

edunet
foundation

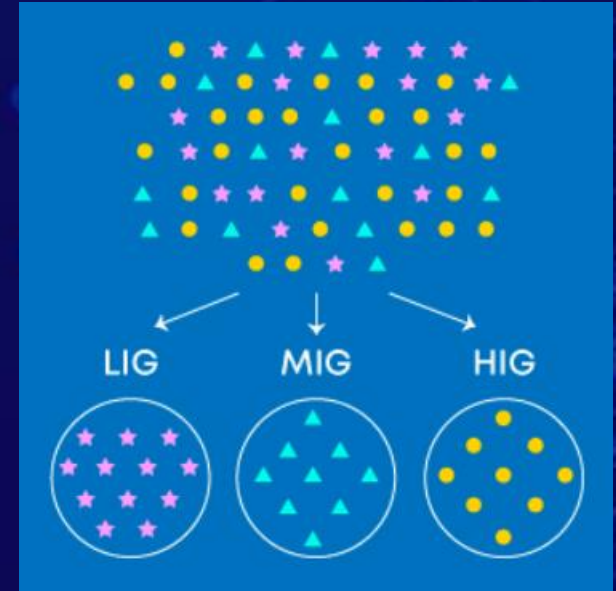
Proudly supported by

SAP



Unit 3.3

Hierarchical Clustering



Disclaimer

The content is curated from online/offline resources and used for educational purpose only

Applications of hierarchical clustering

1. Customer segmentation:
2. Image segmentation:
3. Text analysis:
4. Gene expression analysis:
5. Anomaly detection:
6. Recommendation systems:
7. Risk assessment:
8. Network analysis:
9. Market segmentation:
10. Outlier detection

Learning Objectives

- Introduction
- Types of Hierarchical Clustering
- Agglomerative Hierarchical Clustering
- Steps for Agglomerative Hierarchical Clustering
- Proximity Matrix
- Linkage Methods
- How Agglomerative Clustering work?
- Divisive Hierarchical Clustering
- Steps in Divisive Hierarchical Clustering
- Application of Hierarchical Clustering
- Hands On



Hierarchical Clustering: Introduction

- Hierarchical clustering is unsupervised machine learning algorithm, which is used to group the unlabelled datasets into a cluster.
- It creates groups so that objects within a group are like each other and different from objects in other groups.
- Clusters are visually represented in a hierarchical tree called a dendrogram.
- Sometimes the results of K-means clustering, and hierarchical clustering may look similar, but they both differ depending on how they work.
- An Example of Hierarchical clustering shown in Image.

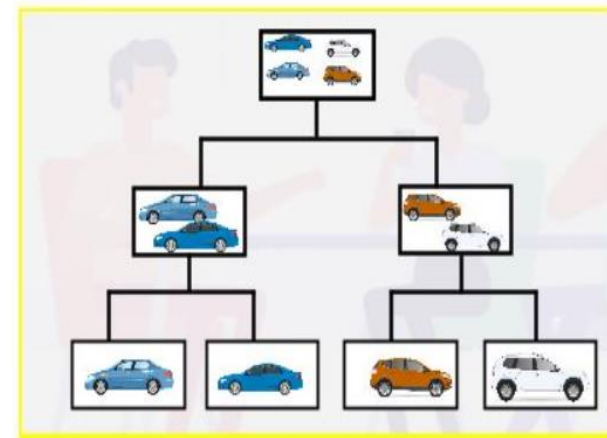


Image: Set of Cars

[Reference](#)

Types of Hierarchical Clustering

Hierarchical clustering is divided into:

- Agglomerative
- Divisive

Agglomerative Clustering: Agglomerative clustering is known as a bottom-up approach. Consider it as bringing things together.

Divisive Clustering: Divisive clustering is known as the top-down approach. We take a large cluster and start dividing it into two, three, four, or more clusters.

