Apprentissage non supervisé

Quentin Fortier

January 21, 2026

Classification supervisée/non supervisée

Définition

Un algorithme de classification est un algorithme qui permet d'associer à chaque donnée une classe (une espèce de fleur, un chiffre...)

Classification supervisée/non supervisée

Définition

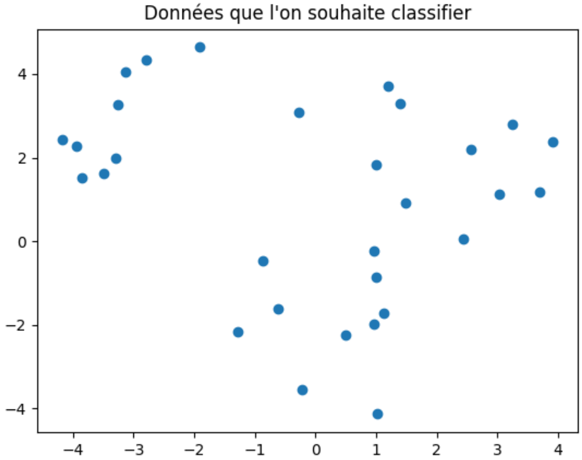
Un algorithme de classification est un algorithme qui permet d'associer à chaque donnée une classe (une espèce de fleur, un chiffre...)

Il y a deux types d'algorithmes de classification :

- Classification supervisée : on connaît les classes de certaines données (données d'entraînement) qui permettent de prédire la classe d'une nouvelle donnée.
  Exemples : k plus proches voisins, ID3.
- Classification non supervisée : Il n'y a pas de donnée d'entraînement et l'ensemble des classes possibles n'est pas connu à l'avance.
  Exemples : k-moyennes, classification hiérarchique ascendante.

Algorithme des k-moyennes : Principe général

Exemple : il semble y avoir $k = 3$ classes de données parmi ces points.



Algorithme des k-moyennes : Principe général

Définition

Le centre (ou : isobarycentre) d'un ensemble de vecteurs $x_1, \ldots, x_n$ est le vecteur

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Algorithme des k-moyennes : Principe général

Soit X un ensemble de données et un entier k.

Inertie

On veut trouver une partition $\mathcal{P}$ de X en k sous-ensembles $X_1, \ldots, X_k$ (classes ou *clusters*) minimisant l'inertie :

$$I(\mathcal{P}) = \sum_{i=1}^{k} \sum_{x \in X_i} d(x, \overline{X_i})^2$$

Plus l'inertie est petite, plus les données sont proches du centre de leur classe et plus le partitionnement est bon.

Algorithme des k-moyennes : Principe général

Algorithme des k-moyennes

Entrée : Des données X, un entier k
Sortie : Une partition de X en k classes

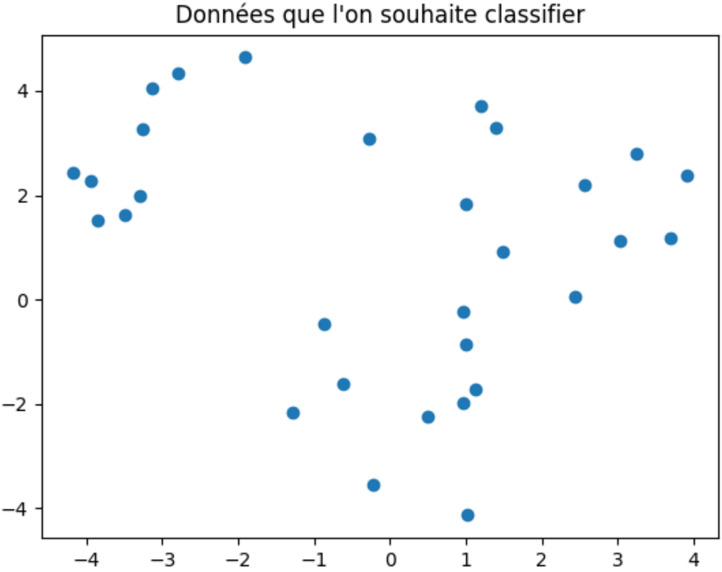
Soient $c_1, ..., c_k$ des vecteurs (centres) choisis aléatoirement
Tant que les centres ont changé **:**
- Associer chaque donnée x à la classe X_i telle que $d(x, c_i)$ soit minimum
- Recalculer les centres des classes $c_i = \overline{X_i}$

