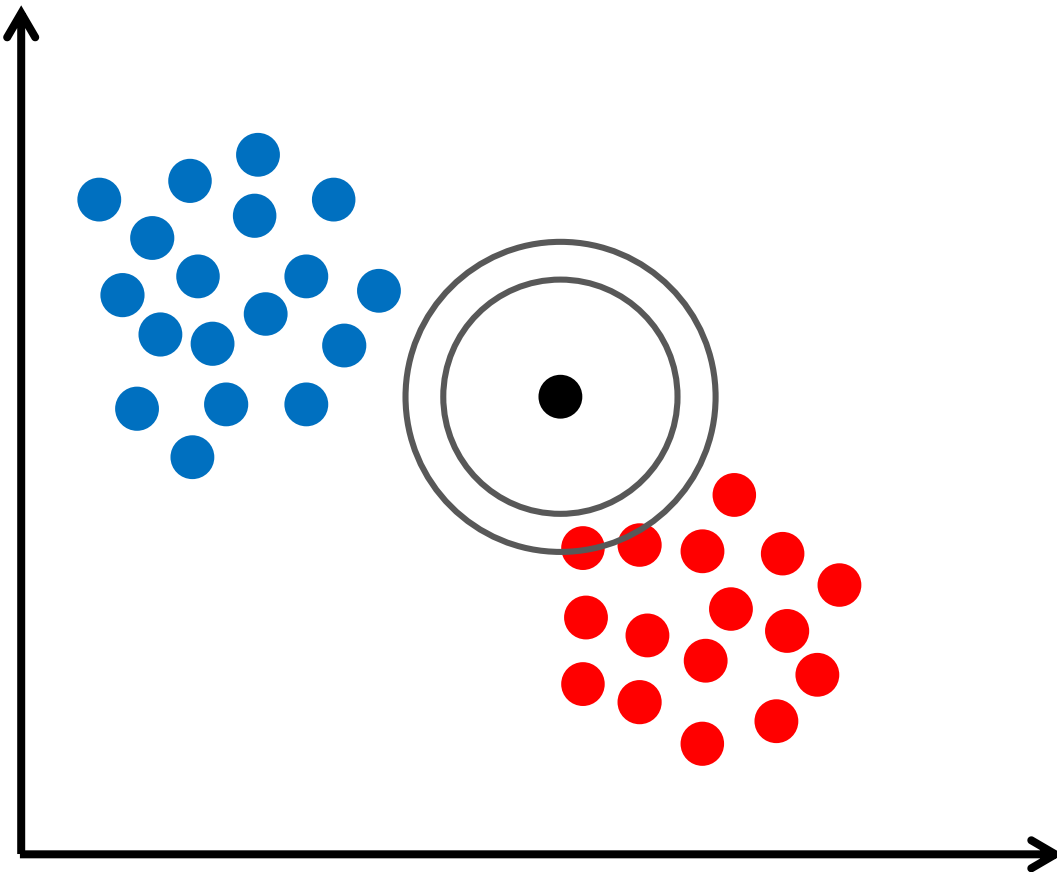
✓ Prédire la classe d'appartenance du point noir



PRINCIPE DES K PLUS PROCHES VOISINS

objectifs :

