

Intelligence Artificielle (IA)

- Apprentissage non supervisé -

Groupe des étudiants : FISA4

Khadidja OULD AMER

E-Mail : khadidja.ouldamer@isen-ouest.yncrea.fr

Apprentissage non supervisé

Apprentissage non supervisé : Méthode d'apprentissage automatique où les données ne possèdent pas d'étiquettes (*labels*). Le système apprend par lui-même à détecter des patterns sans supervision explicite.

Plusieurs applications :



Segmentation des clients



Segmentation d'images



Systèmes de recommandation



Cybersécurité & détection de fraude

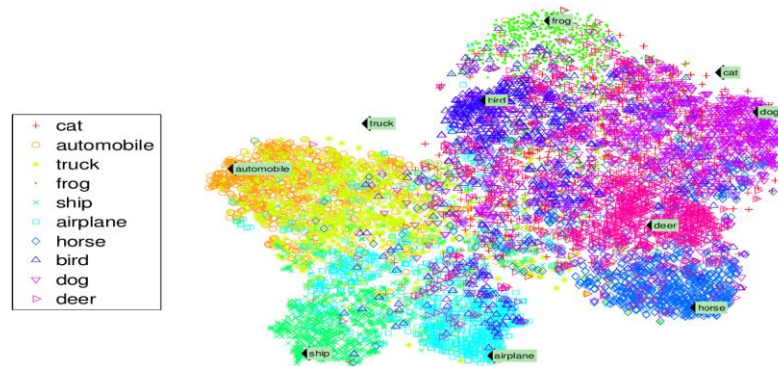


Analyse de données marketing

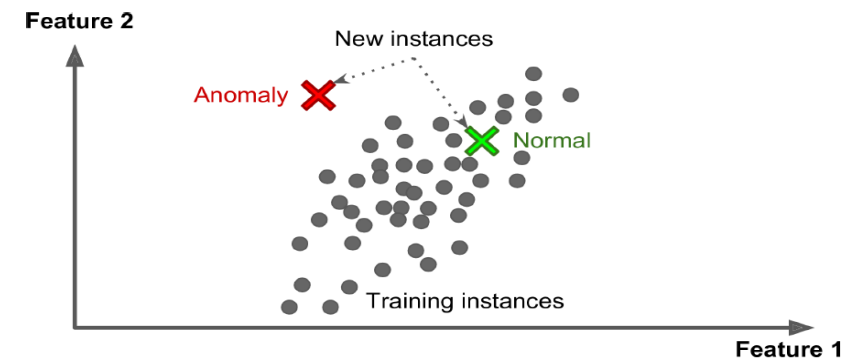
...

Apprentissage non supervisé

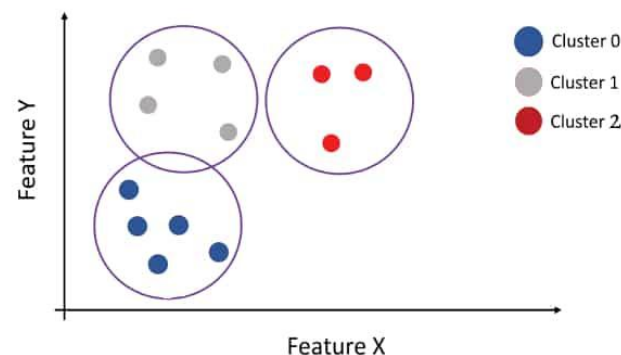
Quelques méthodes de l'apprentissage non supervisé : des approches où l'algorithme n'a pas de labels prédéfinis, et doit découvrir par lui-même des structures cachées dans les données.



Visualisation et réduction de dimensionnalité



Détection d'anomalies (ex. fraudes)



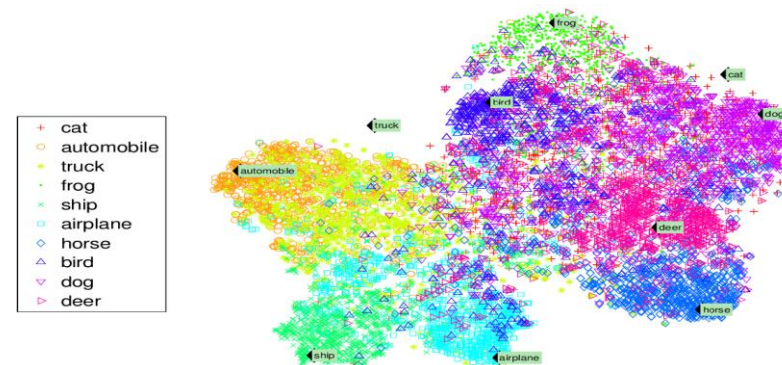
Clustering (ex. segmentation des clients)



Règles d'association (ex. systèmes de recommandations)

Réduction de dimensionnalité

$nD \rightarrow mD$
avec $n > m$



Pourquoi réduire la dimensionnalité ?

En apprentissage automatique, les données sont souvent très riches et complexes, comportant de nombreuses variables (*features*). La réduction de dimension consiste à projeter ces données dans un espace de dimension plus faible tout en préservant l'essentiel de l'information, ce qui offre plusieurs avantages :

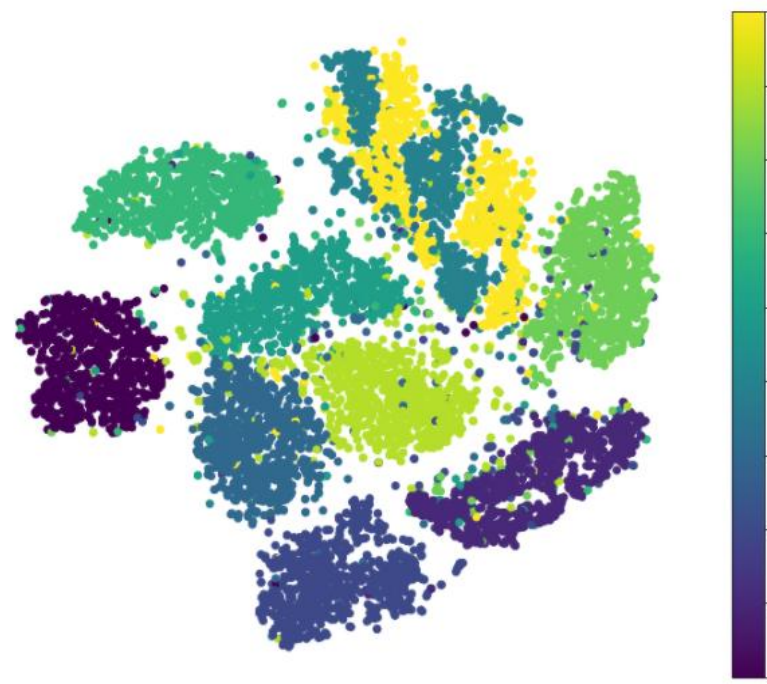
1. Visualisation des données
2. Élimination du bruit
3. Facilitation du clustering
4. Accélération du temps de calcul
5. Amélioration des performances

$nD \rightarrow mD$
avec $n > m$

Pourquoi réduire la dimensionnalité ?

