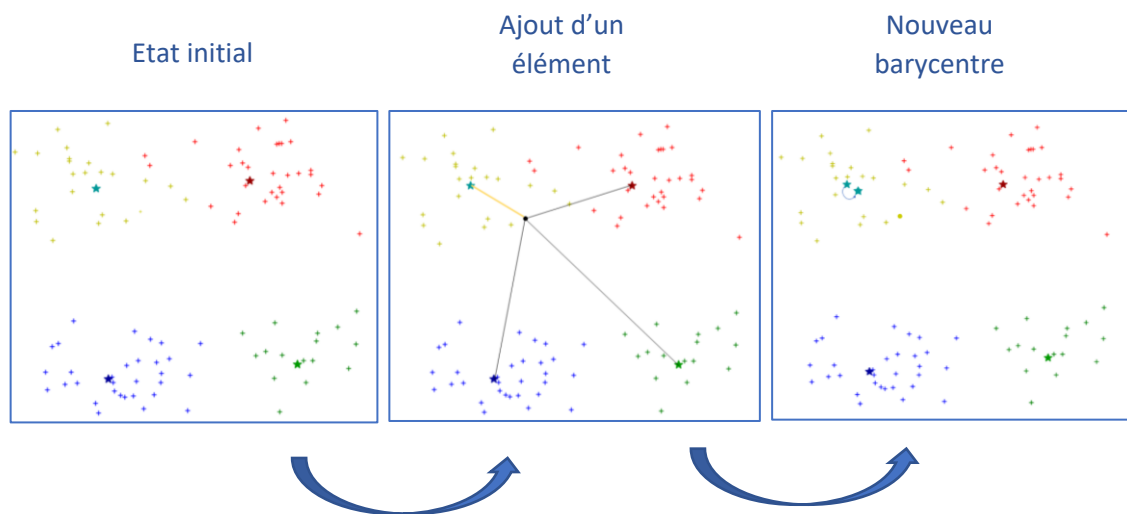


K-means (k-moyennes)

C'est un algorithme non supervisé de clustering. Il permet de regrouper en K clusters distincts les observations du data set.

Son principe est de rajouter chaque élément en entrée au cluster qui lui correspond le mieux parmi des clusters prédéfinis.

On appelle cet algorithme K-means puisqu'il permet de minimiser la somme des distances entre chaque individu et le barycentre du cluster auquel il appartient.

**Algorithme :**

Entrée : un ensemble E de N données à classifier, K le nombre de cluster à former, k clusters prédéfinis (par l'utilisateur)

Sorties : les données réparties en K clusters ;

REPETER

1. Affecter chaque point au barycentre le plus proche

2. Recalculer le barycentre du cluster

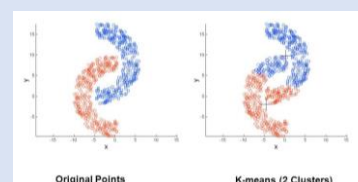
JUSQU'À CONVERGENCE

Avantages il est possible de :

- le comprendre facilement
- l'implémenter facilement
- l'appliquer à des données de grandes tailles, et aussi à tout type de données (Mêmes textuelles), en choisissant une bonne notion de distance

Inconvénients :

- Valeur de K à trouver
- Modèle non adapté à certains types de clusters (non convexes)

**Paramètres importants :**

Le choix du nombre de clusters k, les clusters de départ et de la distance a des impacts parfois significatifs sur les résultats obtenus.