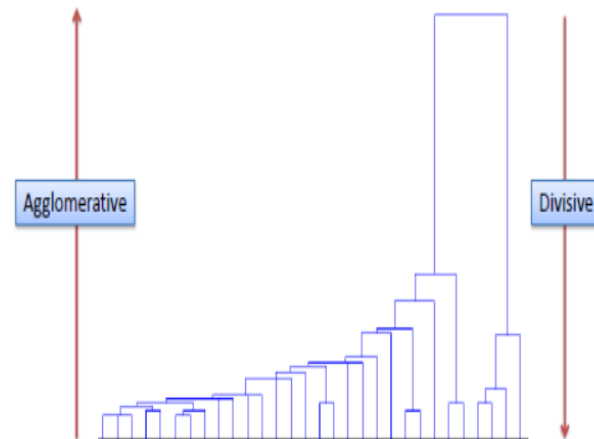
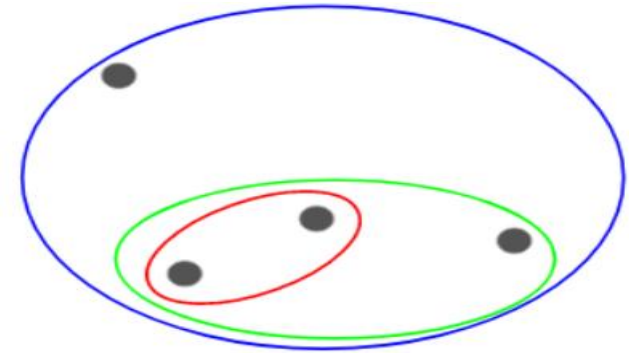


Image: Hierarchical Clustering
[Reference](#)

Agglomerative Hierarchical Clustering

- Agglomerative Hierarchical Clustering is popularly known as a bottom-up approach, where in each data or observation is treated as its cluster.
- A pair of clusters are combined until all clusters are merged into one big cluster that contains all the data.
- In Agglomerative Hierarchical clustering, at each step we are merging or adding clusters. Hence it is also known as additive hierarchical clustering.



Steps for Agglomerative Clustering

The steps for agglomerative clustering are as follows:

1. Start assigning each observation as a single point cluster, so that if we have N observations, we have N clusters, each containing just one observation.
2. Compute the proximity matrix using a distance metric.
3. Use a linkage function to group objects into a hierarchical cluster tree based on the computed distance matrix from the above step.
4. Data points with close proximity are merged together to form a cluster and update proximity matrix.
5. Repeat steps 3 and 4 until a single cluster remains.

Proximity Matrix

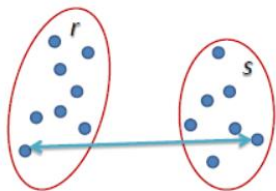
- The proximity matrix is a matrix consisting of the distance between each pair of data points.
- The distance is computed by a distance function. Euclidean distance is one of the most commonly used distance functions.
- The given proximity matrix consists of n points named x_i , and the $d(x_i, x_j)$ represents the distance between the points.
- In order to group the data points in a cluster, a linkage function is used where the values in the proximity matrix are taken and the data points are grouped based on similarity.

	x_1	x_2	x_3	...	x_n
x_1	$d(x_1, x_1)$	$d(x_1, x_2)$	$d(x_1, x_3)$...	$d(x_1, x_n)$
x_2	$d(x_2, x_1)$	$d(x_2, x_2)$	$d(x_2, x_3)$...	$d(x_2, x_n)$
x_3	$d(x_3, x_1)$	$d(x_3, x_2)$	$d(x_3, x_3)$...	$d(x_3, x_n)$
...
x_n	$d(x_n, x_1)$	$d(x_n, x_2)$	$d(x_n, x_3)$		$d(x_n, x_n)$

[Image: Proximity matrix of \$n\$ points](#)

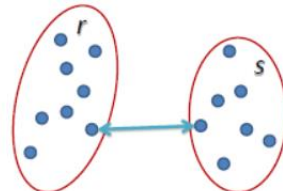
Linkage Methods

- **Complete linkage:** The distance between two clusters is defined as the longest distance between two points in each cluster.
- **Single linkage:** In single linkage hierarchical clustering, the distance between two clusters is defined as the shortest distance between two points in each cluster.



$$L(r, s) = \max(D(x_{ri}, x_{sj}))$$

[Image: Complete linkage](#)

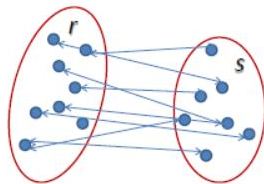


$$L(r, s) = \min(D(x_{ri}, x_{sj}))$$

[Image: Single linkage](#)

Linkage Methods

- **Average linkage:** The average of all pairwise distances between elements in each pair of clusters is used to measure the distance between two clusters.
- **Centroid linkage:** Before merging, the distance between the two clusters' centroids are considered.



$$L(r, s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} D(x_{ri}, x_{sj})$$

[Image: Average linkage](#)