Renvoyer $X_1, \ldots, X_k$

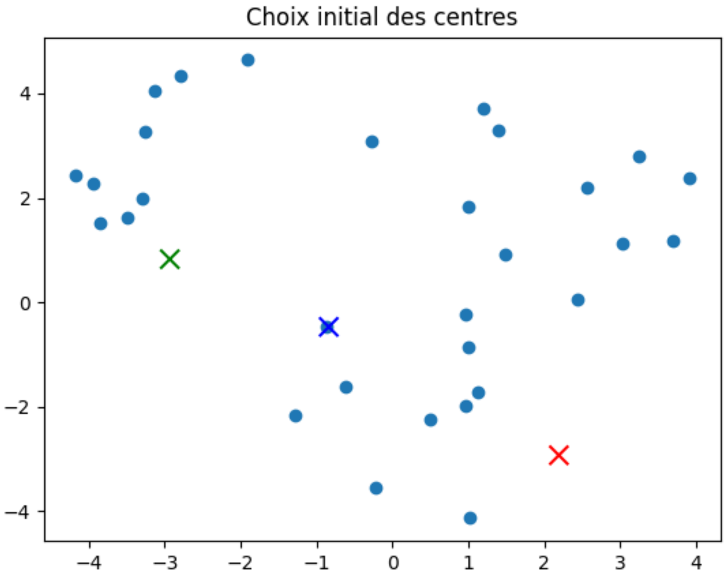
Algorithme des k-moyennes : Principe général

- On peut choisir les centres initiaux aléatoirement dans R^p ou parmi X.
- k est le nombre de classes dans l'algorithme des k-moyennes alors que c'est le nombre de voisins dans l'algorithme des k plus proches voisins.
- Le problème de décision consistant à déterminer s'il existe une partition d'inertie inférieure à un seuil est NP-complet.

