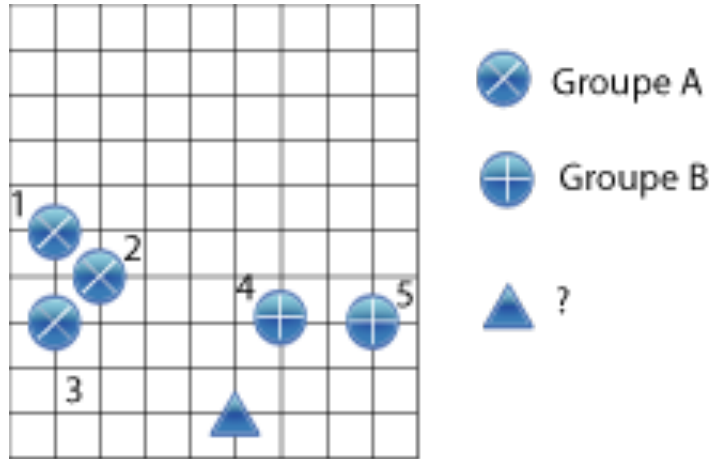


Algorithme des plus proches voisins

1. Le principe



On considère 2 groupes déjà constitués : les groupes A et B.

Un nouvel élément apparaît, comment déterminer à quel groupe il appartient ?

Une méthode souvent utilisée pour ce genre de problématique est l'algorithme des k-plus proches voisins ou **k-nearest neighbors** (k-NN ou KNN).

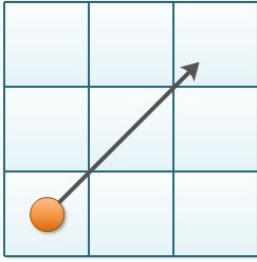
L'algorithme :

- il suppose que les éléments précédents sont déjà classés (selon un classement pré-défini ou selon un classement en cours d'élaboration),
- on fixe une "distance",
- on fixe un nombre k entier,
- on calcule la distance entre chaque élément déjà classé et le nouvel élément,
- on détermine alors les k plus proches voisins du nouvel élément,
- on attribue au nouvel élément la classe majoritaire parmi ces k plus proches voisins.

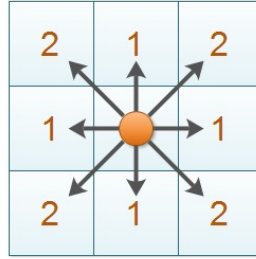
2. Mise en oeuvre de l'algorithme pour différentes distances.

On peut définir en Mathématiques différentes distances :

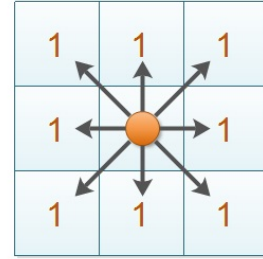
Euclidean Distance



Manhattan Distance



Chebyshev Distance



$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad |x_1 - x_2| + |y_1 - y_2| \quad \max(|x_1 - x_2|, |y_1 - y_2|)$$

- Distance euclidienne :

En considérant la distance euclidienne, déterminer à quelle classe appartient le nouvel élément pour respectivement $k = 1, 3, 5$ voisins.

Calcul des distances :

distance à l'élément 1 :

distance à l'élément 2 :

distance à l'élément 3 :

distance à l'élément 4 :

distance à l'élément 5 :

Plus proches voisins pour $k =$:

k= 1	k=3	k= 5

- Distance de Manhattan

En considérant la distance de Manhattan, déterminer à quelle classe appartient le nouvel élément pour respectivement $k = 1, 3, 6$ voisins.

Calcul des distances :

distance à l'élément 1 :

distance à l'élément 2 :

distance à l'élément 3 :

distance à l'élément 4 :

distance à l'élément 5 :

Plus proches voisins pour $k =$:

$k= 1$	$k=3$	$k= 5$

- Distance de Chebyshev

En considérant la distance de Chebyshev, déterminer à quelle classe appartient le nouvel élément pour respectivement $k = 1, 3, 6$ voisins.

Calcul des distances :

distance à l'élément 1 :

distance à l'élément 2 :

distance à l'élément 3 :

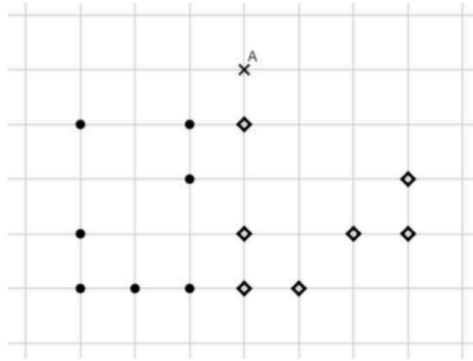
distance à l'élément 4 :

distance à l'élément 5 :

Plus proches voisins pour $k =$:

$k= 1$	$k=3$	$k= 5$

Exercice 1



Dans le quadrillage ci-dessus 14 points sont dessinés, dont 7 de la classe C1, avec des ronds noirs •, et 7 de la classe C2, avec des losanges ◇.

On introduit un nouveau point A, dont on cherche la classe à l'aide d'un algorithme des k plus proches voisins pour la distance géométrique habituelle, en faisant varier la valeur de k parmi 1, 3 et 5.

Quelle est la bonne réponse (sous la forme d'un triplet de classes pour le triplet (1,3,5) des valeurs de k) ?

Réponses possibles

- A C1, C2, C3
- B C2, C1, C2
- C C2, C2, C2
- D C2, C1, C1

3. Usage de cet algorithme

La méthode des k plus proches voisins est une méthode d'apprentissage supervisée en intelligence artificielle.

Elle suppose l'existence au préalable d'une base de données d'apprentissage constituée de N couples "entrée-sortie".

Elle peut être utilisée pour :