1. Visualisation des données

Permet de représenter des données complexes en 2D ou 3D, facilitant ainsi leur visualisation, l'analyse exploratoire et la compréhension des structures sous-jacentes.



- Certaines informations peuvent être perdues.
- Dépend de la méthode de réduction choisie.

Exemple : Base de données de chiffres manuscrits MNIST
Projection en 2D pour visualiser la distribution des chiffres.

Pourquoi réduire la dimensionnalité ?

1. Visualisation des données

Permet de représenter des données complexes en 2D ou 3D, facilitant ainsi leur visualisation, l'analyse exploratoire et la compréhension des structures sous-jacentes.

2. Élimination du bruit

Toutes les dimensions n'apportent pas de l'information utile.

La réduction de dimension filtre les variables redondantes ou bruitées, ce qui rend les structures plus claires.

3. Faciliter le clustering et autres algorithmes non supervisés

Certains algorithmes de clustering souffrent quand la dimension est très grande.

Réduire la dimension rend les distances entre points plus significatives et améliore la qualité des regroupements.

4. Accélération du temps de calcul

Réduire le nombre de variables **peut** accélérer l'apprentissage.

Diminue les besoins en mémoire et puissance de calcul.

5. Amélioration des performances

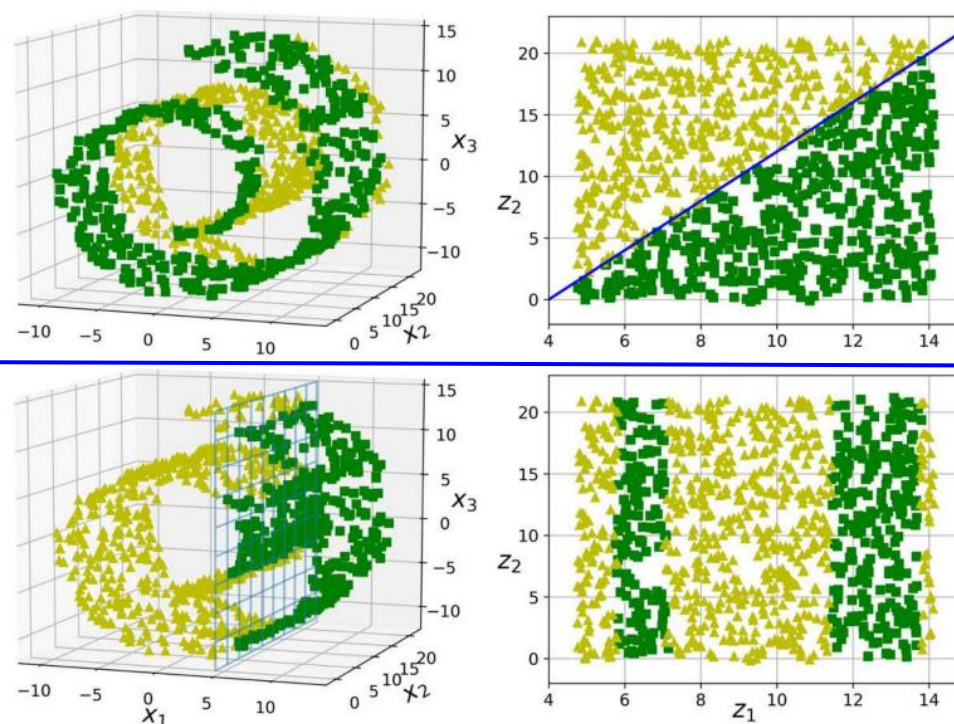
La réduction de dimension peut servir de prétraitement à d'autres tâches (supervisées ou non). Elle peut contribuer à diminuer la variance et à réduire le risque de surapprentissage, ce qui aide parfois le modèle à apprendre des représentations plus stables et à mieux se généraliser, en se concentrant sur les tendances globales plutôt que sur le bruit.



Exemple : l'arrière plan blanc des images de la base de données MNIT n'est pas nécessaire pour la phase d'apprentissage.



La réduction de dimension n'améliore pas toujours les performances ni le temps de calcul. Elle peut entraîner une perte d'informations critiques, et l'hyperplan de décision n'est pas forcément meilleur dans un espace de dimension réduite.



L'hyperplan de décision est meilleur en 2D

L'hyperplan de décision est meilleur en 3D

Peut parfois compliquer la séparation des classes et augmenter le coût computationnel.

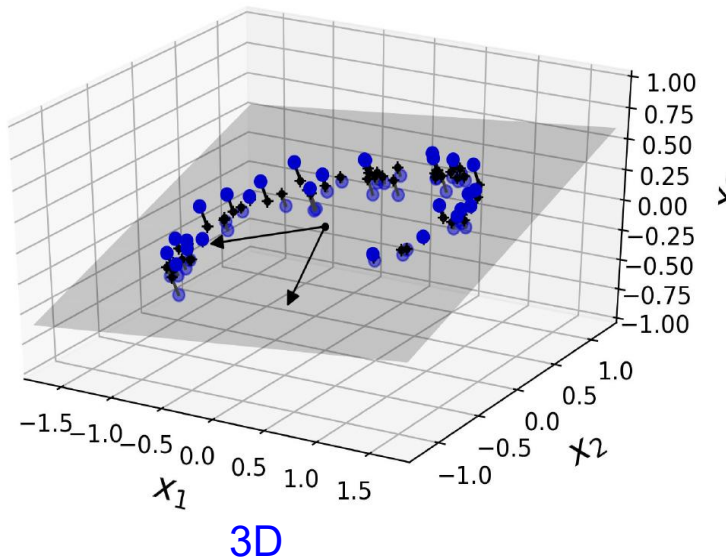
Comment réduire la dimensionnalité ?

1. La Projection

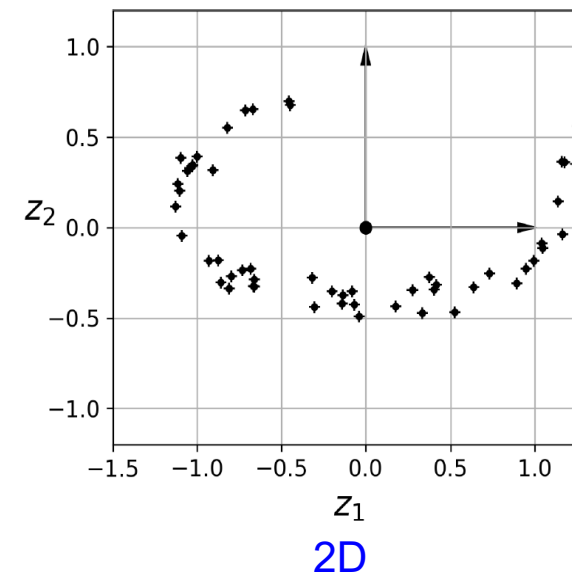
Projection des données dans un espace de plus faible dimension à l'aide de transformations linéaires ou non linéaires, tout en préservant au maximum la structure et l'information.

Quelques méthodes :

- **Principal Component Analysis (PCA)** : Trouve les directions orthogonales de plus grande variance dans les données (*méthode linéaire*)
- **Kernel PCA (KPCA)** : Extension non linéaire de la PCA, adaptée aux structures complexes et non linéaires
- **Random Projections** : Projette les données sur un sous-espace aléatoire, en préservant approximativement les distances



La projection simplifie la représentation des données tout en préservant leur structure essentielle.

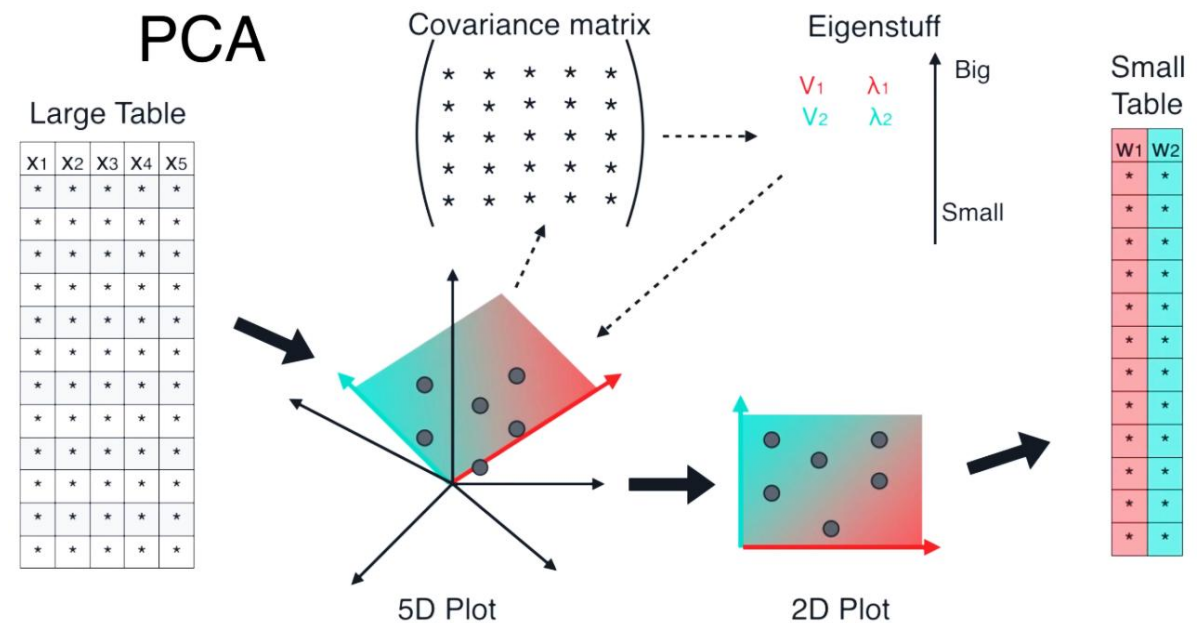


Comment réduire la dimensionnalité ?

Analyse en Composantes Principales (ACP)

L'ACP est une méthode de réduction de dimensionnalité qui transforme les variables originales en nouvelles composantes orthogonales, classées par ordre d'importance variance.

- La variance mesure la dispersion = quantité d'information
- Plus la variance est grande, plus la direction est informative
- L'ACP conserve les directions qui maximisent l'information (plus grandes variances)



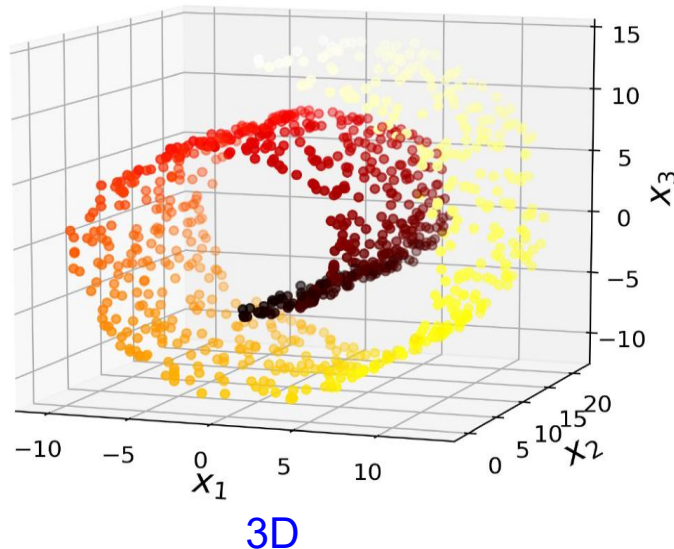
Principe de l'ACP : La matrice de covariance révèle les relations entre variables ; les vecteurs propres définissent les axes principaux (directions de plus forte variance, pondérées par les valeurs propres) ; la projection des données sur ces axes permet de réduire la dimension (ex. de 5D à 2D) tout en conservant l'information essentielle.