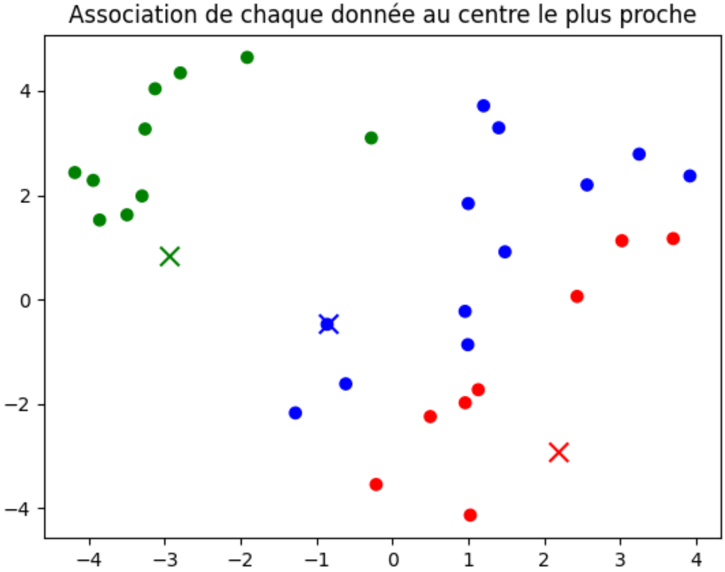
Algorithme des k-moyennes : Exemple



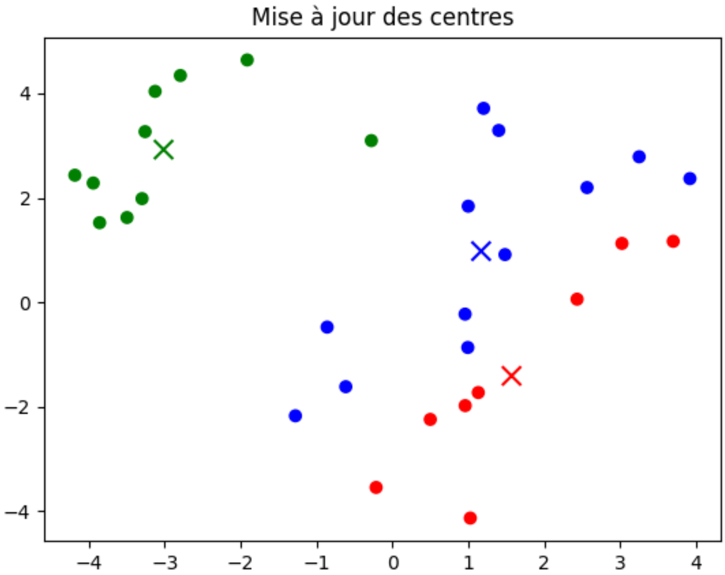
Algorithme des k-moyennes : Exemple



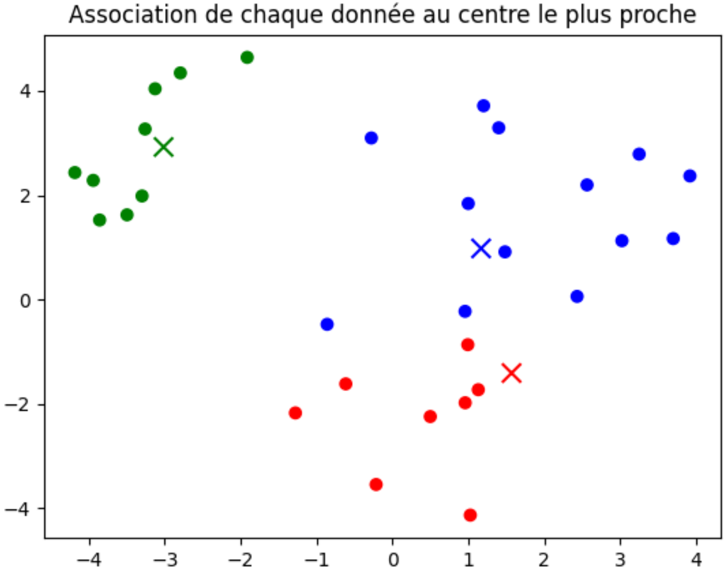
Algorithme des k-moyennes : Exemple



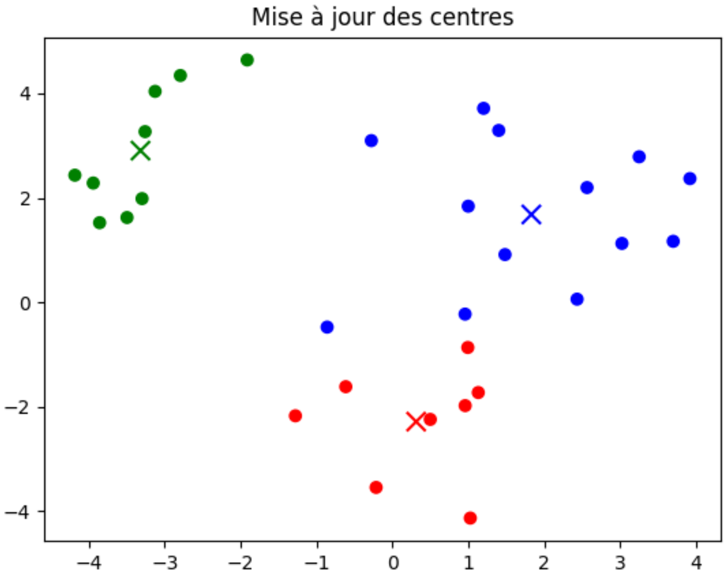
Algorithme des k-moyennes : Exemple



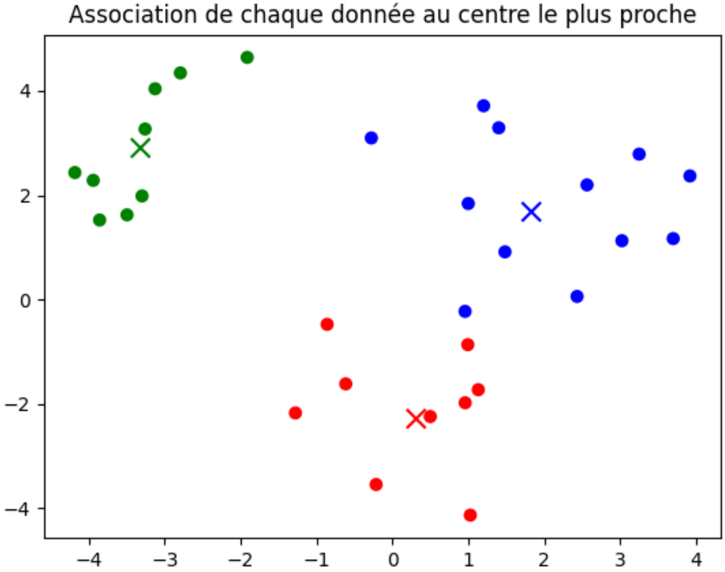
Algorithme des k-moyennes : Exemple



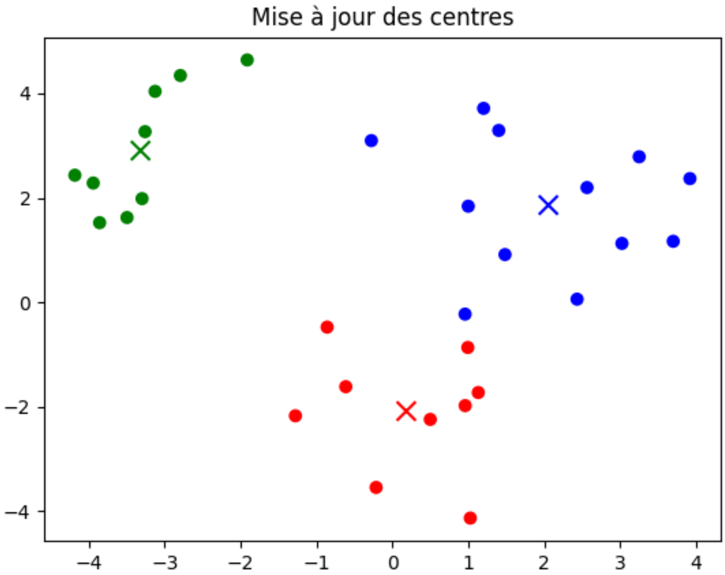
Algorithme des k-moyennes : Exemple



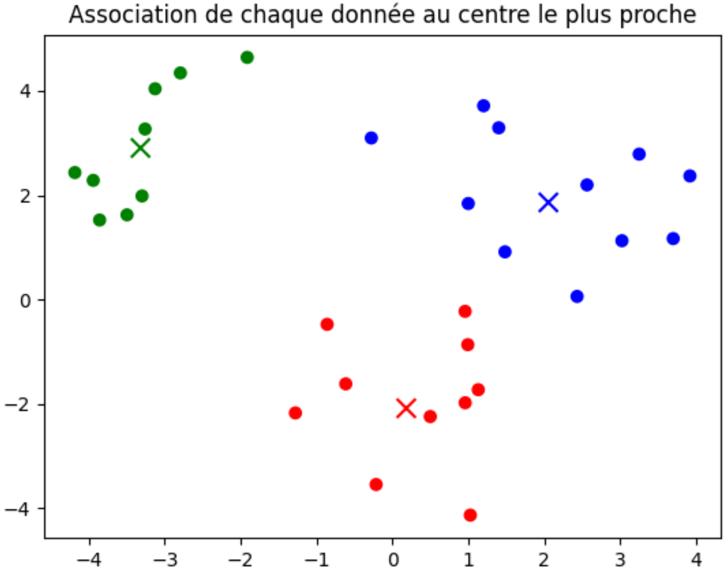
Algorithme des k-moyennes : Exemple



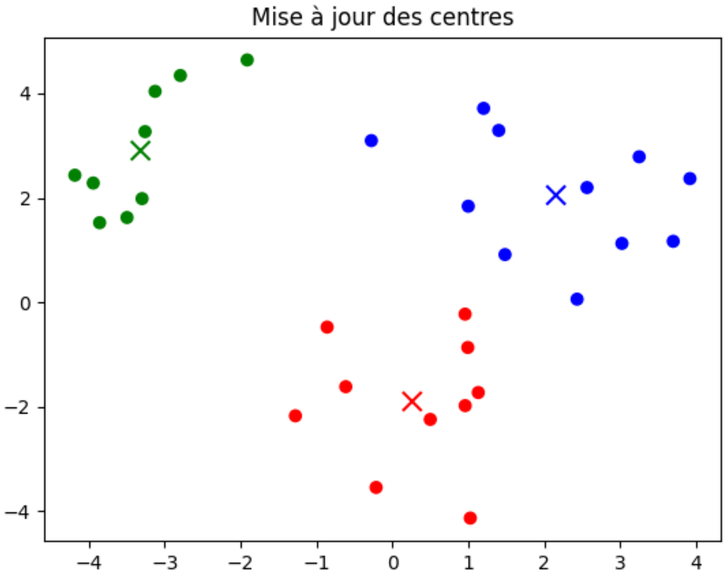
Algorithme des k-moyennes : Exemple



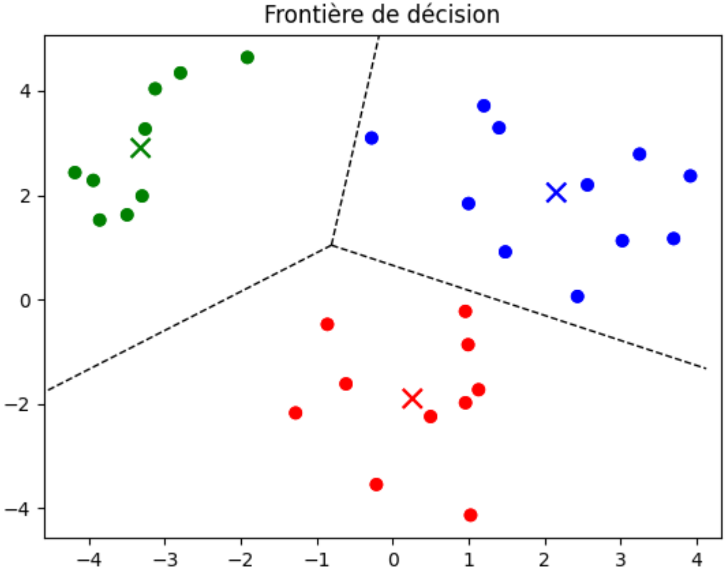
Algorithme des k-moyennes : Exemple



Algorithme des k-moyennes : Exemple



Algorithme des k-moyennes : Exemple



Algorithme des k-moyennes : Terminaison

Théorème (HP)

