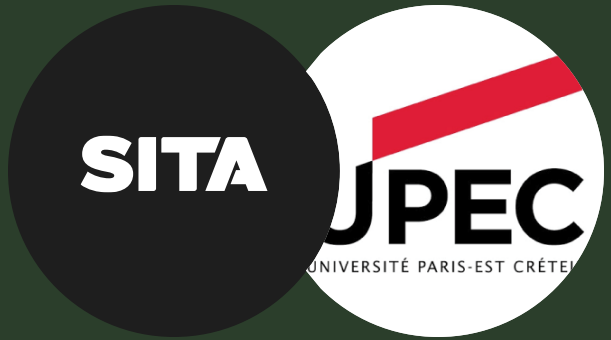


EPISEN – ING3. SI

Machine Learning

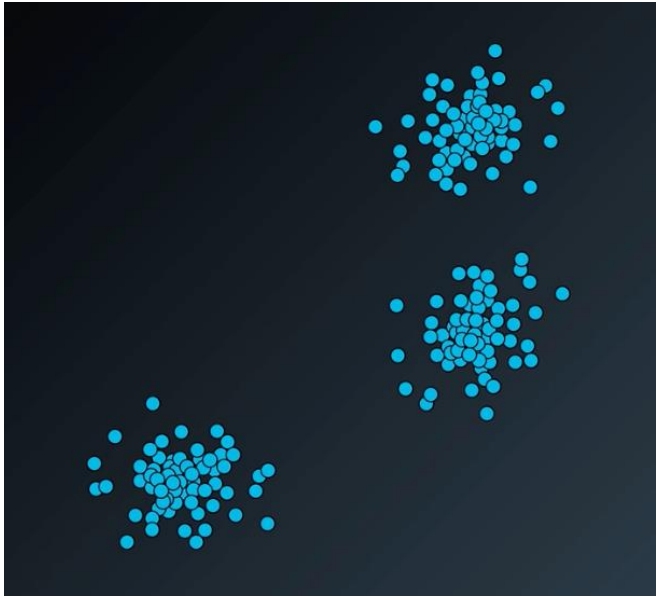
Abdallah EL HIDALI
Tech Lead Sita For Aircraft
abdallah.el-hidali@sita.aero

EPISEN
2024/2025

SITA

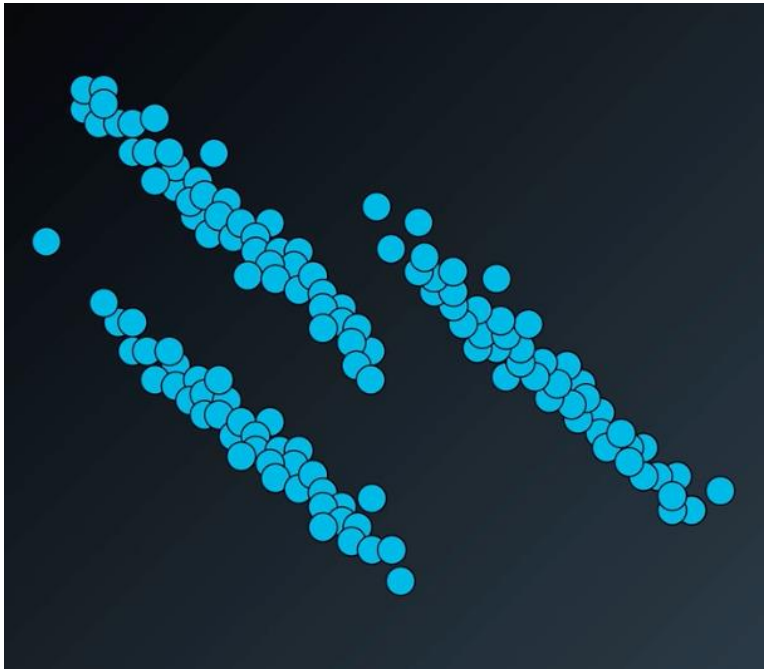
XI. Clustering hiérarchique et clustering basé sur la densité

Les limites du K-means



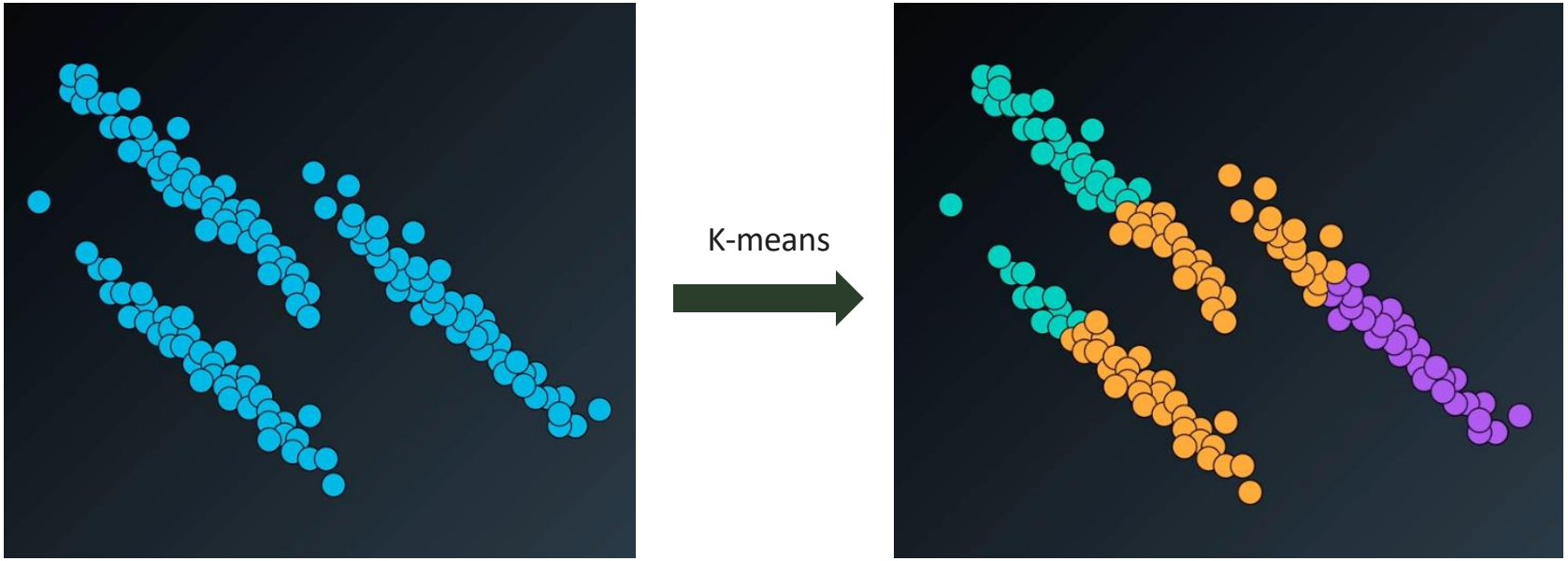
K-means peut facilement trouver les 3 clusters dans cet exemple

Les limites du K-means



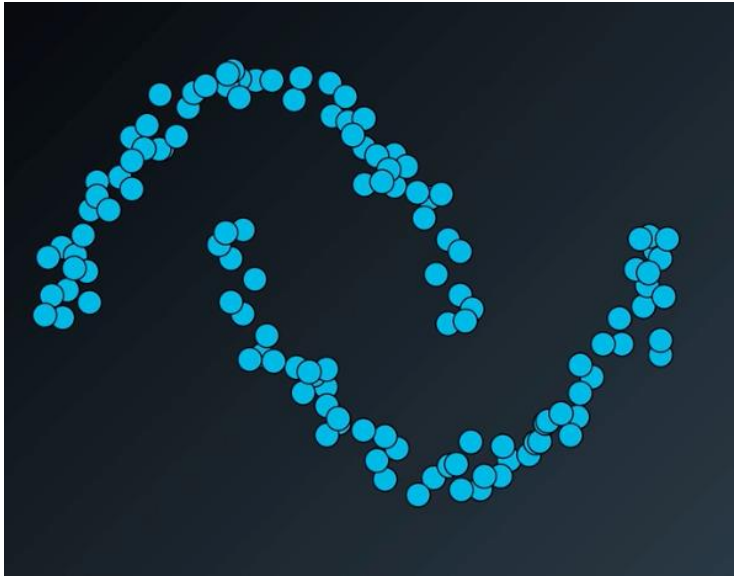
Mais qu'en est-il de cet exemple ? Est-ce que K-means pourrait retrouver les 3 clusters ?

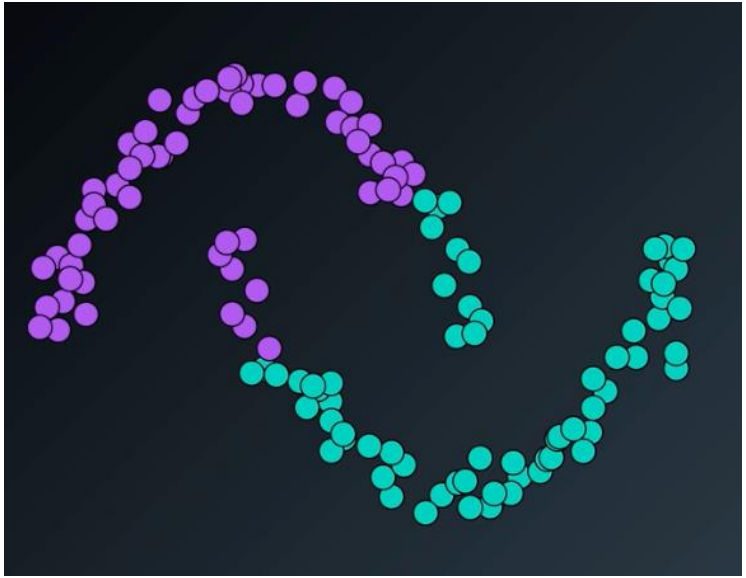
Les limites du K-means



K-means n'arrive pas à identifier les 3 clusters de cet exemple car il se base sur la distance comme moyen de les séparer.

