## Première partie : Afficher les données :

### Code réaliser :

from scipy.io import arff

import matplotlib.pyplot as plt

data=arff.loadarff(open('./artificial/2d-10c.arff','r'))

listX = []

listY=[]

for point in data[0]:

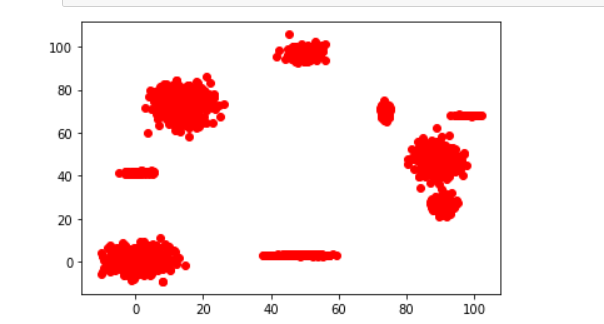
listX.append(point[0])

listY.append(point[1])

plt.plot(listX,listY, 'ro')

plt.show()

### Affichage :



from sklearn.cluster import KMeans

from sklearn.datasets import make\_blobs

import numpy as np

X = []

n\_samples=1500

for point in data[0]:

X.append([point[0],point[1]])

X=np.array(X)

kmeans = KMeans(n\_clusters=3, random\_state=0).fit\_predict(X)

plt.subplot(221)

plt.scatter(X[:, 0], X[:, 1], c=kmeans)

### Metrique

from sklearn import metrics

find = False

nb\_cluster = 2

metric\_best = 0

limit=20

iteration = 0

while(not find & (iteration < limit)):

kmeans = KMeans(n\_clusters=nb\_cluster, random\_state=0).fit\_predict(X)

find\_metric = metrics.silhouette\_score(X, kmeans, metric='euclidean')

if find\_metric > metric\_best:

metric\_best=find\_metric

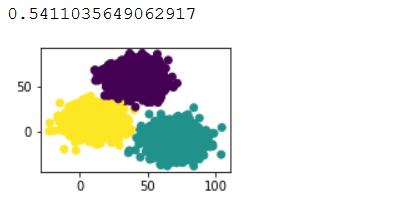
nb\_cluster+=1

else:

find=True

iteration+=1

print(nb\_cluster)



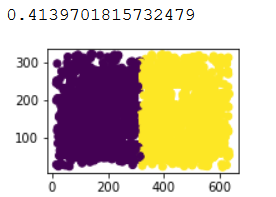
Xclara.arff

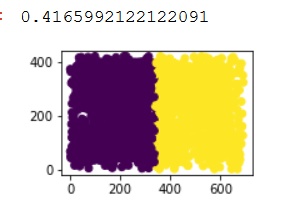
## Limite du k-means

Une image contenant texte, tissu

Description générée automatiquement

K-means -> basé sur la distance euclidienne. Sur les jeux de données au-dessus, les points sont trop proches les uns des autres





# Clustering agglomératif

Une image contenant texte

Description générée automatiquementUne image contenant texte

Description générée automatiquementUne image contenant texte

Description générée automatiquement

Une image contenant texte

Description générée automatiquement

## Métrique

Une image contenant texte

Description générée automatiquement

Une image contenant texte

Description générée automatiquement

Une image contenant texte

Description générée automatiquement

Une image contenant texte

Description générée automatiquement

 -> time

<https://openclassrooms.com/fr/courses/4379436-explorez-vos-donnees-avec-des-algorithmes-non-supervises/4379561-partitionnez-vos-donnees-avec-un-algorithme-de-clustering-hierarchique>

## Limite clustering agglomératif

Une image contenant texte

Description générée automatiquement

La distance entre cluster sur des courbes ne fonctionnent pas

# DBSCAN

Esp 🡪 distance entre les clusters et min\_samples 🡪 nombre min de points par cluster

## Limite :

Densité de variable dans les donnée

Disk-6000n

# Partie 2

## Dataset a :

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | a |  |  |  |  |  | Synthèse |
| k-means |  | |  |  |  |  |  |  |
| Agglomerative cluster |  | |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |