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 connue à l'avance..
  Exemples : k-moyennes, classification hiérarchique ascendante.

I Algorithme des k-moyennes

On note d une distance (par exemple la distance euclidienne).

Définition : Centre

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

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

Définition : Variance

La variance (ou : moment d'inertie) $V(X)$ d'un ensemble de vecteur X est définie par

$$V(X) = \sum_{x \in X} d(x, \overline{X})^2$$

Algorithme des k-moyennes (k-*means*)

Objectif : partitionner X en classes X_1, ..., X_k.

1. Soient $c_1, ..., c_k$ des vecteurs (centres) choisis aléatoirement.
2. Associer chaque donnée x à la classe X_i telle que $d(x, c_i)$ soit minimum.
3. Recalculer les centres des classes $c_i = \overline{X_i}$.
4. Si les centres ont changé, revenir à l'étape 2.

Attention : dans l'algorithme des k-moyennes, k est le nombre de classes alors que dans l'algorithme des k plus proches voisins, k est le nombre de voisins.

I.1 Terminaison (HP)

Théorème

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

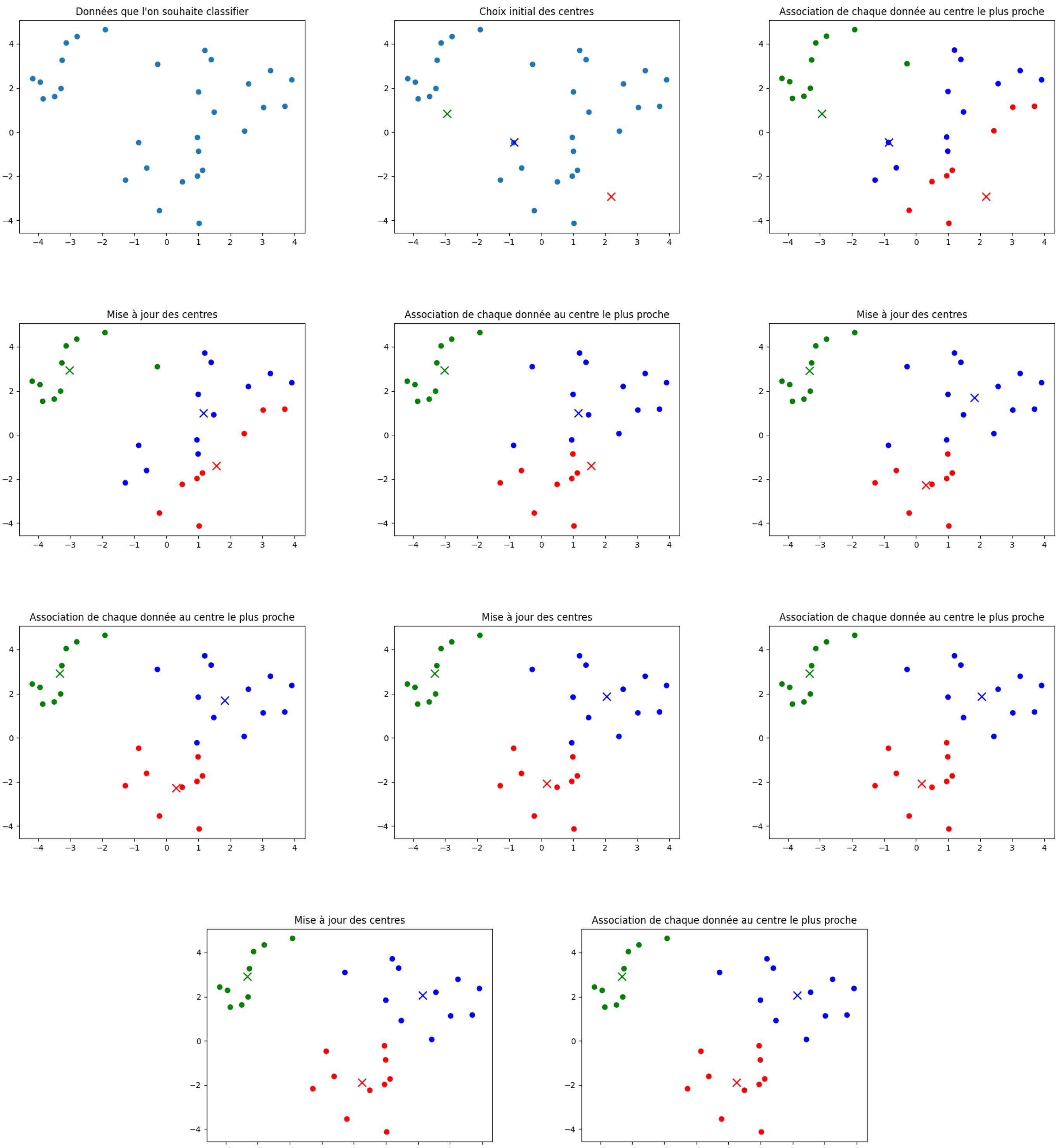
Preuve :