✓ Prédire la classe d'appartenance du point noir

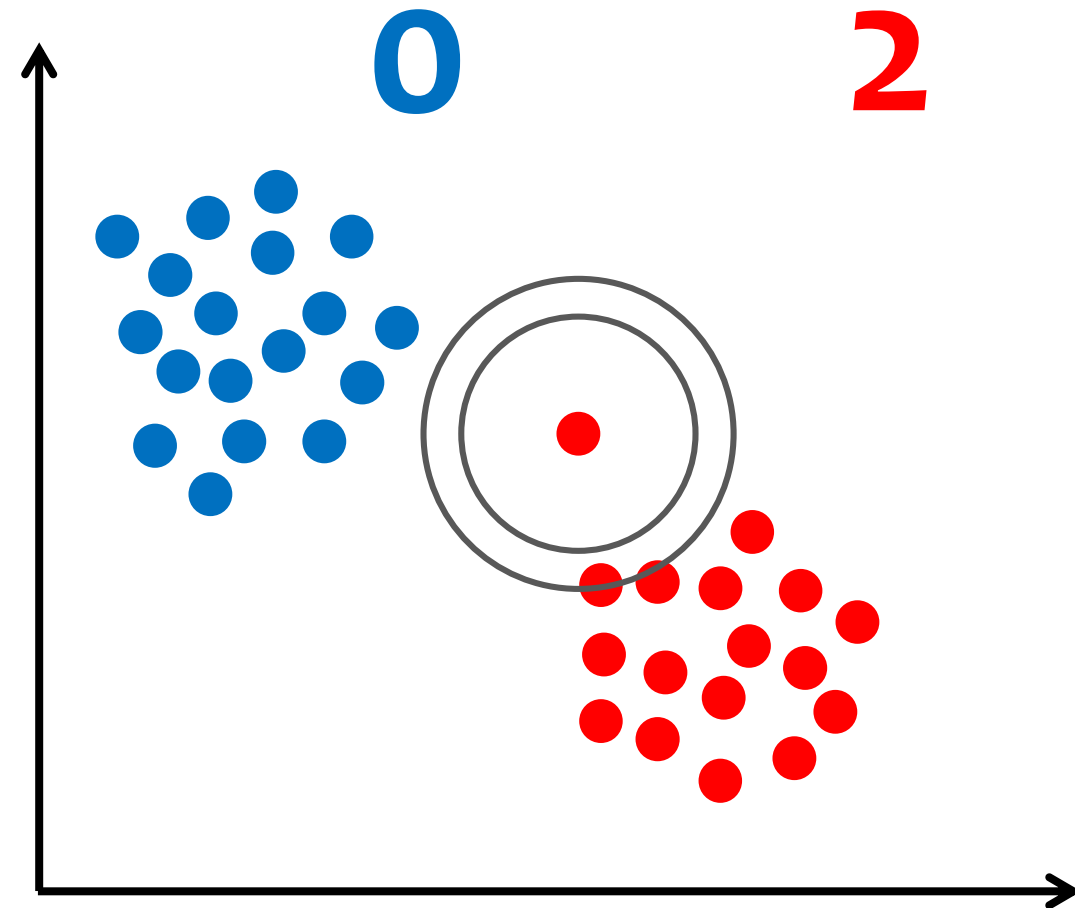


K=2

PRINCIPE DES K PLUS PROCHES VOISINS

objectifs :