Comment réduire la dimensionnalité ?

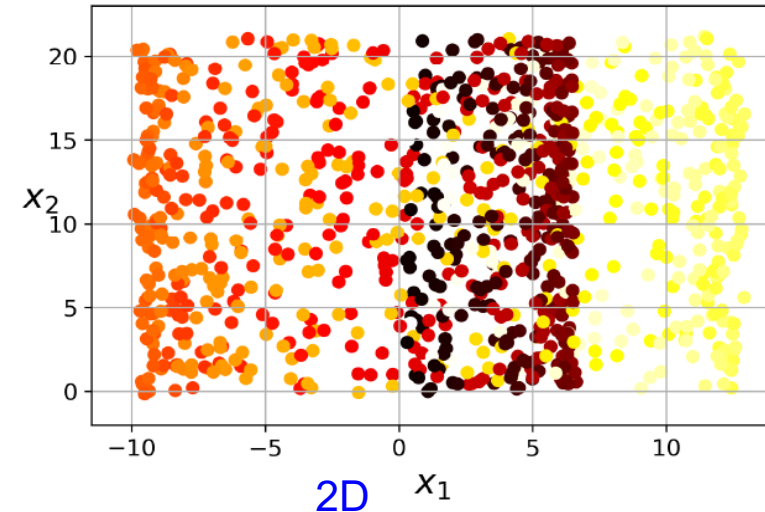
Limite des méthodes linéaires (ex: PCA)

Les projections linéaires échouent sur des structures de données **non-linéaires** complexes (ex: **Swiss Roll**)

Swiss Roll Dataset



Problème de projection
linéaire !



Perte des voisinages locaux et des relations spatiales !

Comment réduire la dimensionnalité ?

1. La projection

2. Apprentissage de Variétés (manifold learning)

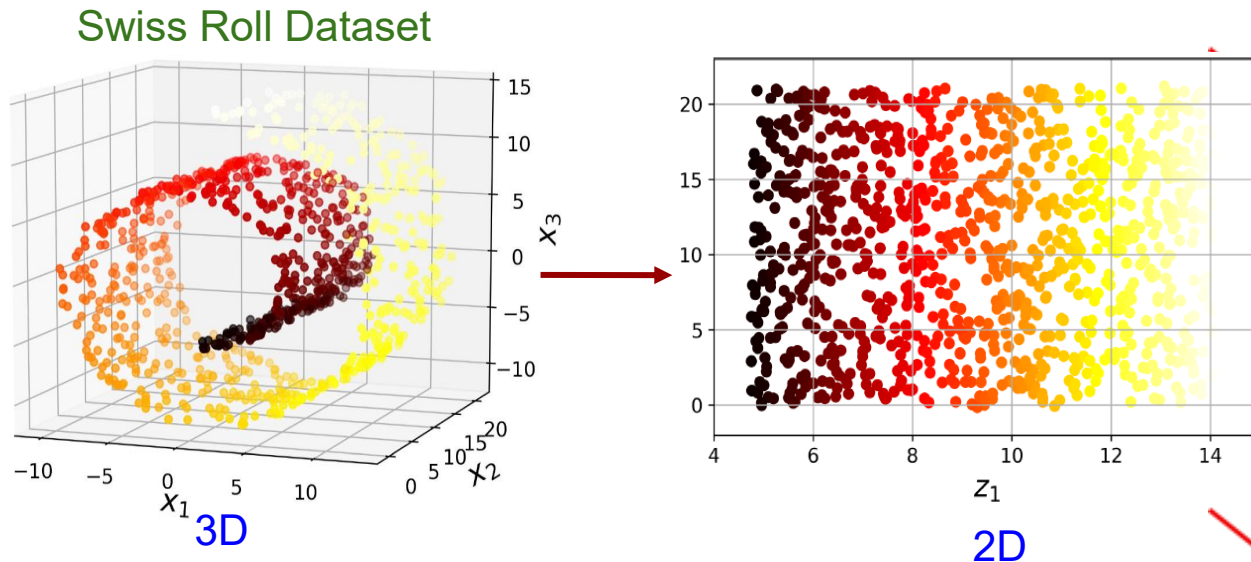
Le manifold learning part de l'hypothèse que les données vivent sur une variété (manifold) de dimension plus faible, plongée dans un espace de grande dimension.

L'objectif est de reconstruire une représentation fidèle de cette variété tout en réduisant la dimensionnalité.

Approches principales :

- Locally Linear Embedding (LLE)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- ...

<https://scikit-learn.org/stable/modules/manifold.html>



TP3



Base de données MNIST

Méthode de
réduction de la
dimensionnalité

Affichage des données

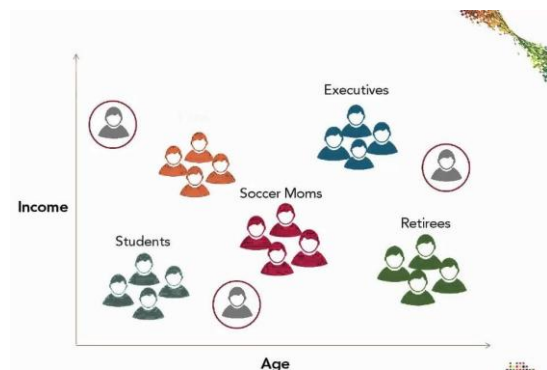
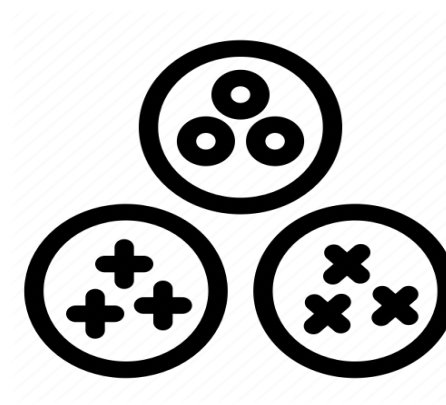
Accélération (ou pas) du
temps d'apprentissage