L'algorithme des k-moyennes termine (pas de boucle infinie).

<u>Preuve</u> :

Algorithme des k-moyennes : Terminaison

Théorème (HP)

L'algorithme des k-moyennes termine (pas de boucle infinie).

Preuve :

Il existe un nombre fini de partitions de X en k classes, donc l'inertie I ne peut prendre qu'un nombre fini de valeurs. De plus, I diminuant strictement à chaque itération (c'est un variant) :

- Réassigner x de X_i à X_j si $d(x, c_i) > d(x, c_j)$ fait diminuer I.
- Recalculer les centres des classes fait diminuer I car $f : y \longmapsto \sum_{x \in X} d(x, y)^2$ est minimum pour $y = \overline{X}$.

Algorithme des k-moyennes : Choisir k

Question

Comment choisir le nombre k de classes ?

Algorithme des k-moyennes : Choisir k

Question

Comment choisir le nombre k de classes ?

On peut calculer l'inertie obtenue pour différentes valeurs de k.

Algorithme des k-moyennes : Choisir k

Question

Comment choisir le nombre k de classes ?

On peut calculer l'inertie obtenue pour différentes valeurs de k. Cependant, plus k est grand, plus l'inertie diminue jusqu'à valoir 0 si k est égal au nombre de données (ce qui n'a aucun intérêt).

Algorithme des k-moyennes : Choisir k

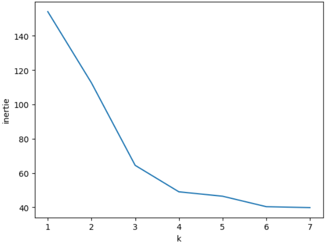
Question

Comment choisir le nombre k de classes ?

On peut calculer l'inertie obtenue pour différentes valeurs de k.
Cependant, plus k est grand, plus l'inertie diminue jusqu'à valoir 0 si k est égal au nombre de données (ce qui n'a aucun intérêt).
On choisit donc la plus grande valeur de k pour laquelle l'inertie diminue de façon significative.

Algorithme des k-moyennes : Choisir k

Méthode du coude (*elbow method*) : On choisit la plus grande valeur de k pour laquelle l'inertie diminue de façon significative.