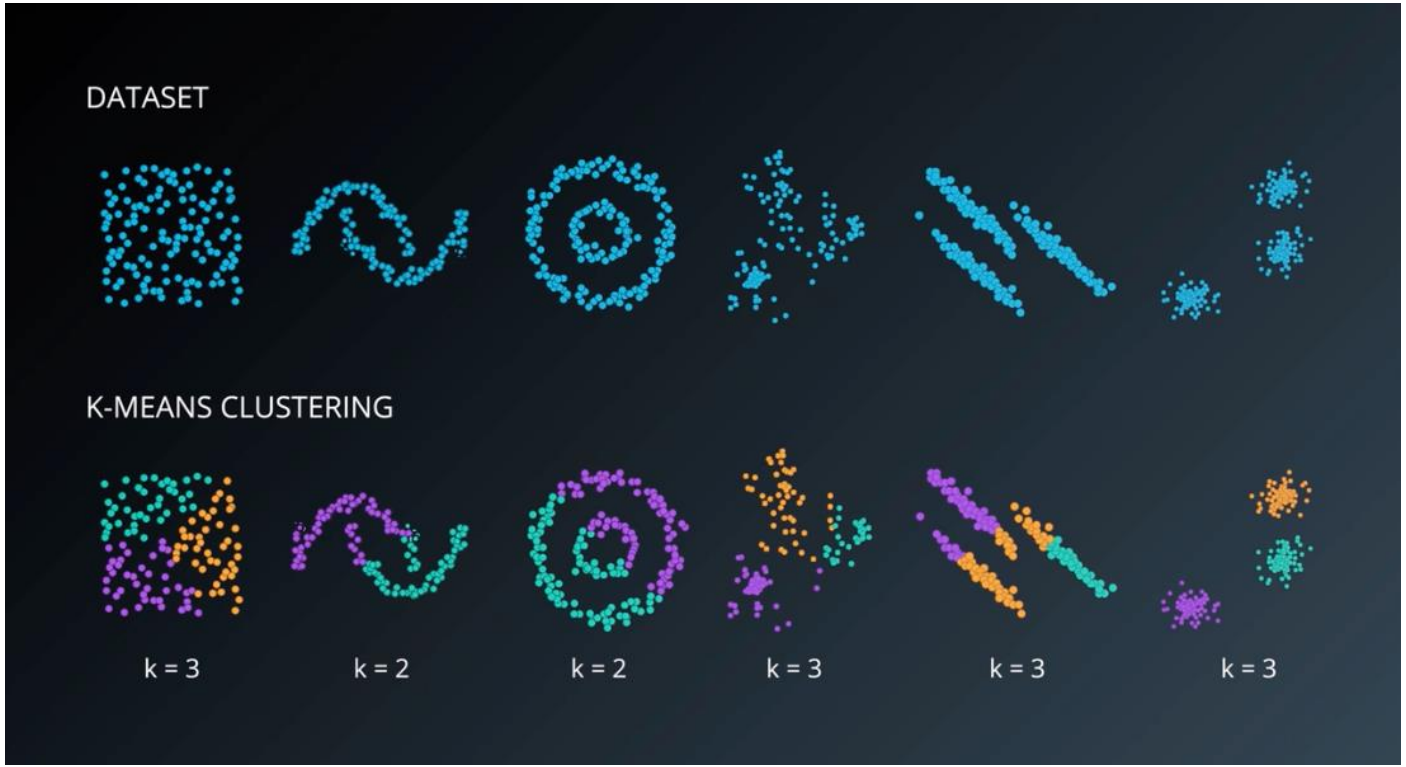
Les limites du K-means





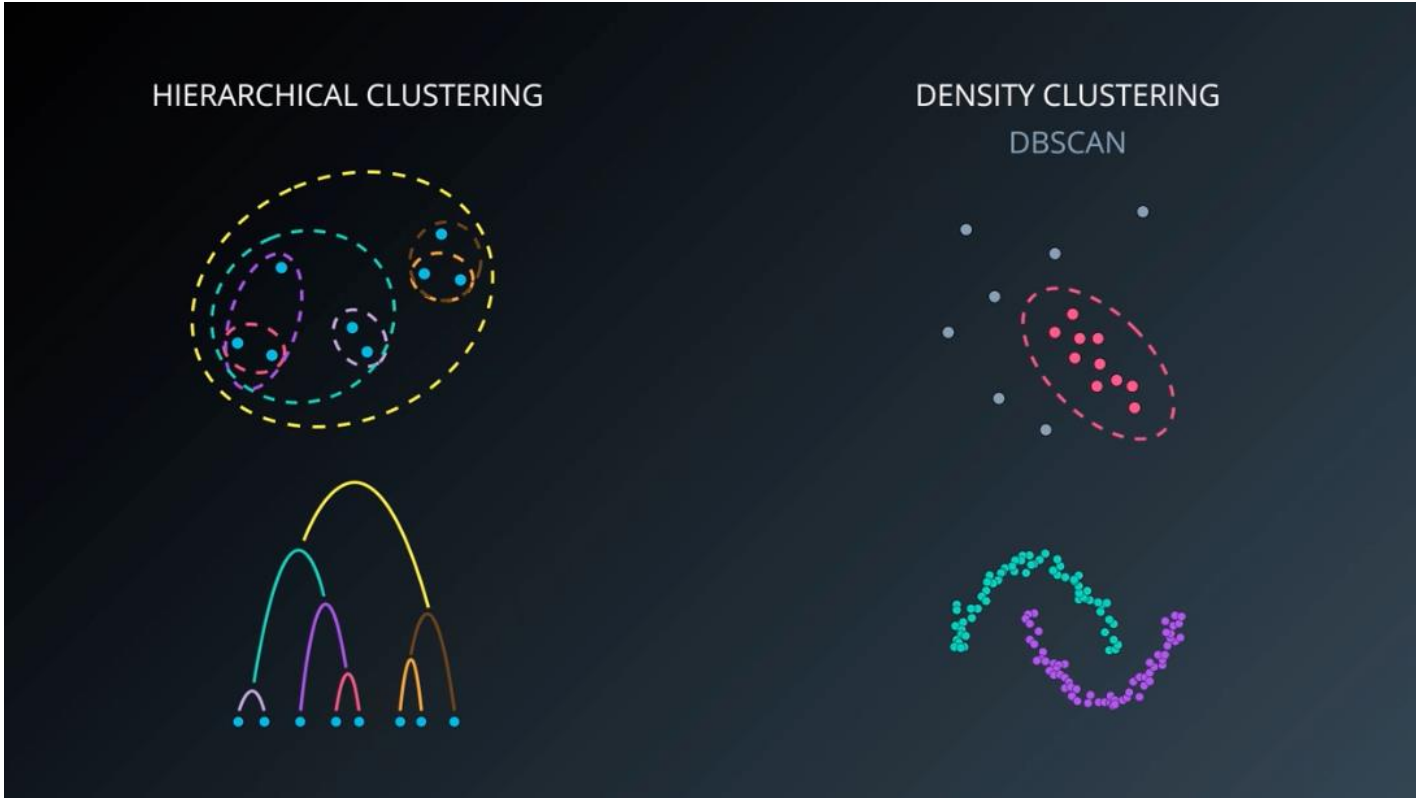
Pour cet exemple également, K-means n'arrive pas à bien séparer les deux clusters.

Les limites du K-means



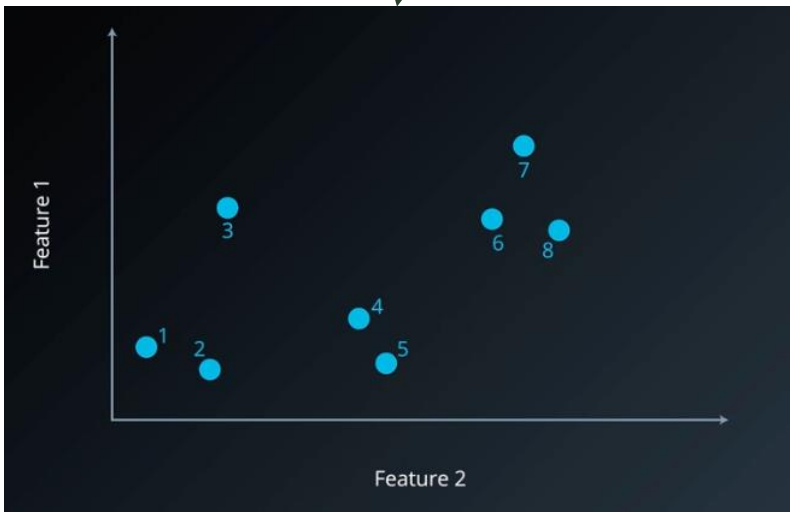
Nous allons essayer les algorithmes de clustering sur différents jeux de données pour comparer leurs performances.

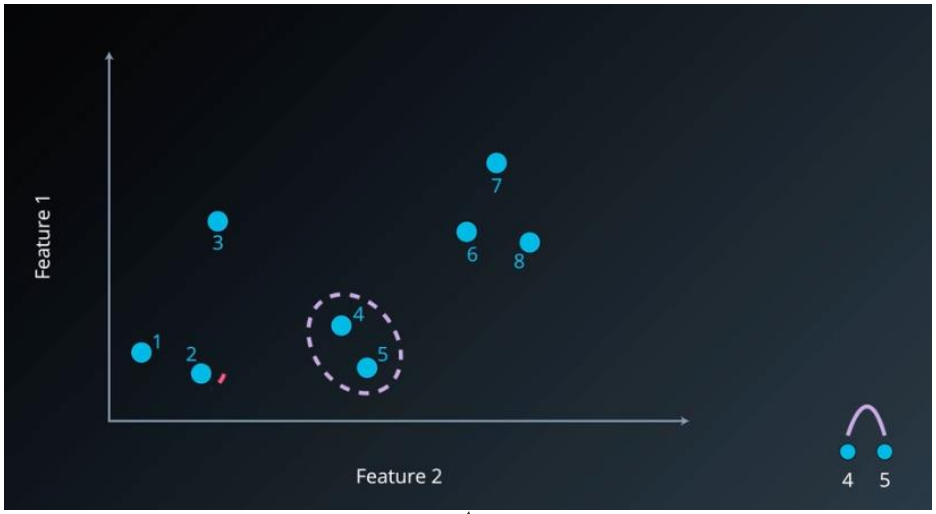
Aperçu des autres méthodes de clustering



Clustering hiérarchique

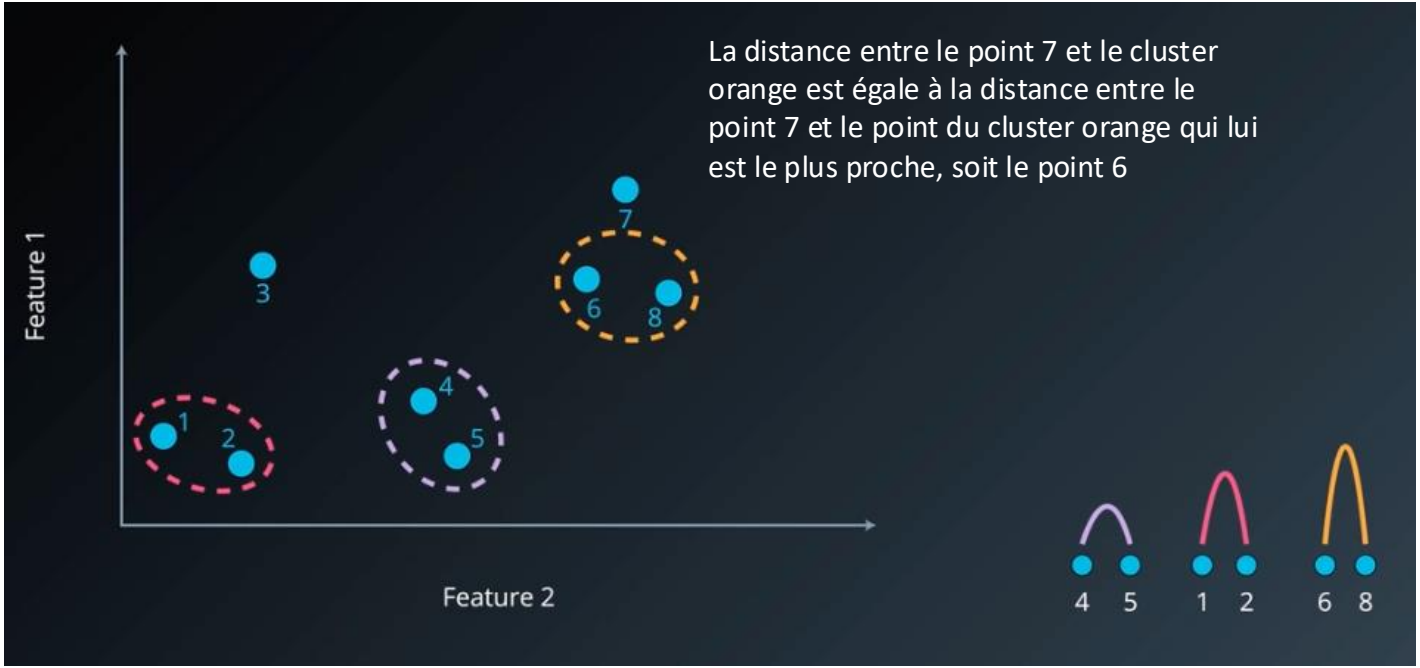
On considère chaque point comme un cluster