Clustering

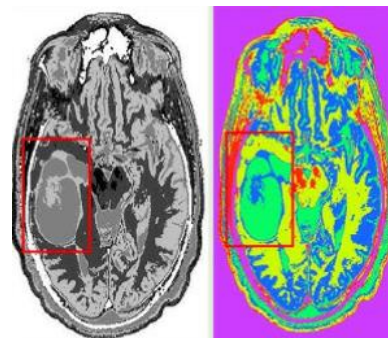
Clustering

Le clustering : Regrouper des données similaires en sous-groupes (clusters) sans étiquettes préalables

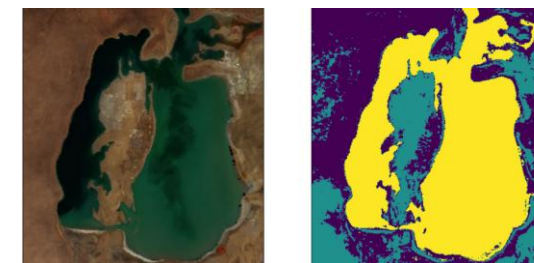
- **En clustering**, on parle de l'indice du cluster
- **En classification**, on parle de la classe (le label)



Segmentation marketing par clustering



Clustering en imagerie médicale



Clustering appliqué à une image satellite

Types de Clustering

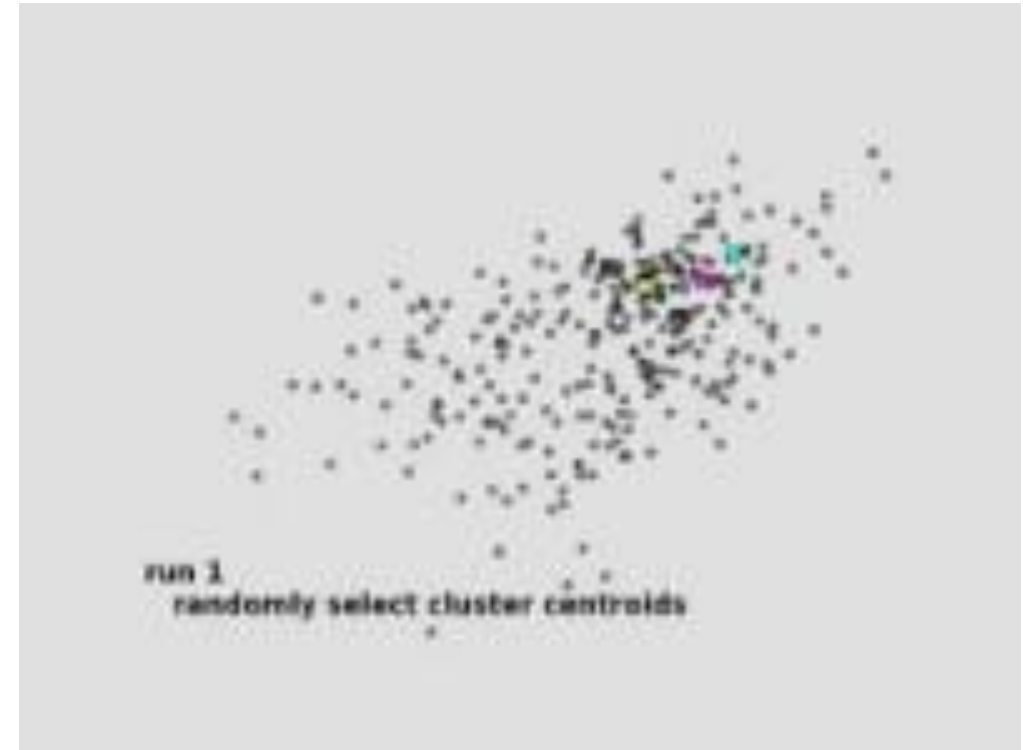
Clustering par Partitionnement (Partitioning Clustering)

Diviser un ensemble de données en un nombre prédéfini **K** de clusters distincts.

Algorithme K-means : Minimiser la somme des distances au carré entre chaque point et le centroïde de son cluster.

Étapes :

1. *Initialiser K centres de clusters de façon aléatoire*
2. *Attribuer chaque point de données au cluster dont le centre est le plus proche*
3. *Recalculer les centres de chaque cluster en prenant la moyenne des coordonnées de tous les points assignés*
4. *Répéter les étapes 2 et 3 jusqu'à convergence (lorsque les centres ne changent plus significativement)*



Types de Clustering

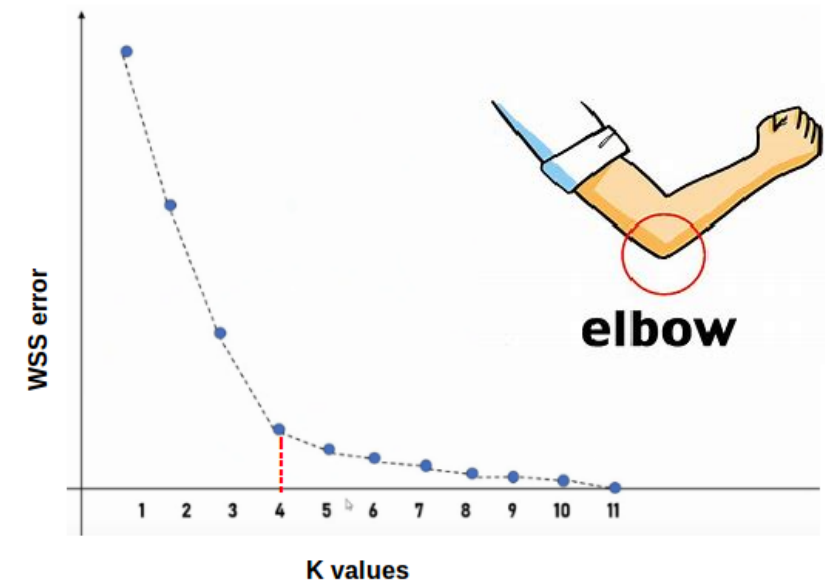
Clustering par Partitionnement : comment déterminer le nombre optimal de clusters (k)?

Méthode du Coude (Elbow Method)

On trace la courbe de la **somme des distances intra-cluster** (*inertia* ou *Within-Cluster Sum of Squares*, WCSS) en fonction du nombre de clusters k .

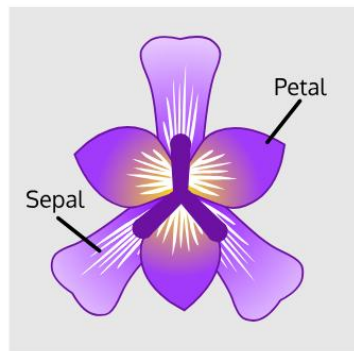
La courbe décroît puis s'aplatit : le “coude” indique le meilleur compromis entre le nombre de clusters et la qualité de regroupement.