✓ Prédire la classe d'appartenance du point noir

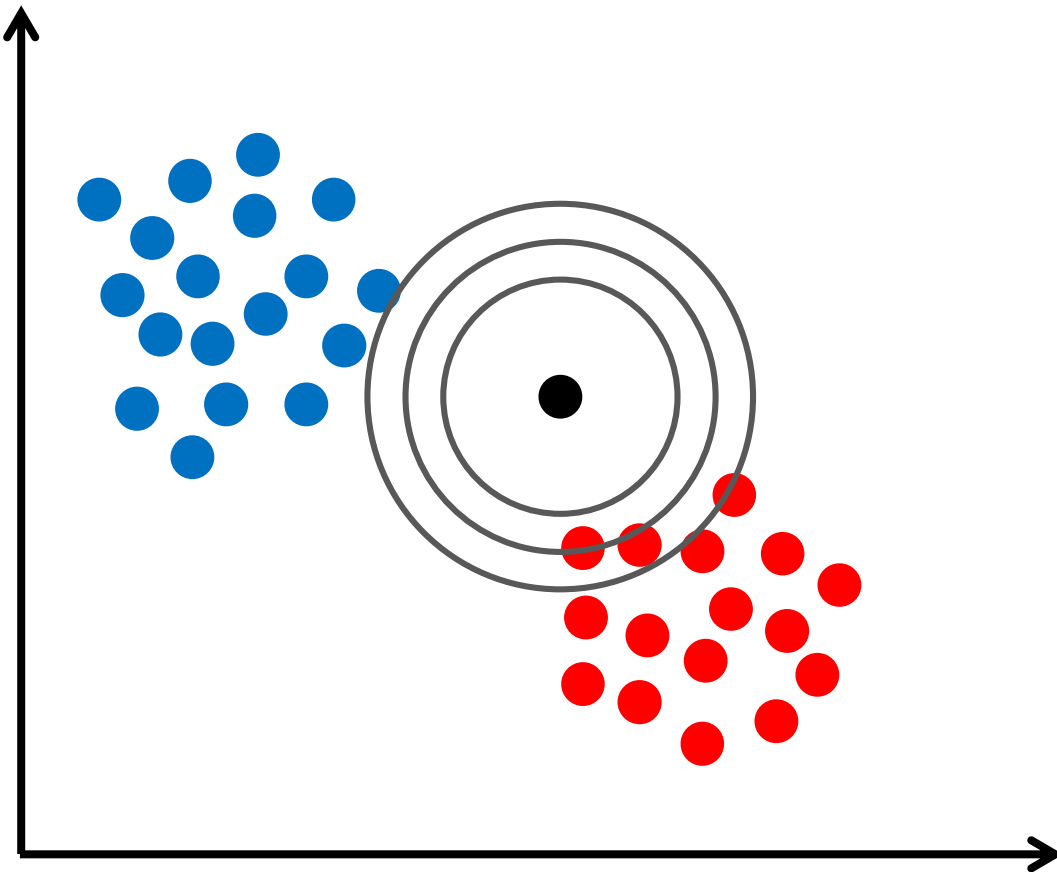


K=2

PRINCIPE DES K PLUS PROCHES VOISINS

objectifs :