On choisit $k = 3$ ou $k = 4$.

Algorithme des k-moyennes : Non optimalité

L'algorithme des k-moyennes converge toujours vers un minimum local, mais pas forcément vers un minimum global de l'inertie.

Question

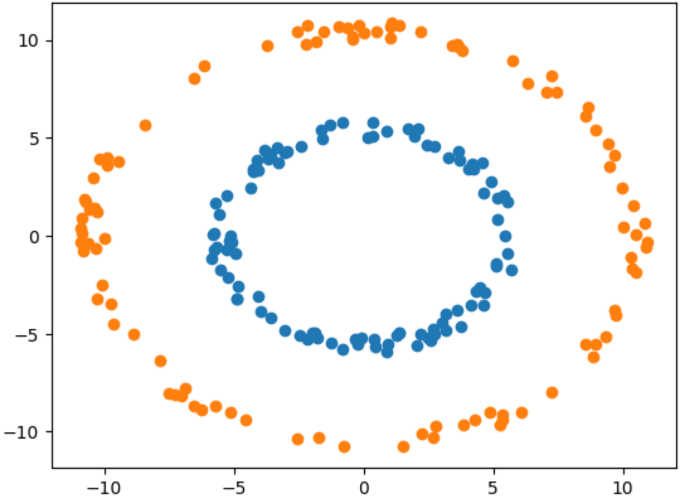
Donner un exemple d'exécution de l'algorithme des k-moyennes qui ne donne pas une partition d'inertie minimum.

Algorithme des k-moyennes : Classification de nouvelles données

On peut utiliser l'algorithme des k-moyennes pour classifier une nouvelle donnée x : on associe x à la classe dont le centre est le plus proche de x.

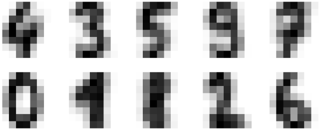
Algorithme des k-moyennes : Limites

L'algorithme des k-moyennes ne marche que sur des données linéairement séparables (pouvant être séparées par un hyperplan).



Algorithme des k-moyennes : Interprétations

Les centres obtenus à la fin de l'algorithme donnent des informations sur les constituants des classes.



Centres obtenus avec $k = 10$ sur des chiffres manuscrits

Algorithme des k-moyennes : Interprétations

Il est également intéressant de regarder l'équation des frontières de décision, pour savoir quels sont les attributs qui permettent de discriminer les données.

Si par exemple l'équation de la frontière de décision entre les classes 1 et 2 est $ax + by = 0$ avec $a \gg b$, alors l'attribut x est plus discriminant que y pour pouvoir distinguer les classes 1 et 2.

Application à la compression d'images

Une image est souvent représentée par une matrice dont chaque élément (pixel) est un triplet de valeurs entre 0 et 255 (rouge, vert, bleu).

Application à la compression d'images

Une image est souvent représentée par une matrice dont chaque élément (pixel) est un triplet de valeurs entre 0 et 255 (rouge, vert, bleu).

Question

Combien y a t-il de couleurs différentes possibles ?

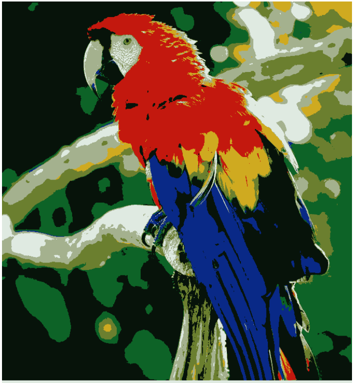
Application à la compression d'images

On peut souhaiter limiter le nombre de couleurs différentes :

- Sur un écran avec un nombre plus limité de couleurs (console...)
- Pour diminuer la taille de l'image : si on utilise k couleurs, on peut utiliser stocker un entier entre 1 et k pour chaque pixel au lieu de trois entiers entre 0 et 255.

Application à la compression d'images

On applique l'algorithme des k-moyennes sur les pixels pour obtenir k couleurs différentes (ici $k = 8$) et on remplace chaque pixel par la couleur la plus proche.



Classification hiérarchique ascendante (CHA)

Classification hiérarchique ascendante

Entrée : Des données X
Sortie : Une partition de X en classes

Mettre chaque $x \in X$ dans une classe différente
Tant que nécessaire **:**
  Fusionner les deux classes les plus proches
Renvoyer Les classes obtenues

Classification hiérarchique ascendante (CHA)

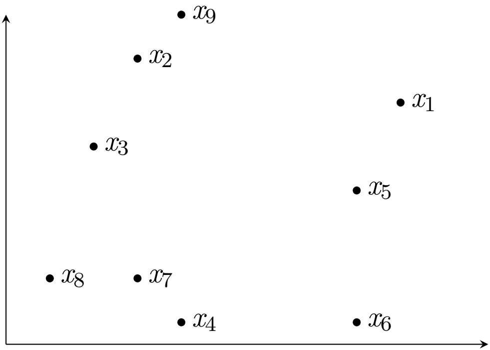
On peut choisir d'arrêter l'algorithme à un certain nombre de classes ou quand la distance minimum entre deux classes est supérieure à un certain seuil.

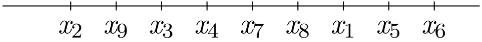
Classification hiérarchique ascendante (CHA)

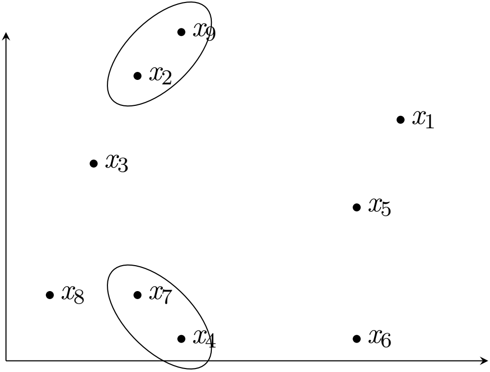
On peut choisir d'arrêter l'algorithme à un certain nombre de classes ou quand la distance minimum entre deux classes est supérieure à un certain seuil.

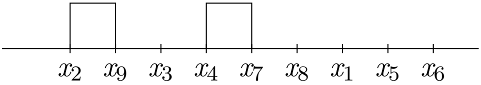
Exemples de distances entre classes A et B :

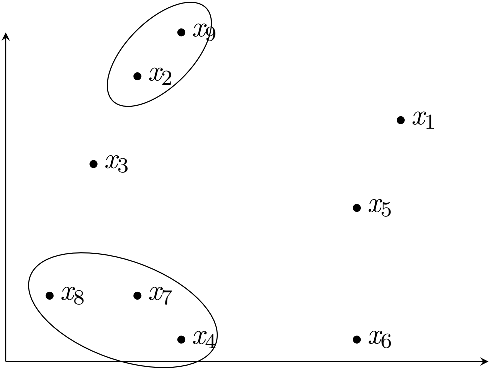
- Distance minimum : $\min_{a \in A, b \in B} d(a, b)$.
- Distance maximum : $\max_{a \in A, b \in B} d(a, b)$.
- Distance moyenne : $\frac{1}{|A||B|} \sum_{a \in A, b \in B} d(a, b)$.
- ...