Elbow method

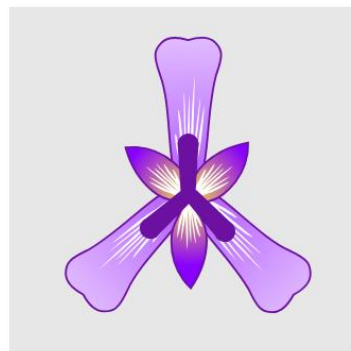


<https://medium.com/@zalarushirajsinh07/the-elbow-method-finding-the-optimal-number-of-clusters-d297f5aeb189>

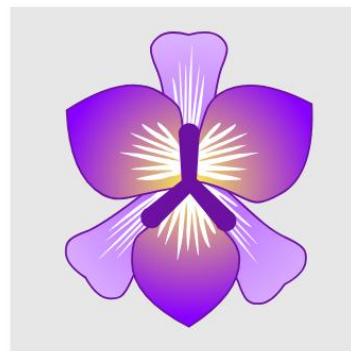
Clustering des fleurs d'Iris



Iris Versicolor



Iris Setosa



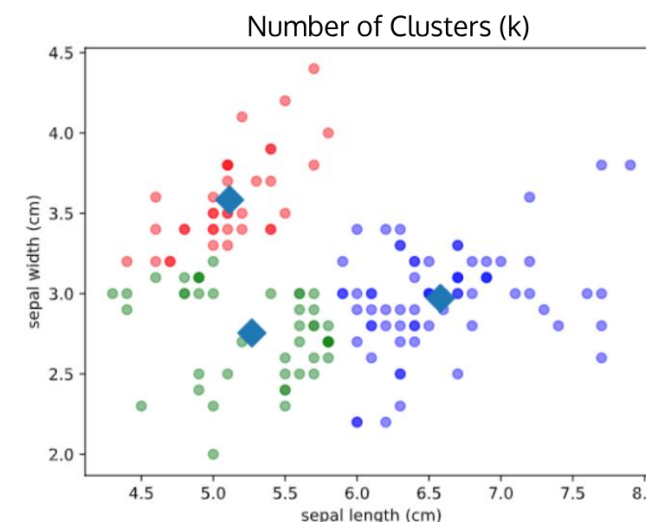
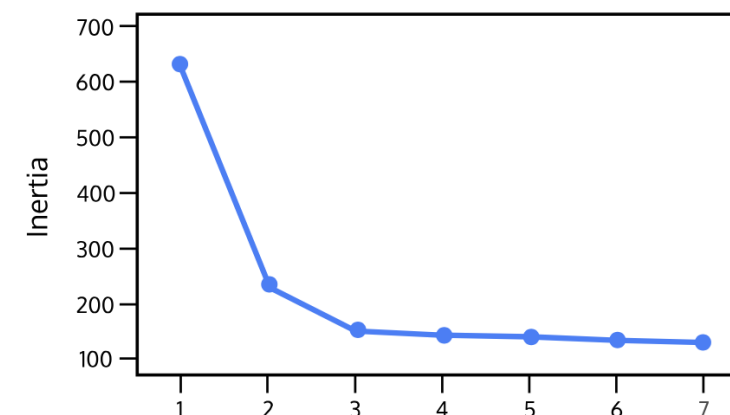
Iris Virginica

sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
...
5.0	3.6	1.4	0.2

Clustering



Optimal Number of Clusters



Ici, nous n'avons pas les noms des classes (type de fleur d'iris) dans le tableau : c'est du non supervisé.

<https://www.codecademy.com/learn/dspath-unsupervised/modules/dspath-clustering/cheatsheet>

Exemple des données de fleurs d'Iris

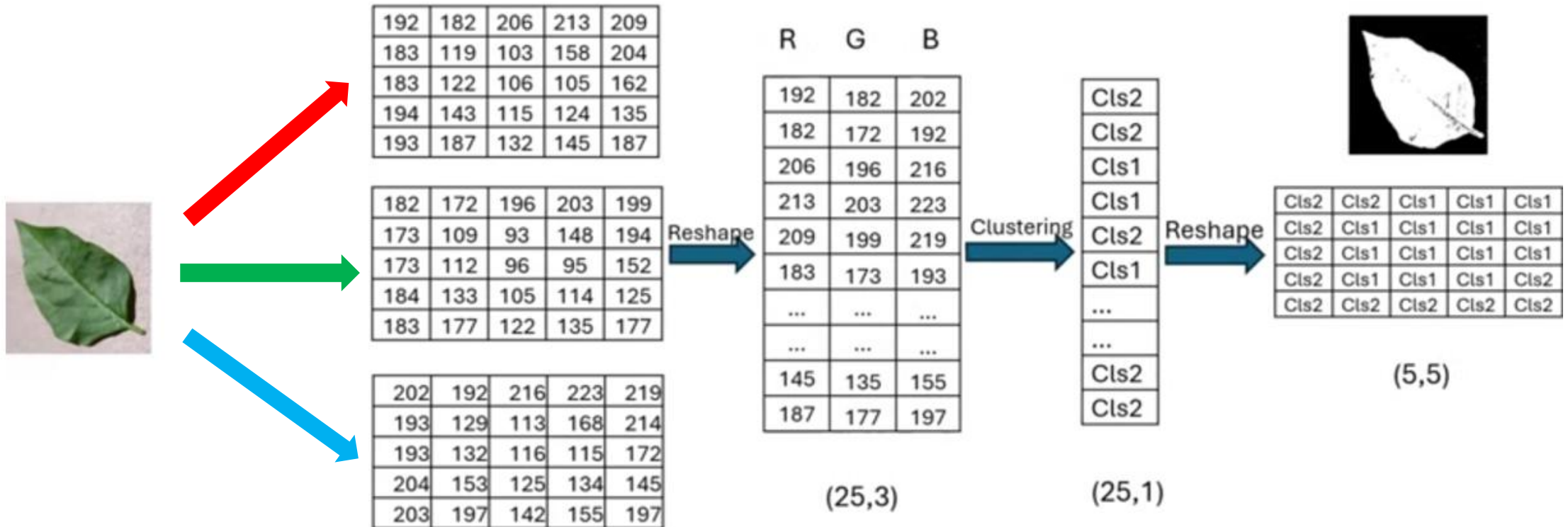


Il peut exister deux cas :

- **Apprentissage supervisé :**
On dispose des **étiquettes** (ex. *Setosa*, *Versicolor*, *Virginica*).
On peut donc entraîner un modèle de classification.
- **Apprentissage non supervisé :**
On ne possède que les **caractéristiques numériques** (longueur/largeur des sépales et pétales), **pas les étiquettes**.
On ne peut pas « prédire une classe connue », mais on peut rechercher des structures ou regroupements naturels dans les données.