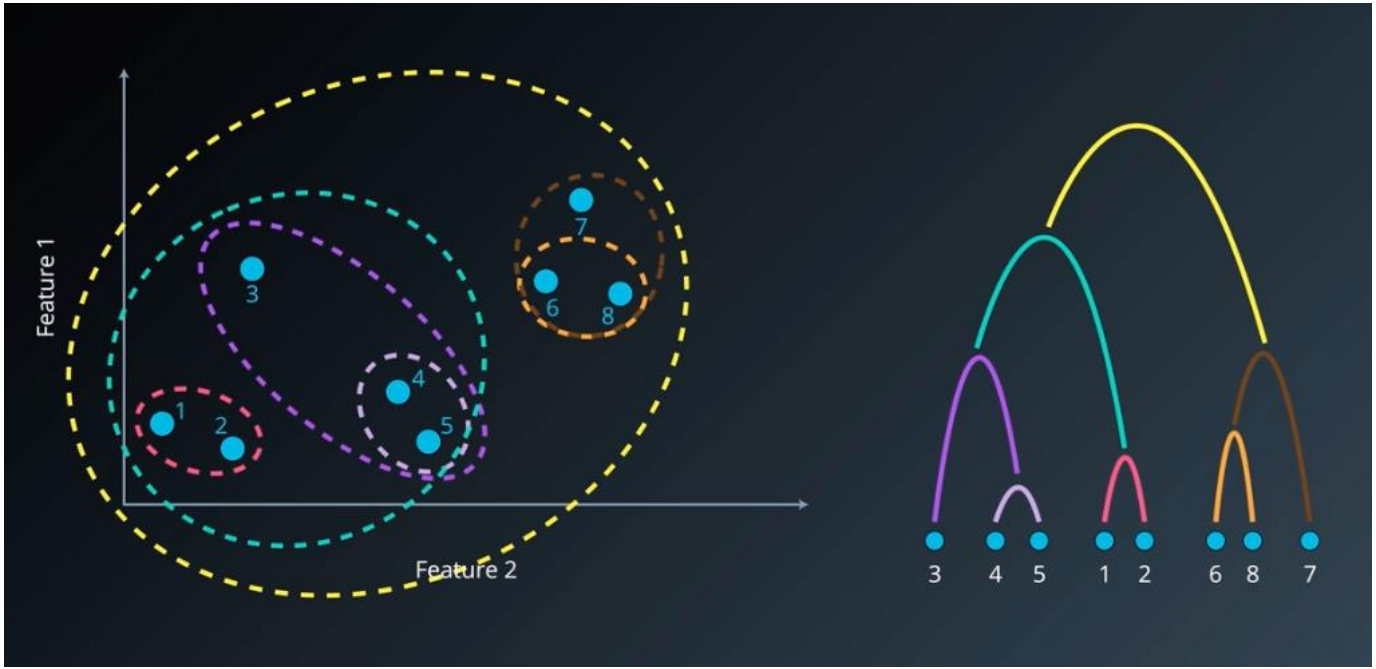


On calcule la distance entre chaque point et le reste des points puis on regroupe les 2 points avec la distance la plus faible dans le même cluster

