**Groupe 2 : Clustering hiérarchique**

TIGA Abdoul-Wakilou

BAH WARAKPE Maguibou

BELLO Conceptia

1. Explication de l’algorithme
2. Application à un cas concret
3. Cas d’utilisation de l’algorithme du clustering hiérarchique
4. Différences avec les autres algorithmes de clustering

**1- Explication de l’algorithme :**

L'algorithme de clustering hiérarchique commence par considérer chaque point de données comme un cluster individuel. Ensuite, il calcule itérativement la similarité entre chaque paire de clusters et fusionne les clusters les plus similaires pour former des clusters plus grands. Ce processus est répété jusqu'à ce qu'un seul cluster contenant toutes les données soit obtenu.

Il existe deux approches principales dans l'algorithme de clustering hiérarchique :

- Agrégatif (ou ascendante) : Dans cette approche, chaque point de données est initialement considéré comme un cluster séparé, puis les clusters sont fusionnés progressivement en fonction de leur similarité. Cette méthode produit un dendrogramme, une structure arborescente représentant la hiérarchie des clusters.

- Divisif (ou descendante) : Contrairement à l'approche agrégative, cette méthode commence avec un seul cluster contenant tous les points de données et divise récursivement ce cluster en sous-clusters plus petits jusqu'à ce que chaque point de données soit dans son propre cluster.

**2- Application à un cas concret :**

Imaginons que nous ayons un ensemble de données de points GPS représentant les emplacements des magasins dans une ville. Nous pourrions utiliser l'algorithme de clustering hiérarchique pour regrouper ces magasins en fonction de leur proximité les uns aux autres. Cela pourrait être utile pour identifier des zones commerciales ou pour optimiser la distribution de la publicité en fonction de la densité de magasins.

**3- Cas d’utilisation de l’algorithme du clustering hiérarchique :**

- **Segmentation de marché :** Identifiez les groupes de clients ayant des comportements d'achat similaires.

**- Biologie :** Classifiez les espèces en groupes en fonction de leurs caractéristiques génétiques ou morphologiques.

**- Analyse de réseaux sociaux :** Identifiez les communautés ou les groupes d'utilisateurs similaires dans un réseau social en fonction de leurs interactions.

**- Cartographie :** Regroupez les zones géographiques similaires en fonction de diverses caractéristiques telles que la démographie ou les infrastructures.

**4- Différences avec les autres algorithmes de clustering :**

**- K-Means :** L'algorithme de clustering hiérarchique ne nécessite pas de spécifier le nombre de clusters à l'avance, contrairement à K-Means. De plus, il produit une hiérarchie de clusters plutôt qu'une partition fixe des données.

**- DBSCAN :** Contrairement à DBSCAN, qui identifie des clusters de formes arbitraires en fonction de la densité des points, l'algorithme de clustering hiérarchique regroupe les points en clusters hiérarchiques basés sur leur similarité.

**- Méthodes de partitionnement spectral :** Ces méthodes tentent de partitionner les données en sous-ensembles en fonction des relations entre les points dans un espace de dimension réduit. L'algorithme de clustering hiérarchique, en revanche, crée une hiérarchie de clusters en fusionnant ou en divisant récursivement les clusters existants.