On montre que I est un variant de boucle : I décroît strictement à chaque itération et ne peut prendre qu'un nombre fini de valeurs, donc le nombre d'itérations est fini. 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. Il suffit donc de montrer que I décroît strictement à chaque itération :

- 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, d'après le résultat suivant :

Théorème

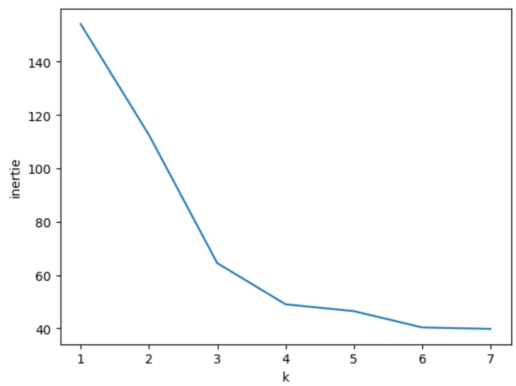
Si X est un ensemble de vecteurs alors $f : y \mapsto \sum_{x \in X} d(x, y)^2$ est minimum pour $y = \overline{X}$.



Exemple d'exécution de l'algorithme des k-moyennes

I.2 Choisir k

On peut calculer l'inertie obtenue pour différentes valeurs de k. La méthode du coude consiste à choisir la plus grande valeur de k pour laquelle l'inertie diminue de façon significative.



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

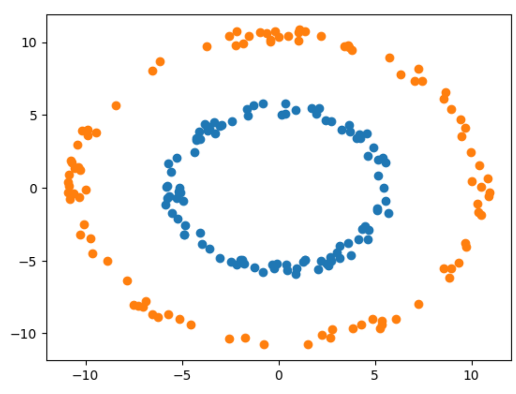
I.3 Non optimalité

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

Exercice 1.
Donner un exemple d'exécution de l'algorithme des k-moyennes qui ne donne pas une partition d'inertie minimum.

I.4 Limites

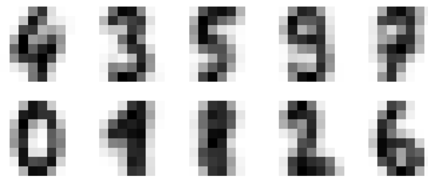
L'algorithme des k-moyennes ne marche que sur des données linéairement séparables (pouvant être séparées par un hyperplan).



L'algorithme des k-moyennes ne permettrait pas de classifier correctement ces données.

I.5 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

II 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

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 :

1. Distance minimum : $\min\limits_{a \in A, b \in B} d(a, b)$.
2. Distance maximum : $\max\limits_{a \in A, b \in B} d(a, b)$.
3. Distance moyenne : $\dfrac{1}{|A||B|} \sum\limits_{a \in A, b \in B} d(a, b)$.

Exercice 2.
Appliquer l'algorithme de classification hiérarchique ascendante sur les données suivantes en dessinant le dendrogramme obtenu. On utilisera la distance 1.