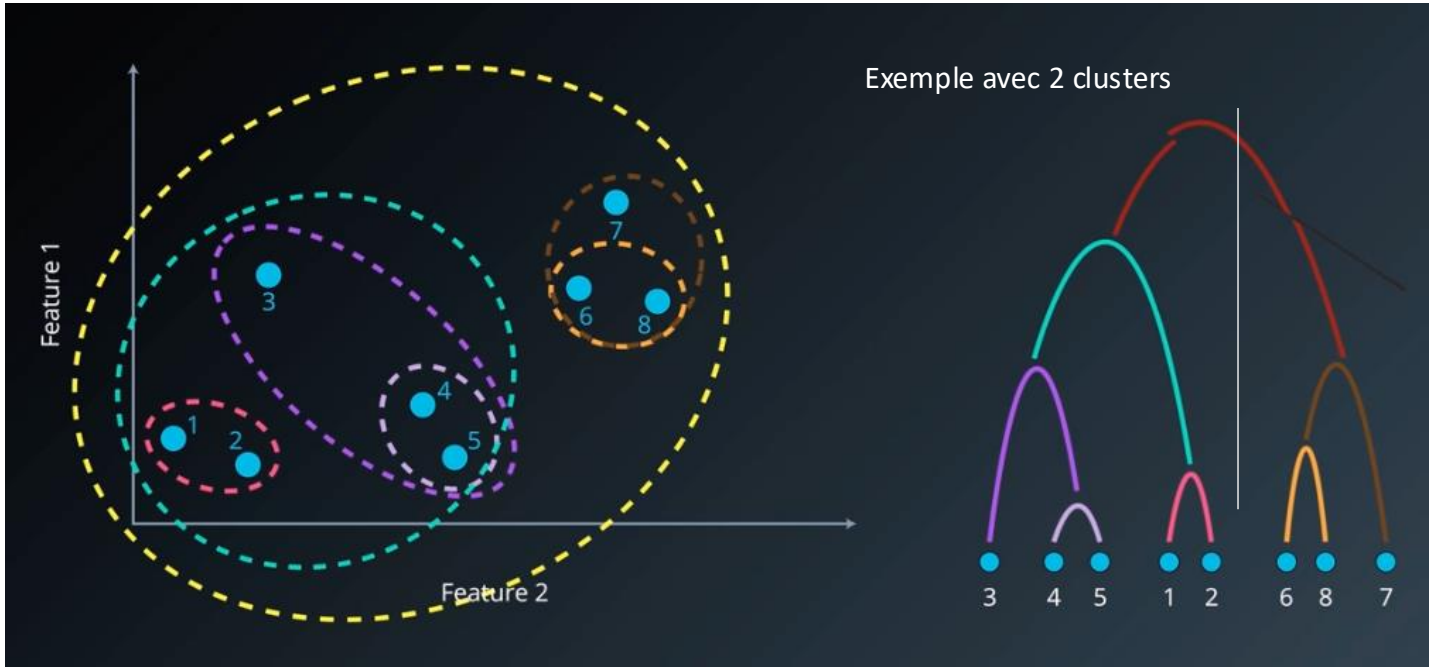
Clustering hiérarchique



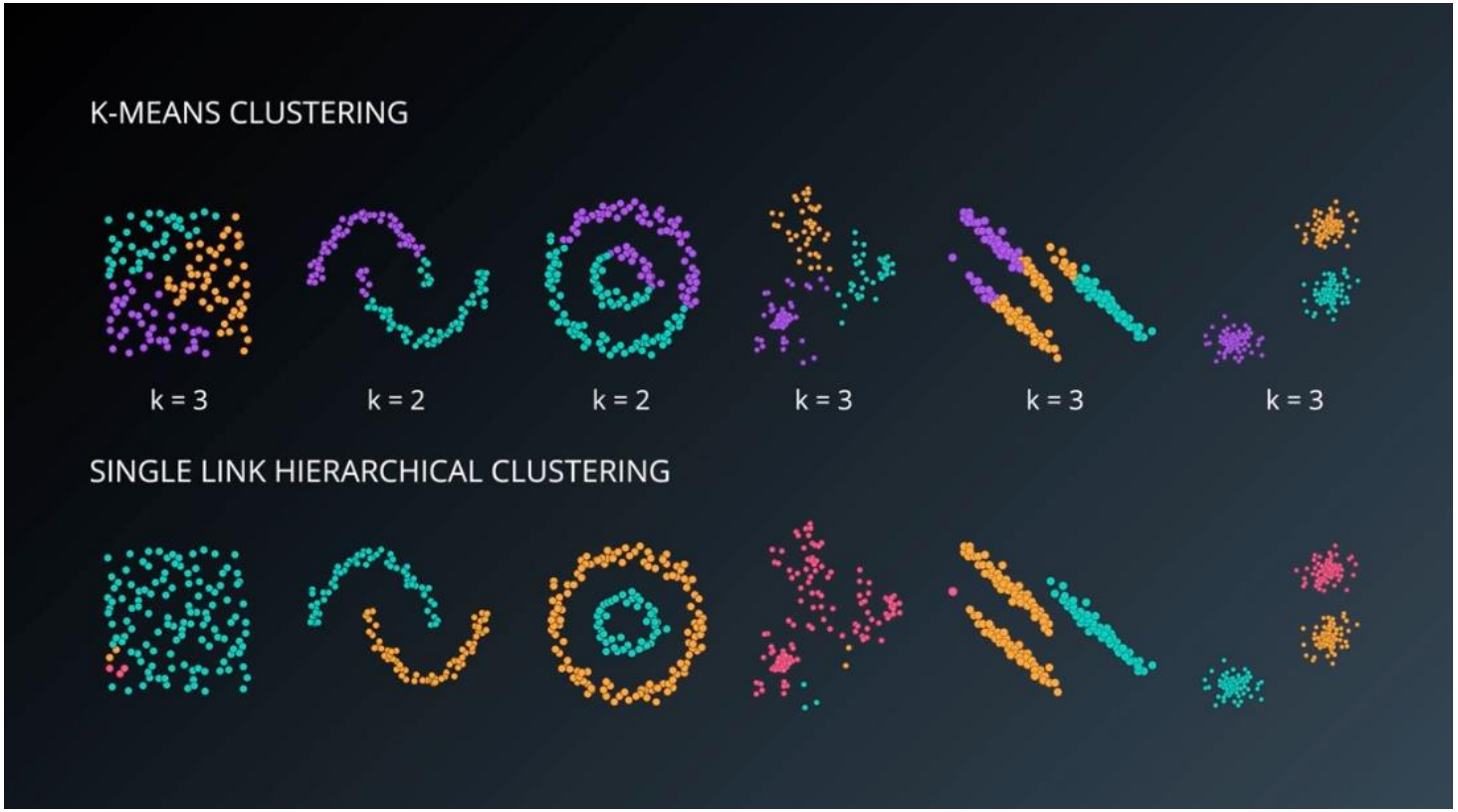
Clustering hiérarchique



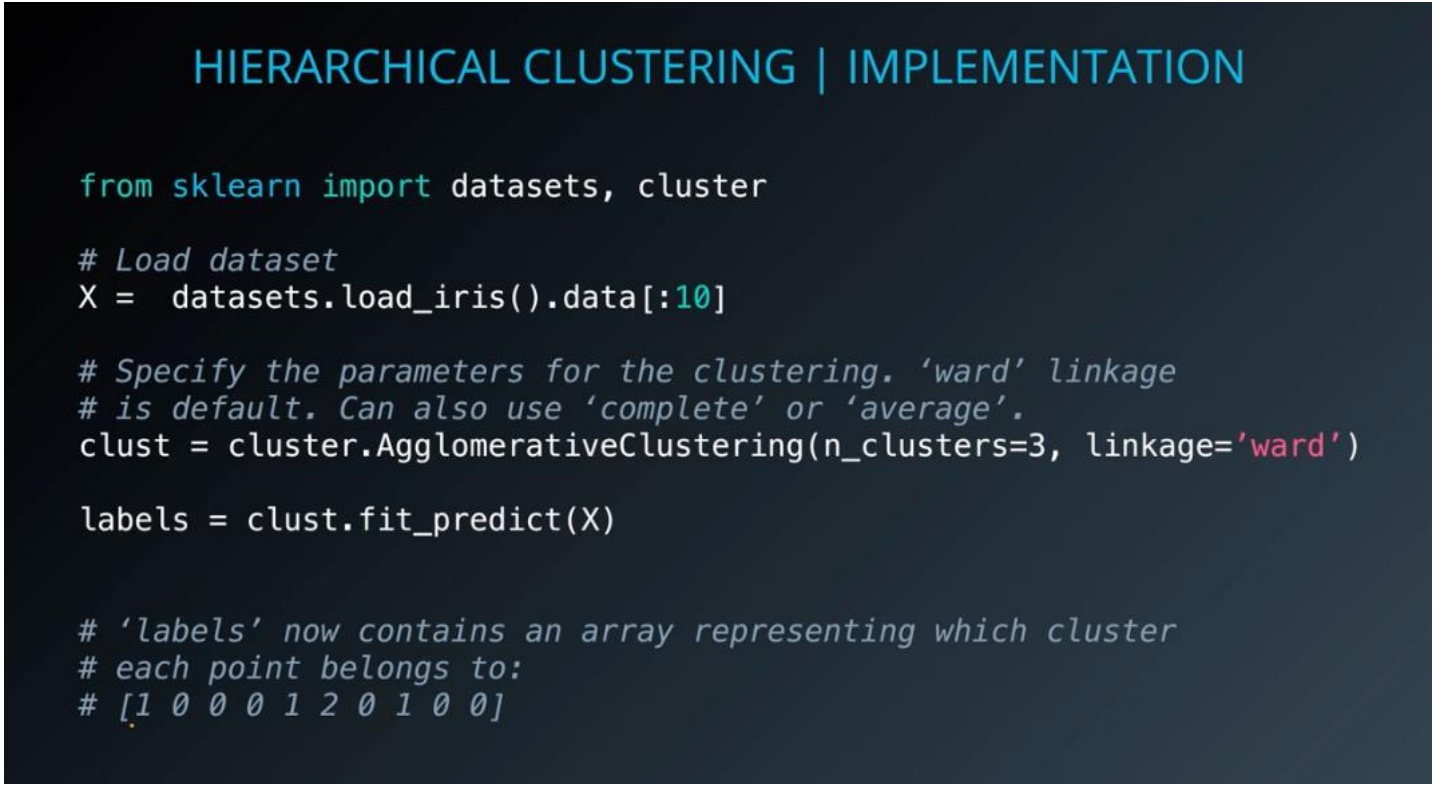
Clustering hiérarchique



Comparaison des méthodes de clustering



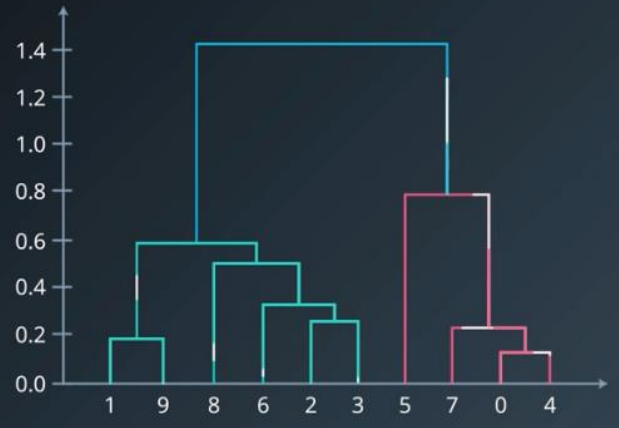
Clustering hiérarchique sur sklearn



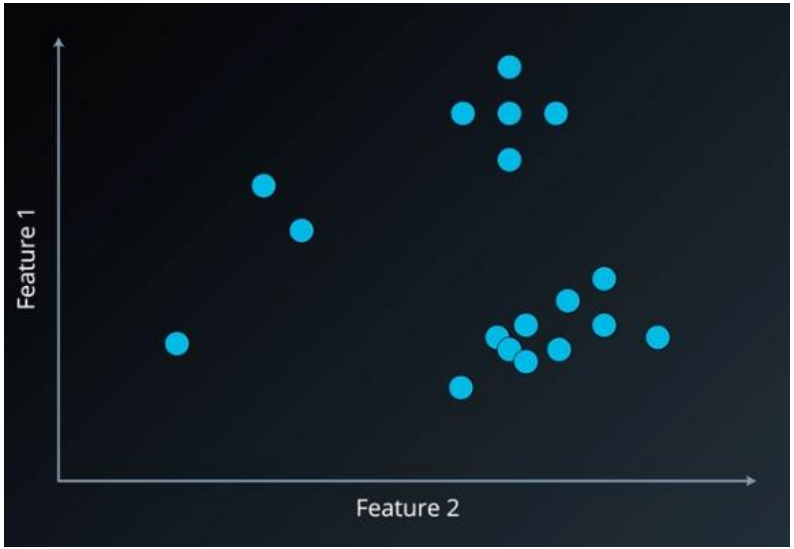
Clustering hiérarchique

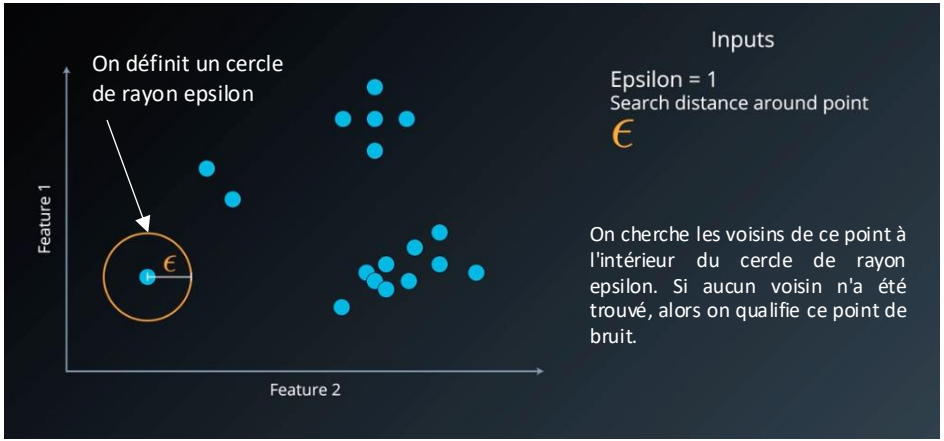