Exemple de clustering (segmentation d'image)

✓ Image couleur dans l'espace RGB: (5,5,3)

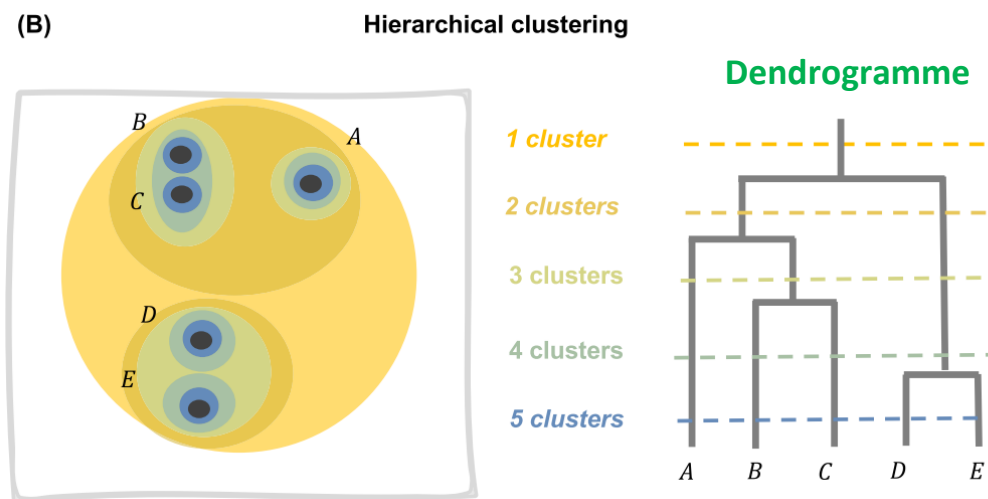


Exemple de clustering appliqué à une image – chaque pixel est transformé en vecteur (R,G,B), puis regroupé par similarité de couleur (clustering). Après reshape inverse, on obtient une image segmentée en classes homogènes, permettant de distinguer la feuille du fond.

Types de Clustering

Clustering hiérarchique (Hierarchical Clustering)

Les données sont organisées selon une hiérarchie de groupes imbriqués. Cette méthode construit une arborescence de clusters représentée sous forme de **dendrogramme**.



<https://blog.gopenai.com/how-hierarchical-clustering-works-d6e66c77102b>

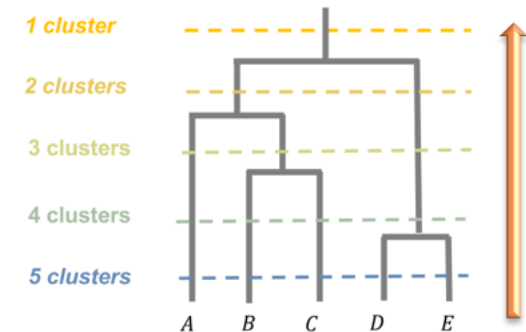
Types de Clustering

Clustering hiérarchique (Hierarchical Clustering)

Les données sont organisées selon une hiérarchie de groupes imbriqués.
Cette méthode construit une arborescence de clusters représentée sous forme de dendrogramme.

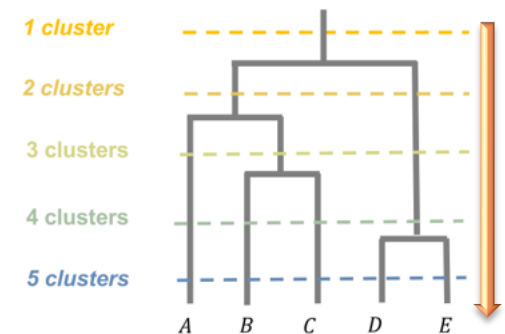
➤ Clustering agglomératif (bottom-up)

Chaque point de données est initialement considéré comme un cluster.
Les clusters les plus proches sont fusionnés progressivement en fonction de leurs similarités.



➤ Clustering divisif (top-down)

Commence avec un seul cluster contenant tous les points
Divise ensuite ce cluster en sous-clusters de plus en plus petits



Types de Clustering

Clustering basé sur la densité (*Density-Based Clustering*)

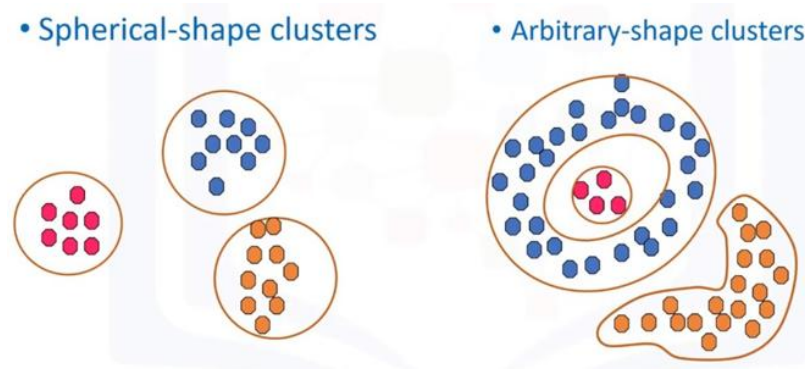
Regroupe les points de données en fonction de la **densité locale**.

Idée principale :

- Les régions de forte densité sont considérées comme des clusters.
- Les régions de faible densité sont considérées comme du bruit (outliers).