✓ Prédire la classe d'appartenance du point noir

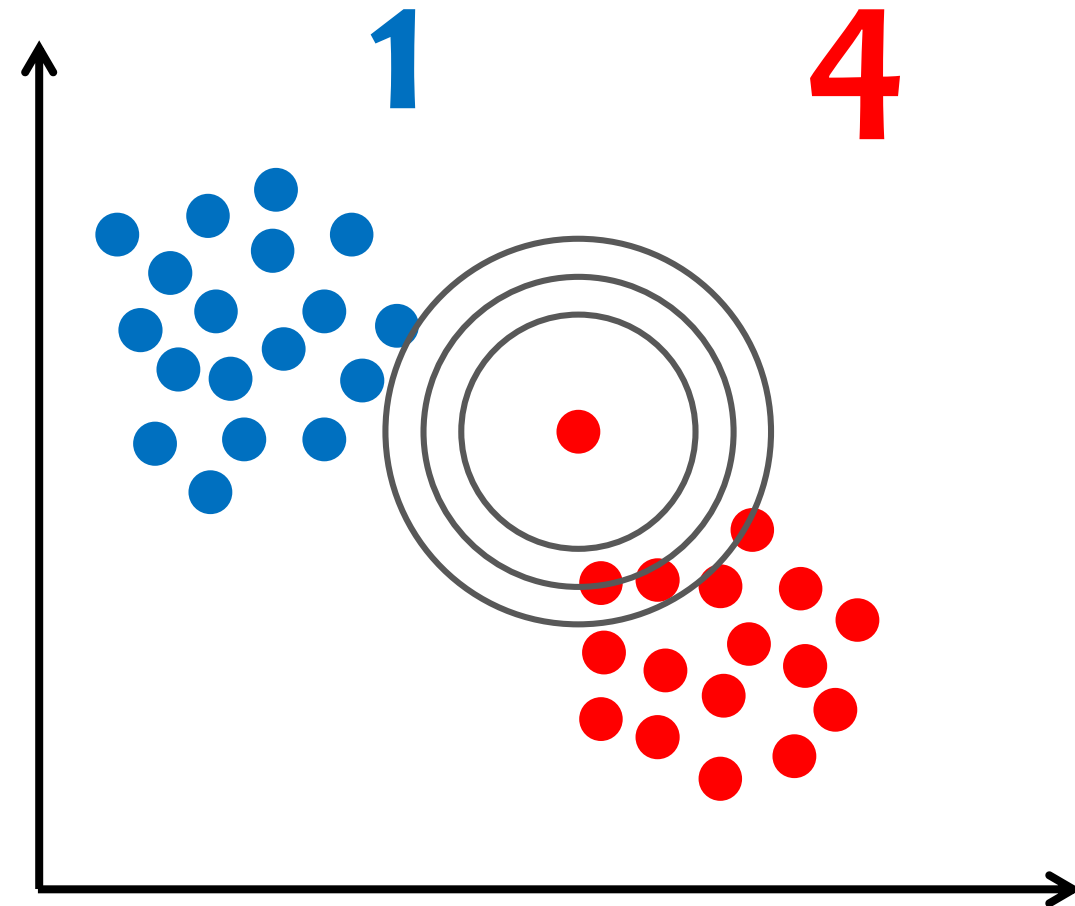


K=3

PRINCIPE DES K PLUS PROCHES VOISINS

objectifs :

✓ Prédire la classe d'appartenance du point noir

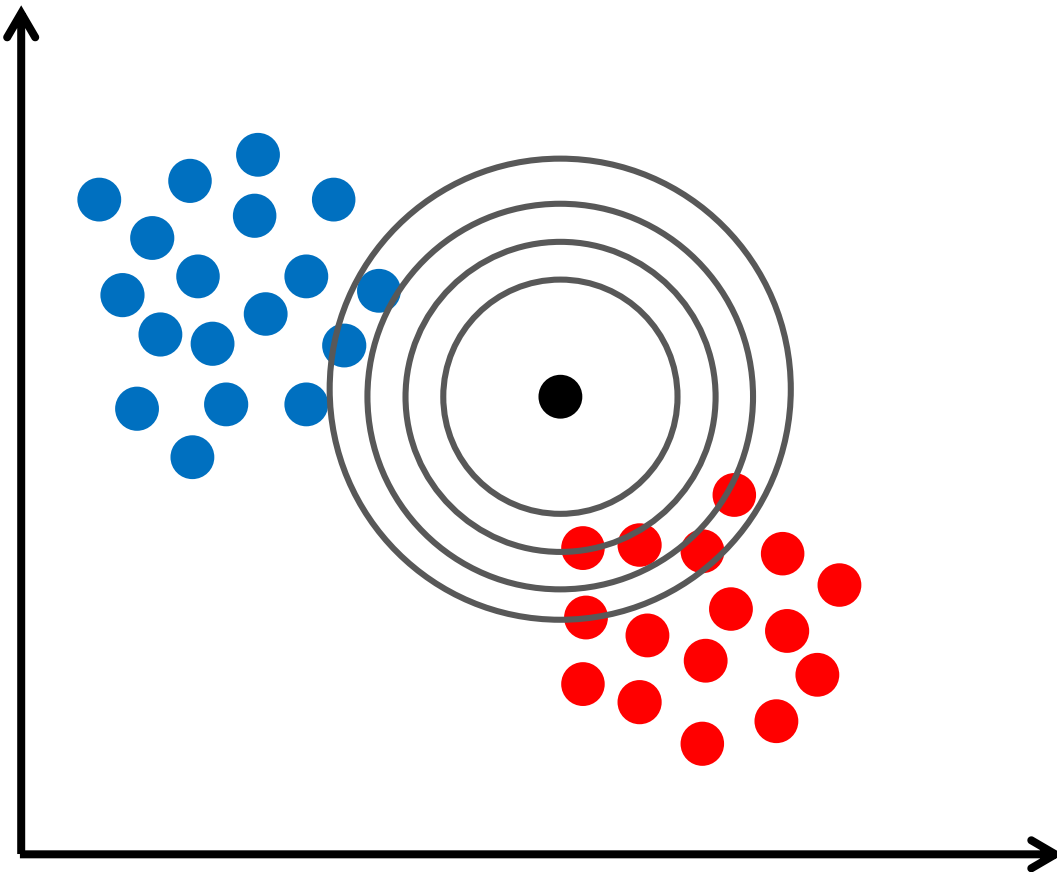


K=3

PRINCIPE DES K PLUS PROCHES VOISINS

objectifs :