HIERARCHICAL CLUSTERING | IMPLEMENTATION

```
from scipy.cluster.hierarchy import dendrogram, ward, single
from sklearn import datasets
import matplotlib.pyplot as plt

# Load dataset
X =  datasets.load_iris().data[:10]

# Perform clustering
linkage_matrix = ward(X)

# Plot dendogram
dendogram(linkage_matrix)

plt.show()
```



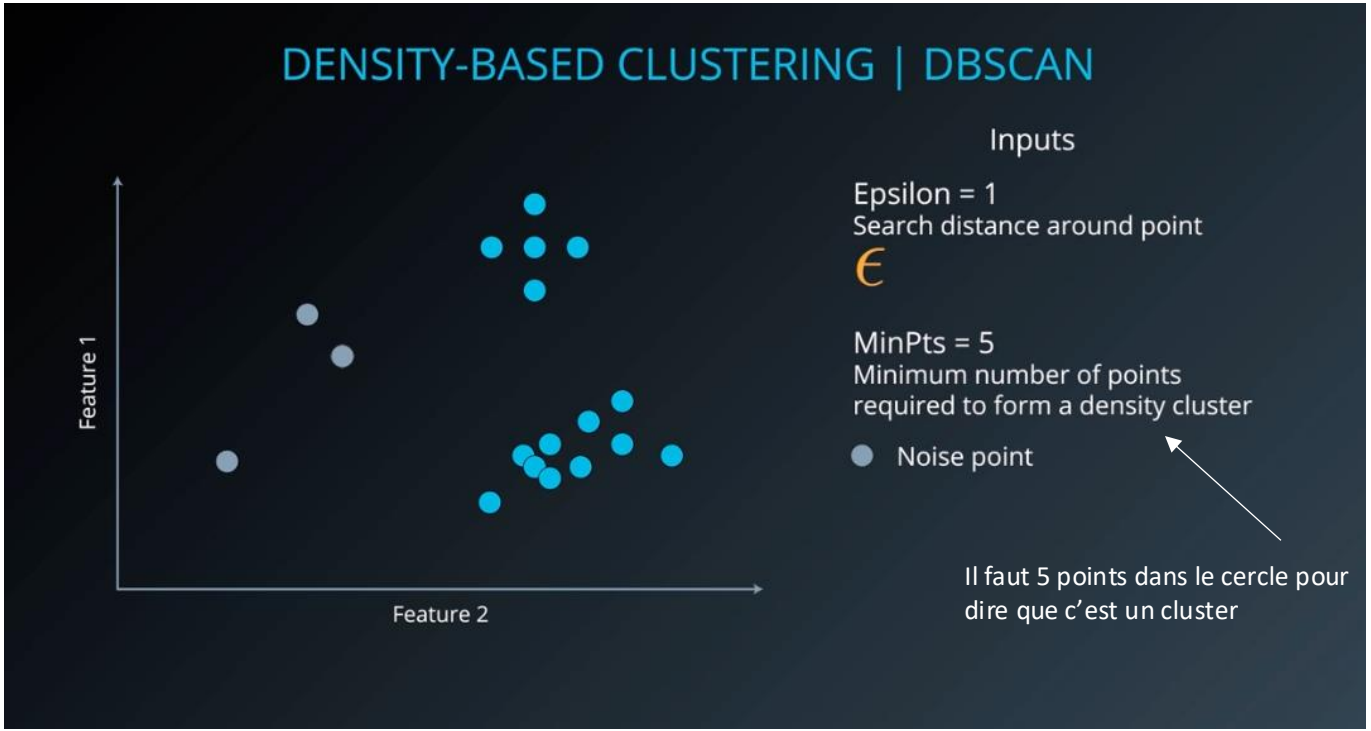
DBSCAN: Density-Based Spatial Clustering of Applications with Noise