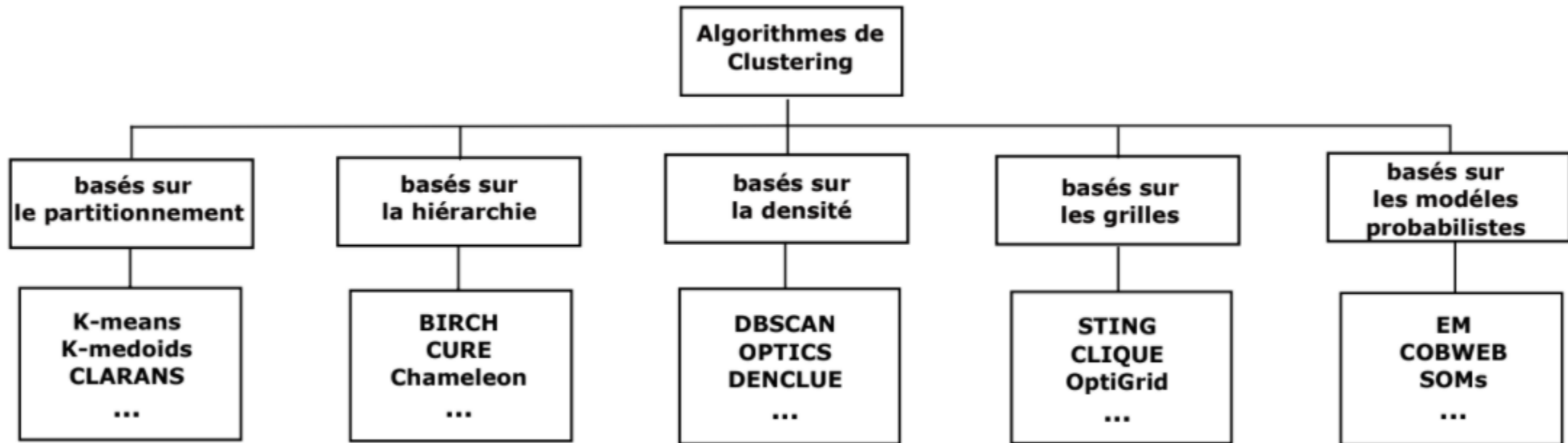
Algorithmes :

- DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- Mean Shift
-



<https://www.youtube.com/watch?v=6jl9KkmgDIw>

Types de Clustering



Source : Rehioui, H., & Idrissi, A. (2019). **New clustering algorithms for twitter sentiment analysis**. IEEE Systems Journal, 14(1), 530-537.

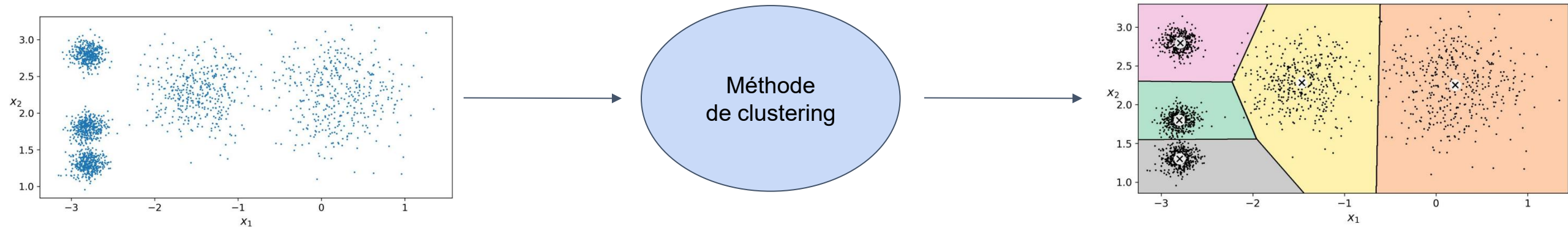
Mesures de performance

	Mesures	Description	Valeur	Performance
internes	Indice de Dunn (DI)	Cet indice évalue le degré de séparation entre les individus d'un même cluster, c'est-à-dire la similarité intra-cluster.	↗	↗
	Indice de David Bouldin (DBI)	Cet indice, similaire à DI, évalue également le degré de séparation, mais cette fois ci entre les clusters (dissimilarité inter-clusters)	↘	↗
	Compactness (CP)	CP mesure la distance moyenne entre chaque paire dans le cluster, puis entre tous les clusters ; les membres de chaque cluster devraient être aussi proches les uns des autres que possible.	↘	↗
externes	Précision (CA)	CA mesure le pourcentage d'objets correctement classifiés dans un cluster, en se basant sur des étiquettes prédéfinies.	↗	↗
	Information Mutuelle Normalisée (NMI)	cet indice permet de mesurer les informations statistiques partagées par les points représentant les affectations des clusters et les affectations d'étiquettes prédéfinies des instances. La valeur de NMI varie entre 0 et 1.	↗	↗
	Entropy	C'est le degré auquel chaque cluster est constitué d'objets appartenant à une seule classe (objets étiquetés par la même étiquette).	↘	↗

Source : Rehioui, H., & Idrissi, A. (2019). **New clustering algorithms for twitter sentiment analysis**. IEEE Systems Journal, 14(1), 530-537.

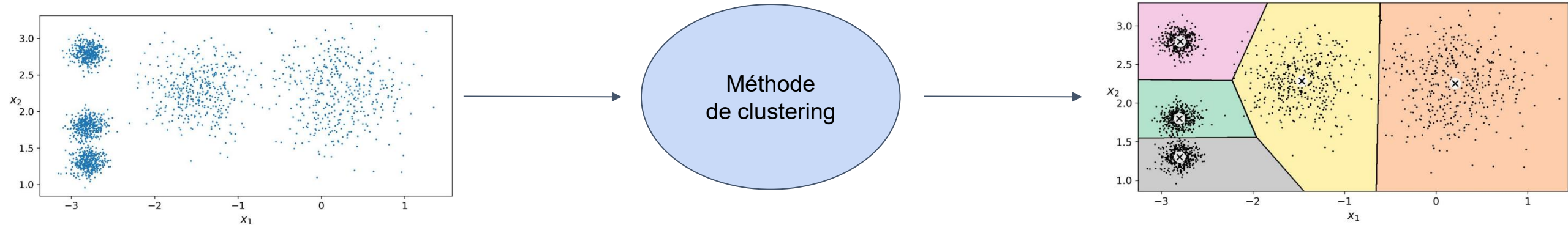
TP4

■ Partie 1



TP4

■ Partie 1



■ Partie 2

Images faciales