✓ Prédire la classe d'appartenance du point noir

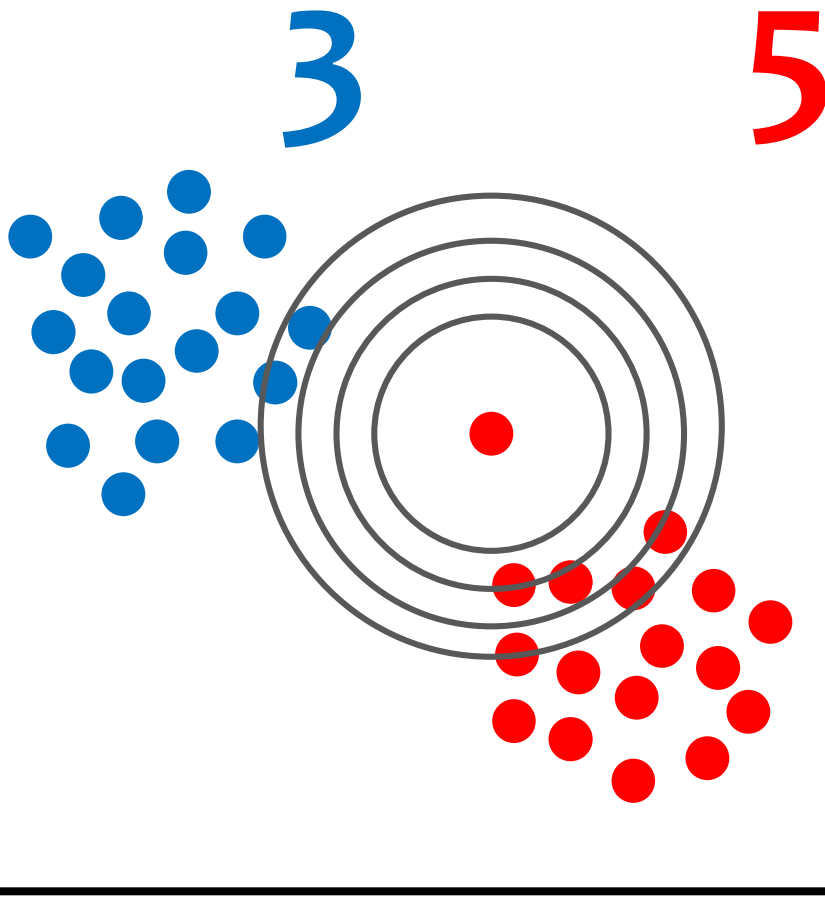


K=4

PRINCIPE DES K PLUS PROCHES VOISINS

objectifs :

✓ Prédire la classe d'appartenance du point noir



K=4

ALGORITHME

notations :

- Soit $L = \{(x', c) \mid x' \in R^d, c \in C\}$ l'ensemble d'apprentissage
- Soit x l'exemple qu'on désire déterminer la classe d'appartenance

algorithme kNN

DEBUT kNN

POUR chaque exemple de $(x', c) \in L$ **FAIRE**
 calculer la distance $D(x, x')$

FIN POUR

POUR chaque $\{x' \in kppv(x)\}$ **FAIRE**
 compter le nombre d'occurrences de chaque classe

FIN POUR

attribuer à x la classe la plus fréquente

FIN kNN

FORCE ET FAIBLESSE DU KNN

Les attributs ont le même poids

centrer et réduire pour éviter les biais

certaines peuvent être moins classant que d'autres

Apprentissage paresseux

rien n'est préparé avant le classement

tous les calculs sont fait lors du classement

nécessité de technique d'indexation pour large BD

Calcul du score d'une classe

peut changer les résultats; variantes possibles