MEGA Adrian 5SDBD

RABARY Anaïs

TP Apprentissage Non Supervisé

# Introduction

Ce rapport a pour but de rendre compte de nos travaux en TP d’apprentissage non supervisé. Les deux chapitres traiteront des deux parties du TP, respectivement sur DBSCAN et SNN. Le code peut être trouvé sur le dépôt suivant

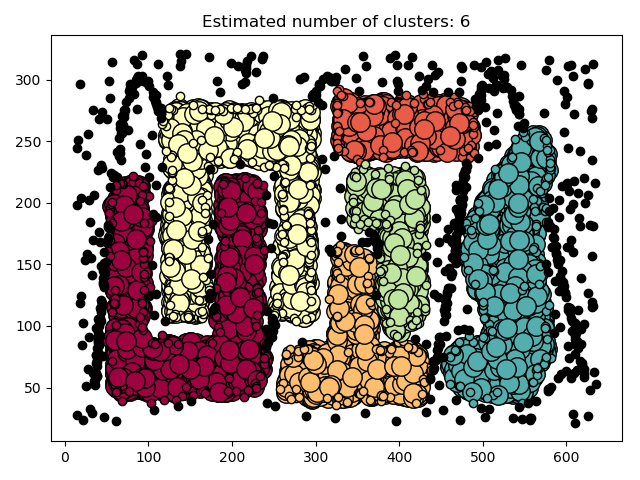
<https://github.com/dx07/5SDBD-Apprentissage/tree/master/Apprentissage-Non-Supervise>

Nous parlerons dans ce rapport des résultats de nos algorithmes de clustering sur les 4 jeux de données proposés. Il sera question d’appliquer les algorithmes développés et d’étudier les paramétrages, la qualité du clustering, etc…

# I – DBSCAN de scikit-learn

DBSCAN présente l’intérêt de ne pas obliger de fixer le nombre de clusters à l’avance, tout en étant robuste au bruit et aux anomalies. De plus, il va nous permettre d’identifier des clusters concaves et d’apprécier leur qualité. En faisant varier les paramètres de la fonction DBSCAN de scikit-learn, nous allons tenter d’affiner la qualité de nos clusters.

## Dataset 1



Epsilon : 10

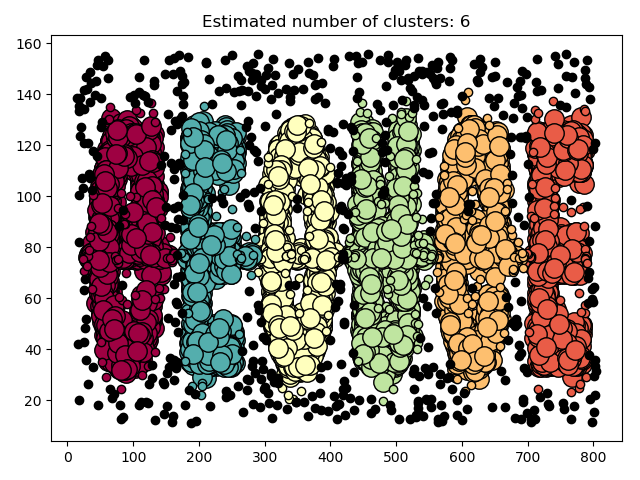
Min Samples : 18

Estimated number of clusters : 6

Estimated number of noise points : 630

Silhouette Coefficient : 0.246

## Dataset 2



Epsilon : 9.8

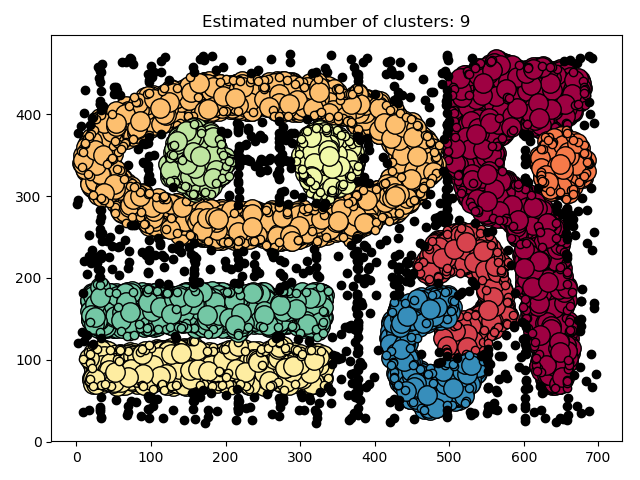
Min Samples : 18

Estimated number of clusters: 6

Estimated number of noise points: 566

Silhouette Coefficient: 0.490

## Dataset 3



Epsilon : 10

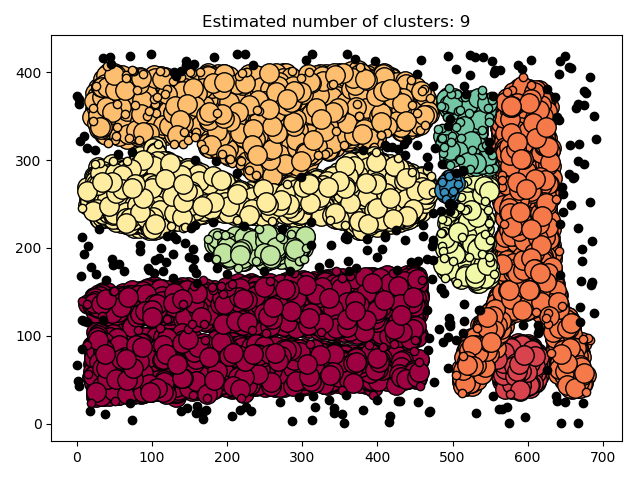
Min Samples : 13

Estimated number of clusters: 9

Estimated number of noise points: 774

Silhouette Coefficient: -0.068

## Dataset 4



Epsilon : 10.1

Min Samples : 7

Estimated number of clusters: 9

Estimated number of noise points: 277

Silhouette Coefficient: -0.078

Notes Evaluation :

* Formes non-convexes, passer d’un point à un autre en restant dans le cluster
* Densité (nb de voisins proches)
* Forme des clusters
* Stabilité des clusters
* Cohérence