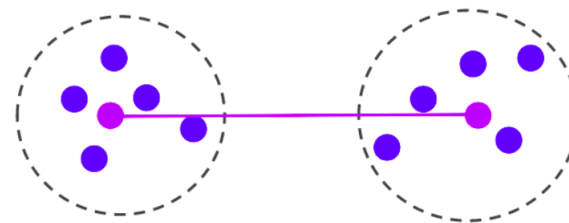
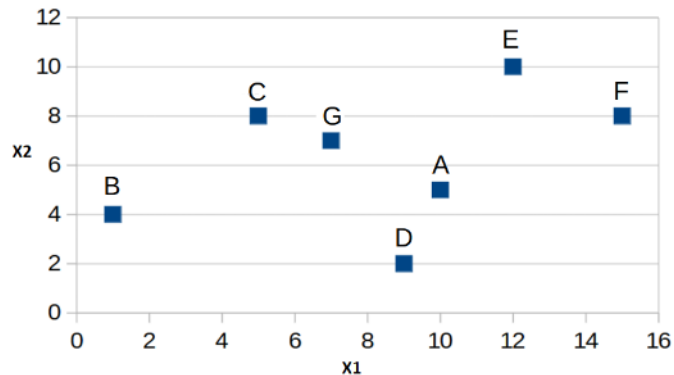


Image: Centroid linkage

How Agglomerative clustering work?

- Clustering the following 7 data points

	X1	X2
A	10	5
B	1	4
C	5	8
D	9	2
E	12	10
F	15	8
G	7	7



[Reference](#)

How Agglomerative clustering work?

- **Step 1:** Calculate distances between all data points using Euclidean distance function. The shortest distance is between data points C and G.

	A	B	C	D	E	F
B	9.06					
C	5.83	5.66				
D	3.16	8.25	7.21			
E	5.39	12.53	7.28	14.42		
F	5.83	14.56	10.00	16.16	3.61	
G	3.61	6.71	2.24	8.60	5.83	8.06



[Reference](#)

How Agglomerative clustering work?

- **Step 2:** We use "Average Linkage" to measure the distance between the "C,G" cluster and other data points.

	A	B	C,G	D	E
B	9.06				
C,G	4.72	6.10			
D	3.16	8.25	6.26		
E	5.39	12.53	6.50	14.42	
F	5.83	14.56	9.01	16.16	3.61

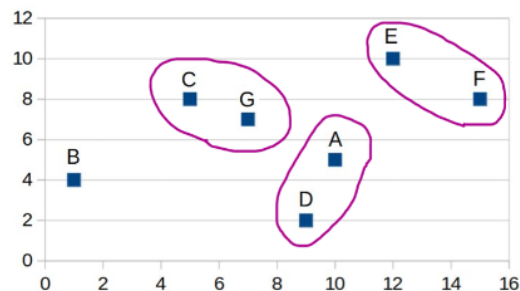


[Reference](#)

How Agglomerative clustering work?

- Step 3:

	A,D	B	C,G	E
B	8.51			
C,G	5.32	6.10		
E	6.96	12.53	6.50	
F	7.11	14.56	9.01	3.61

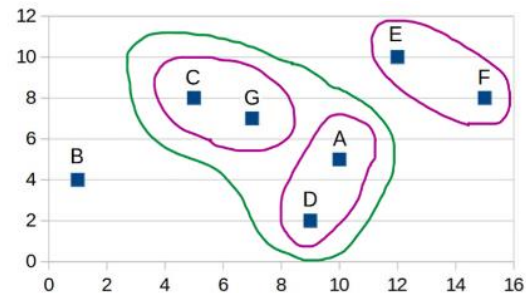


[Reference](#)

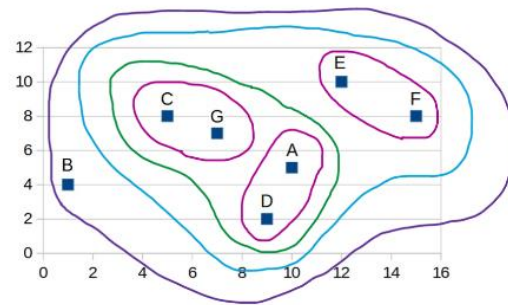
How Agglomerative clustering work?

- Step 4 and 5:

	A,D	B	C,G
B	8.51		
C,G	5.32	6.10	
E,F	6.80	13.46	7.65

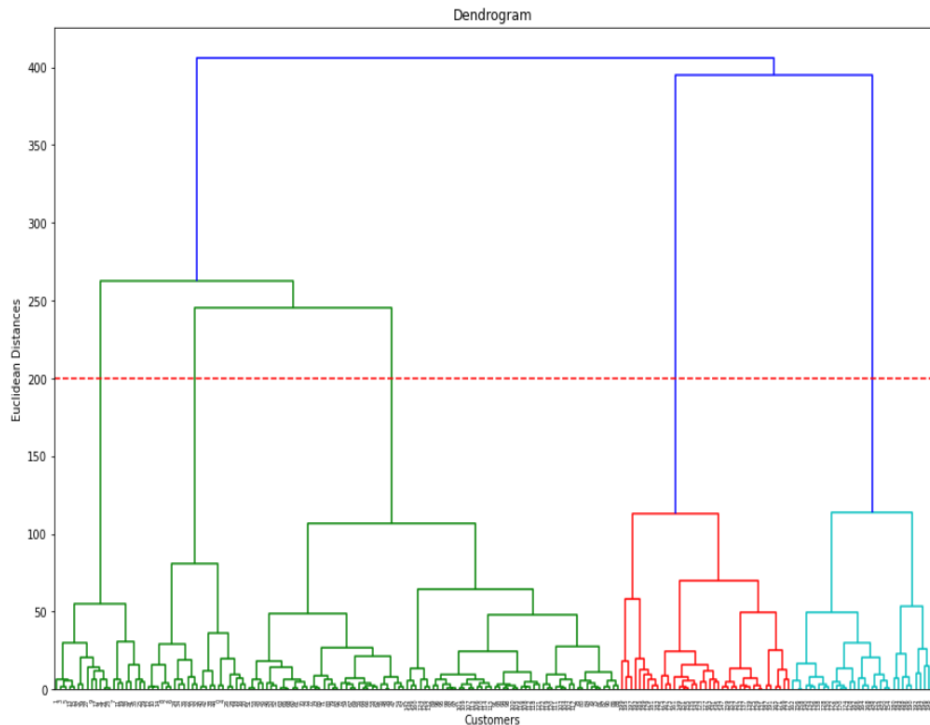


	A,D,C,G,E,F
B	9.07



Dendrogram

- A dendrogram is a tree-like diagram that records the sequences of merges or splits. More the distance of the vertical lines in the dendrogram, more the distance between those clusters.
- We can set a threshold distance and draw a horizontal line (Generally, we try to set the threshold in such a way that it cuts the tallest vertical line).

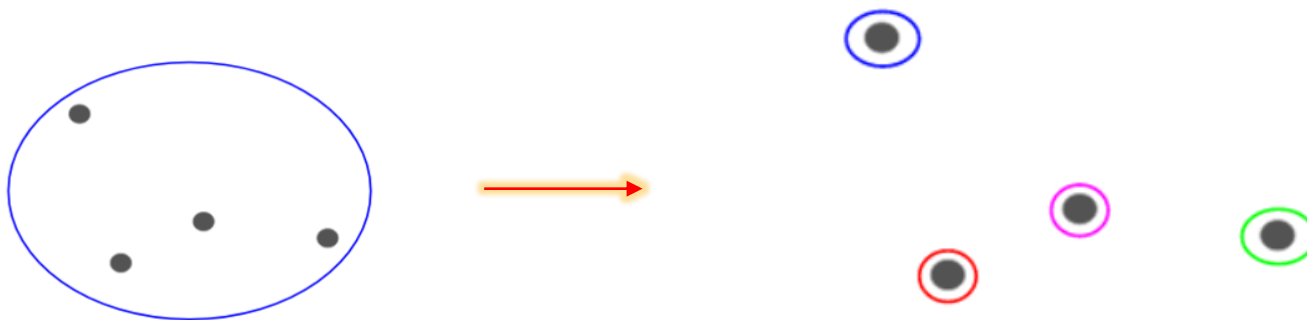


Divisive Hierarchical Clustering

- The Divisive Hierarchical clustering is the opposite of the Agglomerative hierarchical clustering.
- Initially, all objects are considered in a single cluster.
- Then the division process is performed step by step until each object forms a different cluster.
- The cluster division or splitting procedure is carried out according to some principles that maximum distance between neighbouring objects in the cluster.
- It follows a top-down approach and is more efficient than agglomerative clustering. But, due to its complexity in implementation, it doesn't have any predefined implementation in any of the major machine learning frameworks.

Divisive Hierarchical Clustering

- For Example: Consider 4 points belong to same cluster at the beginning.
- At each iteration, we split the farthest point in the cluster and repeat this process until each cluster only contains a single point:



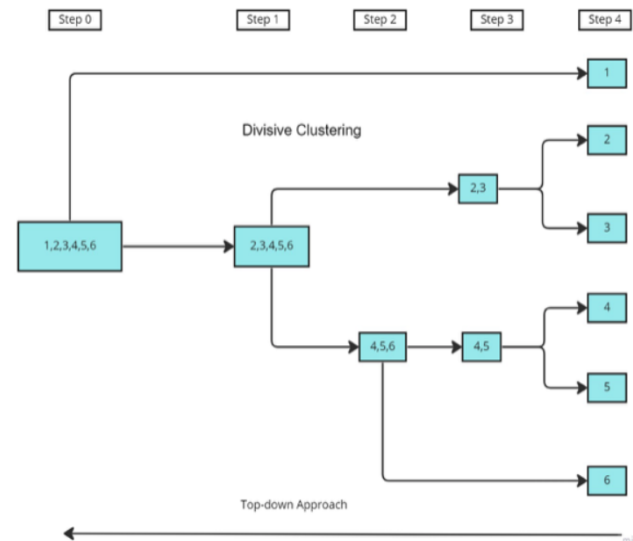
[Reference](#)

Steps in Divisive Hierarchical Clustering

Consider all the data points as a single cluster.

1. Split into clusters using any flat-clustering method, say K-Means.
2. Choose the best cluster among the clusters to split further, choose the one that has the largest Sum of Squared Error (SSE).
3. Repeat steps 2 until a single cluster is formed.

The proximity matrix and linkage function follow the same procedure as agglomerative clustering,



[Image: Divisive Clustering](#)

Application of Hierarchical Clustering

- Identifying fake news
- Identifying criminal activity
- Document Analysis:
- Phylogenetic trees analysis
- Tracking viruses through phylogenetic tree

Lab 1 Implementation of Hierarchical Clustering Using Mall Customer Data

Summary

- Certainly! In this discussion, we've explored hierarchical clustering, a powerful technique in unsupervised machine learning. It's a method to group similar data points into clusters, revealing inherent patterns without predefined labels.
- There are two main types of hierarchical clustering: agglomerative and divisive. Agglomerative clustering starts by treating each data point as a single cluster and progressively merges them into larger clusters. Divisive clustering, on the other hand, begins with all data points in one cluster and divides them into smaller clusters step by step.
- Linkage methods determine how cluster distances are calculated. Common methods include complete linkage (longest distance between points), single linkage (shortest distance), average linkage (average of pairwise distances), and centroid linkage (distance between cluster centroids).
- Divisive hierarchical clustering starts with all data points in one cluster and splits them into smaller groups, guided by the maximum distance between points in a cluster.

Quiz

Question 1: Hierarchical clustering is primarily utilized in:

- a) Data labeling
- b) Unsupervised learning
- c) Reinforcement learning
- d) Dimensionality reduction

Quiz

Question 1: Hierarchical clustering is primarily utilized in:

- a) Data labeling
- b) Unsupervised learning
- c) Reinforcement learning
- d) Dimensionality reduction

Answer: b) Unsupervised learning

Quiz

Question 2: Which type of hierarchical clustering starts with individual data points as clusters and progressively merges them?

- a) Agglomerative clustering
- b) Divisive clustering
- c) K-Means clustering
- d) Spectral clustering

Quiz

Question 2: Which type of hierarchical clustering starts with individual data points as clusters and progressively merges them?

- a) Agglomerative clustering
- b) Divisive clustering
- c) K-Means clustering
- d) Spectral clustering

Answer: a) Agglomerative clustering

Quiz

Question 3: What does a dendrogram represent in hierarchical clustering?

- a) A linear regression model
- b) A scatter plot of data points
- c) A tree-like diagram of merge/split sequences
- d) A matrix of distance values

Quiz

Question 3: What does a dendrogram represent in hierarchical clustering?

- a) A linear regression model
- b) A scatter plot of data points
- c) A tree-like diagram of merge/split sequences
- d) A matrix of distance values

Answer: c) A tree-like diagram of merge/split sequences

Quiz

Question 4: The linkage method that considers the average of pairwise distances between cluster elements is:

- a) Single linkage
- b) Complete linkage
- c) Average linkage
- d) Centroid linkage

Quiz

Question 4: The linkage method that considers the average of pairwise distances between cluster elements is:

- a) Single linkage
- b) Complete linkage
- c) Average linkage
- d) Centroid linkage

Answer: c) Average linkage

Quiz

Question 5: Divisive hierarchical clustering follows a:

- a) Bottom-up approach
- b) Randomized approach
- c) Top-down approach
- d) Hierarchical approach

Quiz

Question 5: Divisive hierarchical clustering follows a:

- a) Bottom-up approach
- b) Randomized approach
- c) Top-down approach
- d) Hierarchical approach

Answer: c) Top-down approach

References

- <https://www.geeksforgeeks.org/hierarchical-clustering-in-data-mining/>
- https://en.wikipedia.org/wiki/Hierarchical_clustering
- <https://www.javatpoint.com/hierarchical-clustering-in-machine-learning>

Thank you...!