

Lab Clustering n°3

Lets build a Kmeans clustering program

Prérequis

1. Ecrire une fonction **D(x,y)** qui retourne la distance entre deux points x et y (peut importe leur dimension)
2. Ecrire une fonction **G(liste_pts)** où *liste_pts* est une liste de points, qui retourne le barycentre de ces pts.
3. Ecrire une fonction **Rand(liste_pts,k)** qui retourne K éléments pris au hasard parmi la liste des points *liste_pts*.

Algorithme

1. Implémentez l'algorithme suivant :

Var X : liste de **n** points de dimension **p**

Var k : nombre de classes

Prendre **k points** aléatoires comme **centres** des **k classes**.

Répéter

 Pour **point** dans **X** faire

 Affecter **point** à la **classe** dont le **centre** est le plus proche.

 FinPour

 Recalculer le **centre** de chaque **classe**

FinRepeter **Condition**

La condition d'arrêt peut varier d'algorithme en algorithme. Dans notre cas, nous prendrons la condition d'arrêt :

Aucun point n'a changé de classe durant la dernière itération.

Application

1. Appliquer le programme obtenu au dataset du Lab n°1. Comparez avec le résultat donné par l'algorithme implémenté par **sikitlearn**.