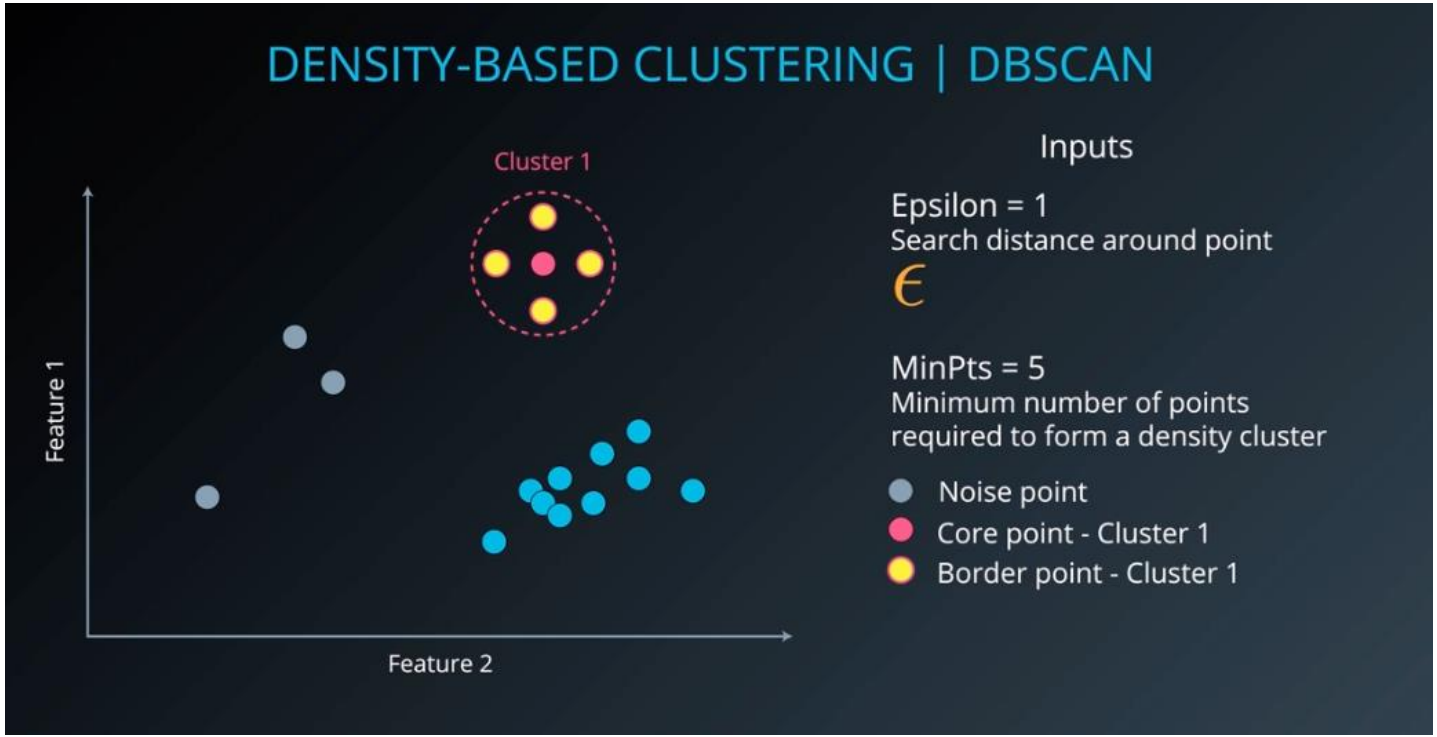


https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/

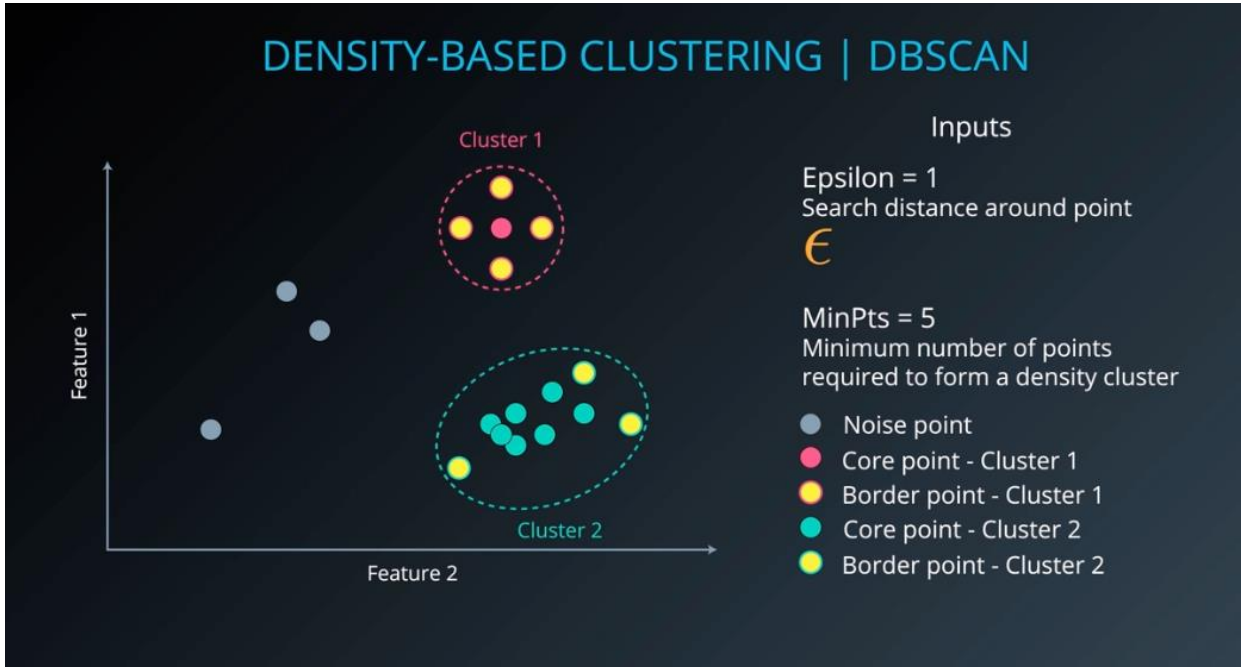
DBSCAN: Density-Based Spatial Clustering of Applications with Noise



