

# \*GELOMER\*TIVE CLUSTERING



#### What is Agglomerative Clustering?



- The agglomerative clustering is the most common type of hierarchical clustering used to group objects in clusters based on their similarity.
- It's also known as AGNES (Agglomerative Nesting).
- Pairs of clusters are successively merged until all clusters have been merged into one big cluster containing all objects.

### Classification of Clustering Procedures (1)

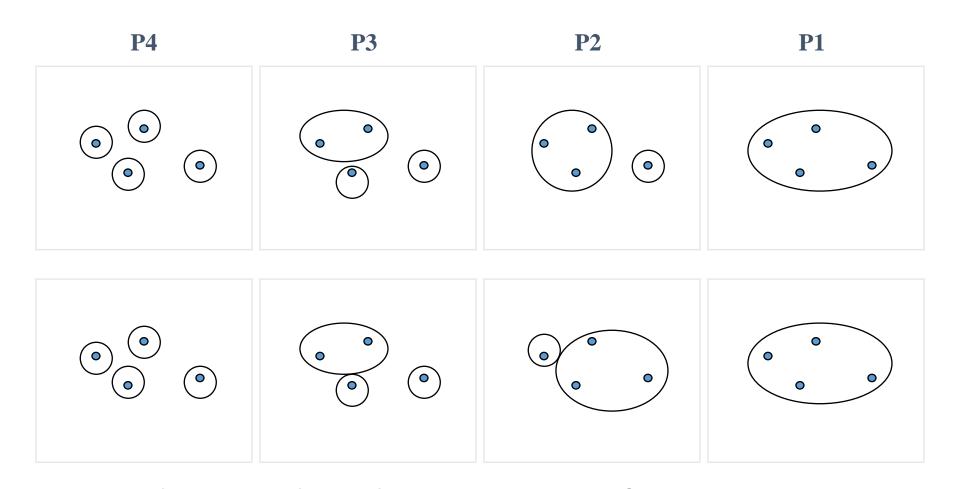
**ML** Labs Pvt Ltd **Clustering Procedures** Hierarchical **Nonhierarchical Agglomerative** Divisive Centroid Sequential Paralle **Optimizing** Varianc Linkage Methods Methods Methods **Threshold** Thre\$hold Partitioning Ward' Method Single Complete Averag Linkage Linkage Linleage





- When 'n' objects or samples present
- Start with partition P<sub>n</sub>, where each object forms its own cluster.
- Merge the two closest clusters, obtaining  $P_{n-1}$ .
- Repeat merging until only one cluster is left.





**Top:** hierarchical sequence of partitions **Bottom:** non-hierarchical sequence

#### **Hierarchical Clustering Methods**



- -Hierarchical clustering is characterized by the development of a hierarchy or tree-like structure.
- 1. Agglomerative clustering
- 2. Divisive clustering
  - -Agglomerative clustering starts with each object in a separate cluster.

Clusters are formed by grouping objects into bigger and bigger clusters.

#### **Hierarchical Clustering Methods**



- -Divisive clustering starts with all the objects grouped in a single cluster.
- Clusters are divided or split until each object is in a separate cluster.
- Agglomerative methods are commonly used in marketing research.
- They consist of linkage methods, variance methods, and centroid methods.

#### Agglomerative-Linkage Methods



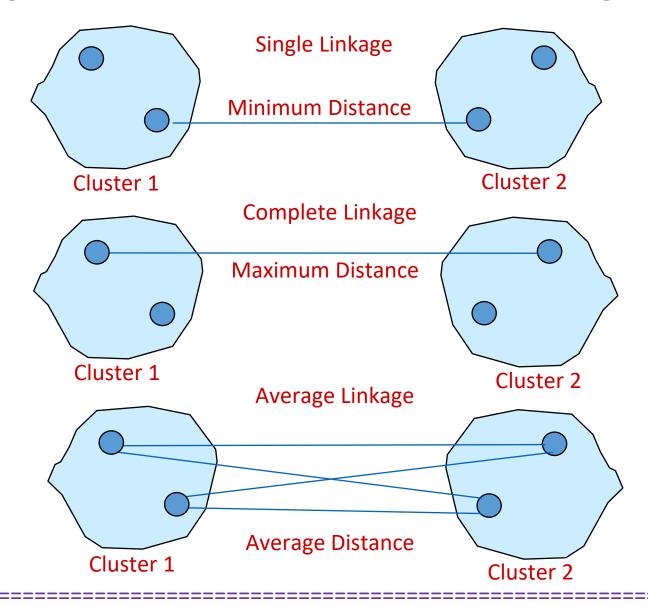
 The single linkage method is based on minimum distance, or the nearest neighbor rule.

 The complete linkage method is based on the maximum distance or the furthest neighbor approach.

 The average linkage method the distance between two clusters is defined as the average of the distances between all pairs of objects

#### Linkage Methods of Clustering







#### Agglomerative Clustering-Variance Method

 Variance methods generate clusters to minimize the within-cluster variance.

Ward's procedure is commonly used variance method. For each cluster, the sum of squares is calculated. The two clusters with the smallest increase in the overall sum of squares within cluster distances are combined.

ML Labs Pvt Ltd | WhatsApp: 91-7338339898 | www.analytics6.com | bharath@themllabs.com

#### Agglomerative Clustering-Centroid Method



- In the centroid methods, the distance between two clusters is the distance between their centroids.
- Of all the hierarchical methods, average linkage and Ward's methods have been shown to perform better than other procedures.

## Agglomerative Clustering-Variance and Centroid Methods



