Types of Clustering Algorithms

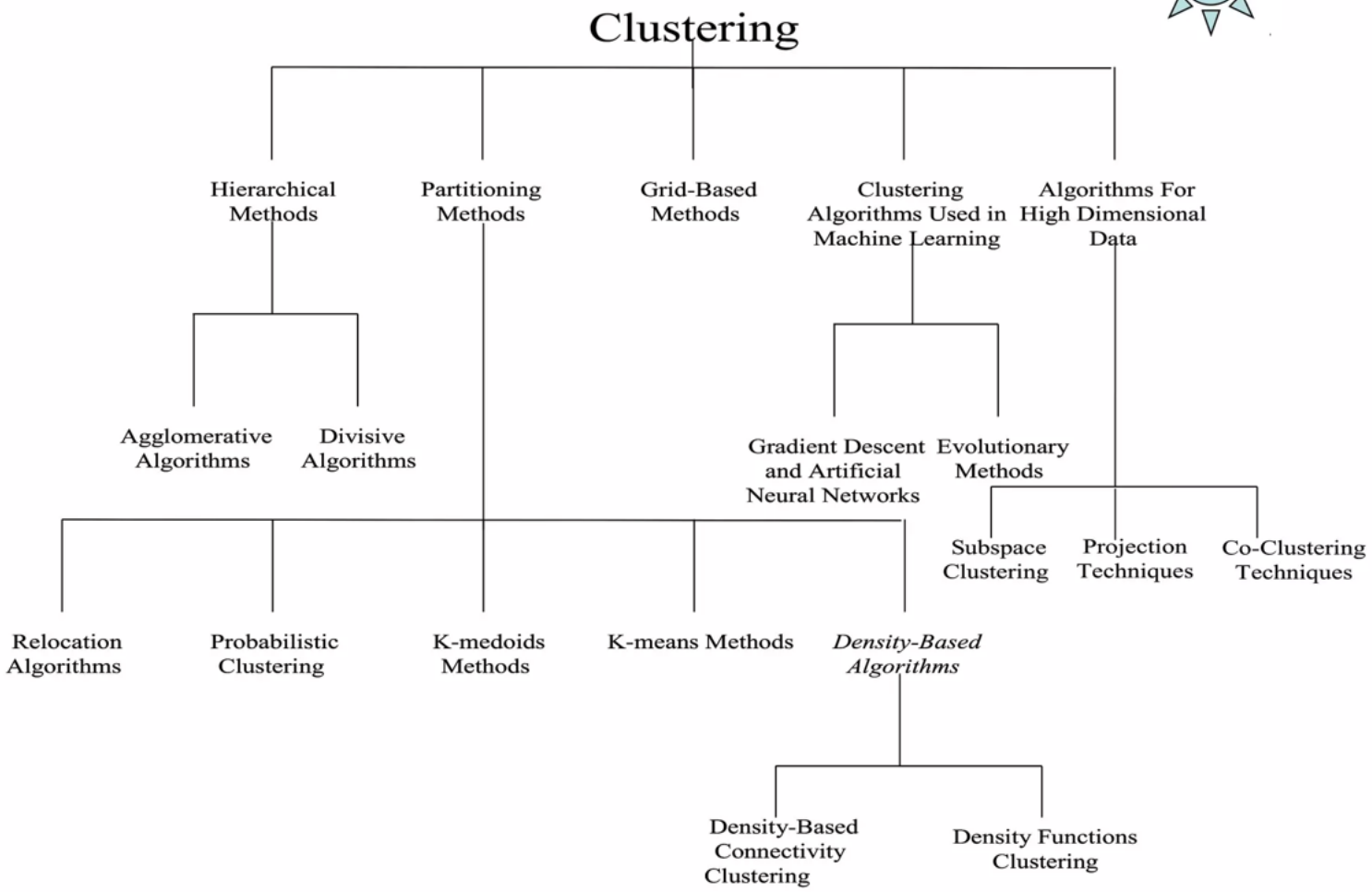

Table of Contents

1. How clustering is different from classification
2. Applications of Clustering
3. What are density based methods
4. What are Hierarchial based methods
5. What are partitioning methods
6. What are Grid Based methods
7. Main Requirements for Clustering Algorithms

Unsupervised learning Algorithms. (Pg no of 5 Reference)

i) K-Means
ii) Hierarchical clustering
iii) DBSCAN
} clustering - Evaluation metrics for clustering are important...

iv) Performance measurement
v) Principal component Analysis
vi) Dimensionality Reduction.

(*) clustering :-

→ First we need to understand clustering is different from classification.

Parameter	classification	clustering
1) Type	used for supervised learning.	used for unsupervised learning.
2) Basic	Process of classifying the input instances based on their corresponding "labels"	Grouping the instances based on their similarity without the help of class "labels"
3) complexity	more complex	less complex
4) Example Algo	Logistic Regression, Naive Bayes, SVM etc— KNN	K-means clustering, Hierarchical clustering, DBScan, BIRCH Algorithm etc

Applications of clustering

→ For customer segmentation:

You can cluster your customers based on their purchases, their activity on website and so And this can be used in Recommender systems to suggest content that other users in same cluster enjoyed (Bingo).

Eg: marketing, Insurance, Libraries etc.

→ Earth-quake studies:

By learning old data it can make clusters and determine dangerous zones.

→ For search engines:

As you search, similar images would end up on same cluster.

→ To segment an Image:

By clustering pixels according to their color, then replacing each pixel's color with the mean color of its cluster, it is possible to reduce no. of different colors in image.

Eg: used in object detection & tracking systems.

→ For Anamoly detection (outliers)

→ As a dimensionality Reduction Technique.

→ For Data Analysis

clustering methods

① Density-Based methods:-

These methods consider the clusters as th dense region having some similarity and different from the lower dense region or spa

These methods have good accuracy & ability to merge two clusters.

Eg:-

i) DBSCAN [Density Based Spatial Clustering of Applications with Noise]

ii) OPTICS [Ordering Points to Identify Clustering Structure.]

② Hierarchial-Based methods:-

The clusters formed in this method forms a tree-type structure based on the hierarchy. New clusters are formed using the previously formed one.
It has two categories

i) Agglomerative [Bottom up approach - forms a single cluster & expandi

ii) Divisive [Top down approach - forms a big cluster & divides]

Eg:

iii) CURE [Clustering Using Representatives]

iv) BIRCH [Balanced Iterative Reducing Clustering and using Hierarchies]

③ Partitioning methods:-

- these methods partition the objects into k-clusters and each partition forms one cluster. this method is used to optimize an objective criterion similarity function such as when the distance is a major parameter

Eg:
i) K-means
ii) CLARANS [Clustering Large Applications based upon Randomized Search]

④ Grid-Based methods:-

In this method the data space is formulated into a finite no. of cells that form a grid like structure. All the clustering operation done on these grids are fast and independent of the no. of data objects.

Eg:
i) STING [statistical Information Grid]
ii) CLIQUE [Clustering In QUEst]
iii) Wave cluster, etc.

Main Requirements for clustering Algorithms

1. It should be scalable.
2. It should be able to deal with attributes of different types.
3. It should be able to discover arbitrary shape clusters.
4. It should have an inbuilt ability to deal with Noise & outliers.
5. The cluster should not vary with the order of input records.
6. It should be able to handle data of high dimensions.
7. It should be easy to interpret & use.