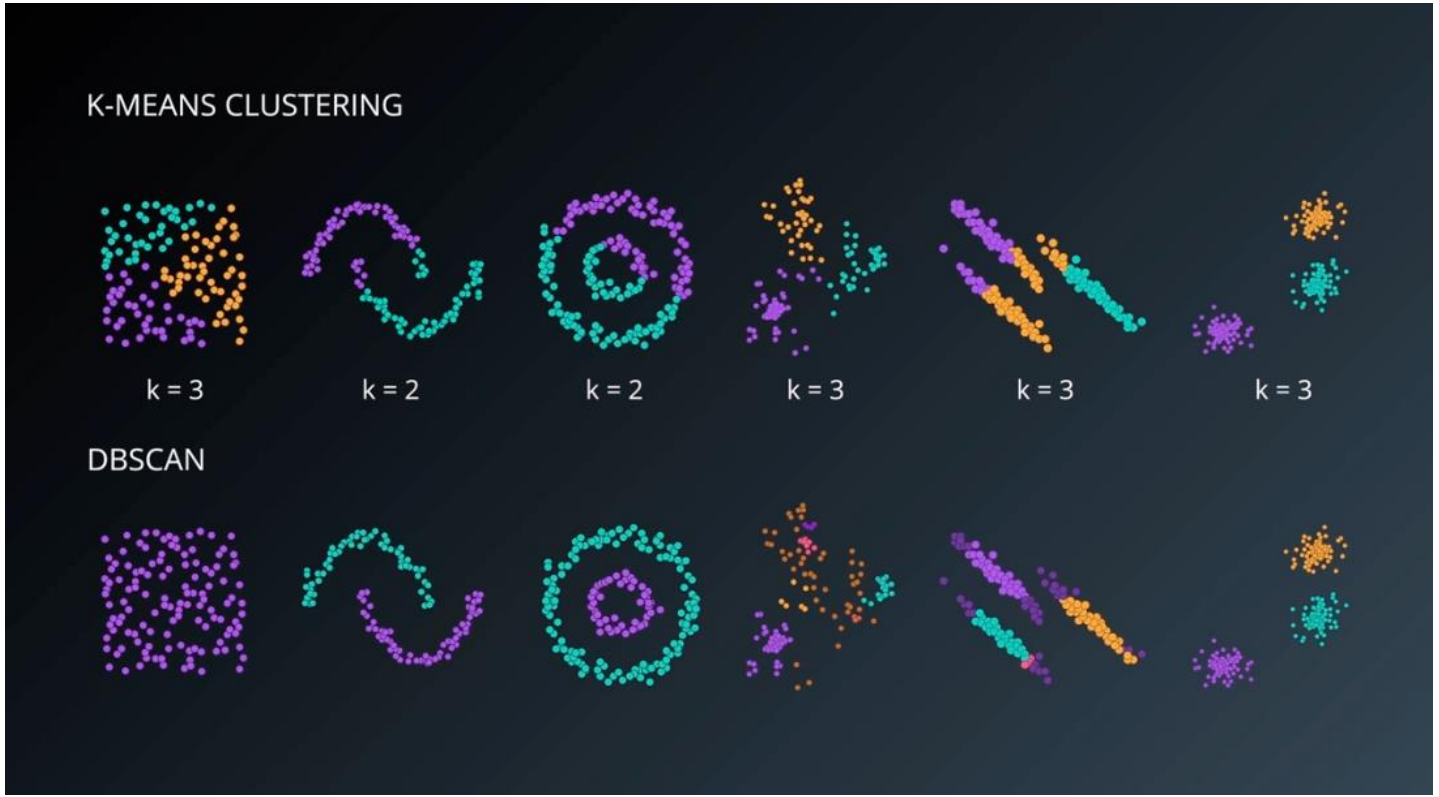
DBSCAN: Density-Based Spatial Clustering of Applications with Noise



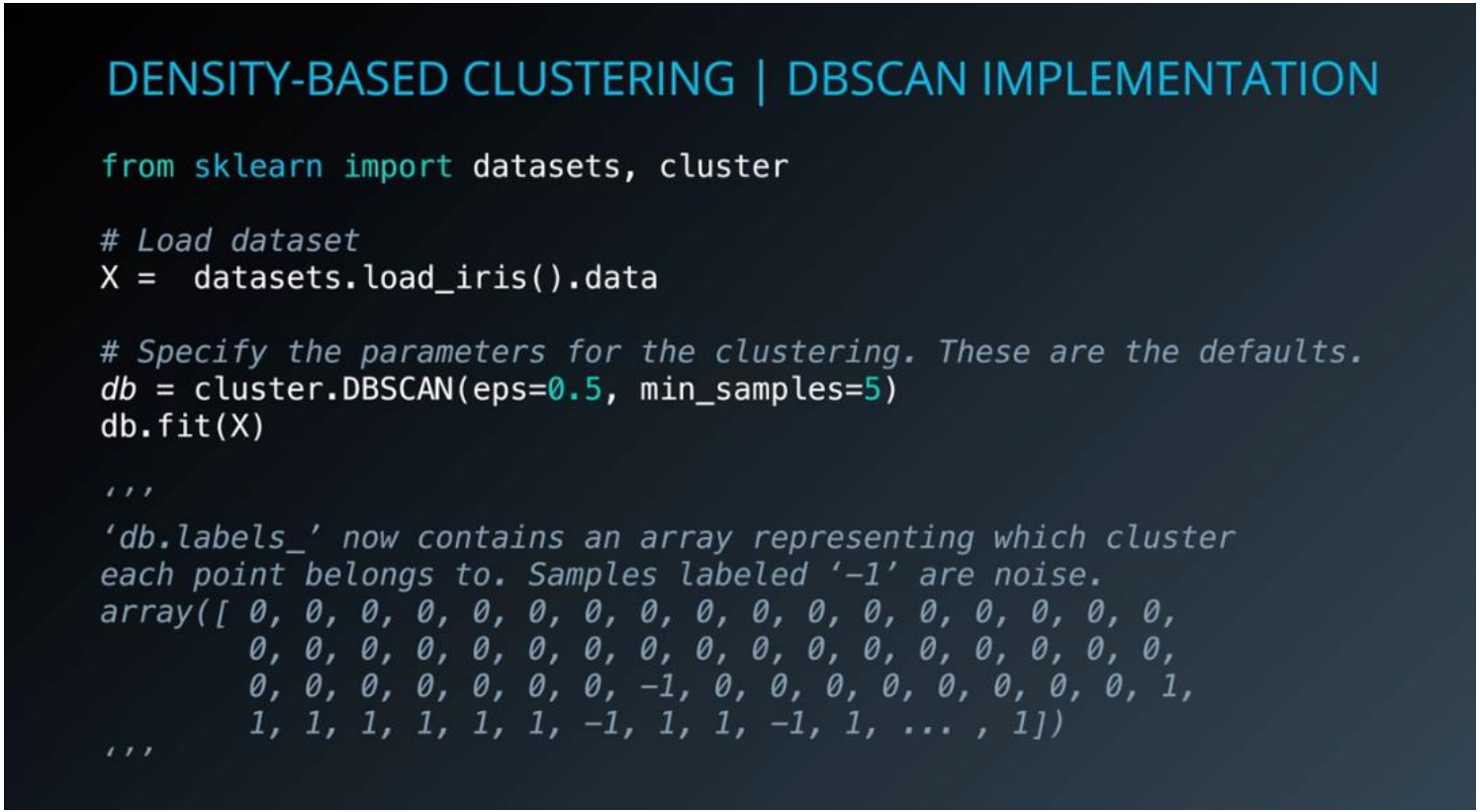
DBSCAN: Density-Based Spatial Clustering of Applications with Noise



Comparaison des méthodes de clustering



DBSCAN sur sklearn



DBSCAN: exercice

https://github.com/elhidali/EPISEN-2024/tree/main/exercice_session_5/dbscan