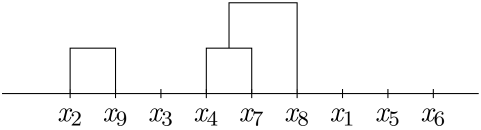

x_9
x_2
x_1
x_3
x_5
x_8
x_7
x_4
x_6

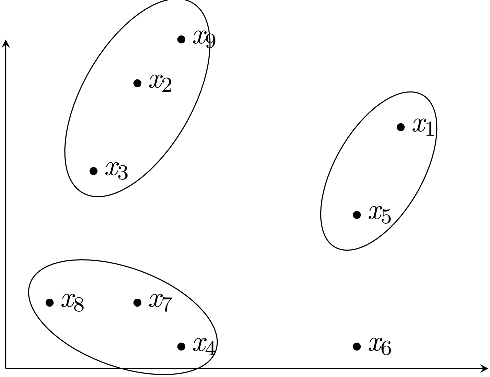

x_2 x_9 x_3 x_4 x_7 x_8 x_1 x_5 x_6

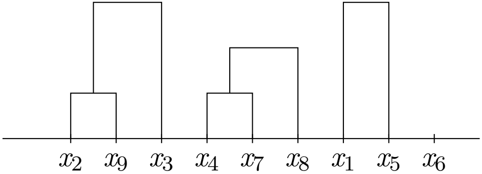

x_9
x_2
x_1
x_3
x_5
x_8
x_7
x_4
x_6

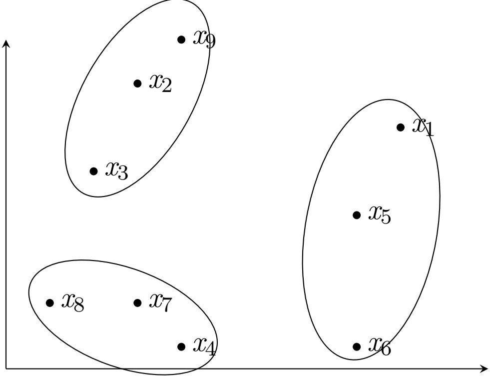

x_2 x_9 x_3 x_4 x_7 x_8 x_1 x_5 x_6

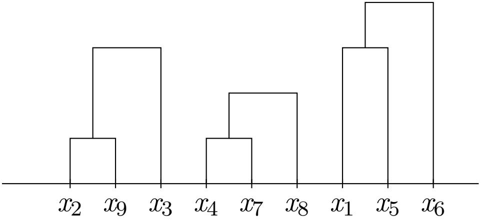

x_9
x_2
x_1
x_3
x_5
x_8
x_7
x_4
x_6

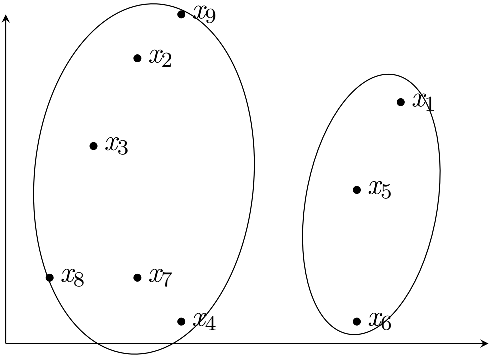

x_2 x_9 x_3 x_4 x_7 x_8 x_1 x_5 x_6


x_9
x_2
x_1
x_3
x_5
x_8
x_7
x_4
x_6

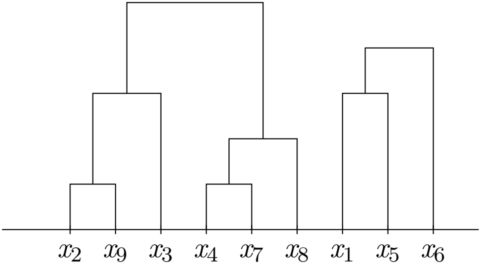

x_2 x_9 x_3 x_4 x_7 x_8 x_1 x_5 x_6


x9
x2
x3
x1
x5
x8
x7
x4
x6

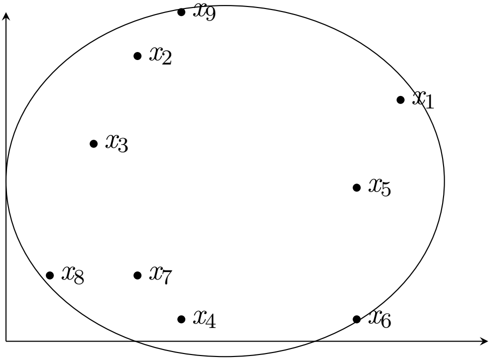

x2 x9 x3 x4 x7 x8 x1 x5 x6



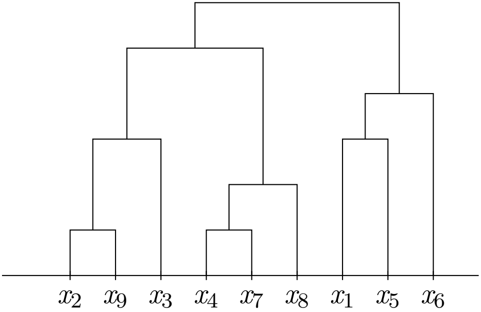
x_9
x_2
x_1
x_3
x_5
x_8
x_7
x_4
x_6



x_2 x_9 x_3 x_4 x_7 x_8 x_1 x_5 x_6



x_9
x_2
x_1
x_3
x_5
x_8
x_7
x_4
x_6



x_2 x_9 x_3 x_4 x_7 x_8 x_1 x_5 x_6