Data Science and Artificial Intelligence

Machine Learning

Unsupervised learning

Lecture No.4

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

Topic Kmedoid

Topic Aglomerative clustering

Topic Dendogram

Topic Linkages

Topic

Turn on Slide map

Topics to be Covered

Topic Example

Topic Divisive clustering

Topic Kmean & DC

Topic DC using MST.

Topic

WHEN YOU FOCUS
ON THE GOOD
THE GOOD GETS
BETTER

K-Means Clustering

Single link.	min (Interpoint dist)
Complete link.	max ( )
Avg link	avg ( )
Centroid link	distance of centroid.

Hierarchical Clustering : Bottom Up Approach

- We have Leaf Nodes : Clusters with Single point
- We have Internal Nodes : Clusters
- Tree ends at the root node @Root node we have all points.

Hierarchical Clustering : Bottom Up Approach

- (no need to specify K)
- (we find Best 'K' using dendogram)

So How we initialise and start this method ...

We don't need K ...

Clustering

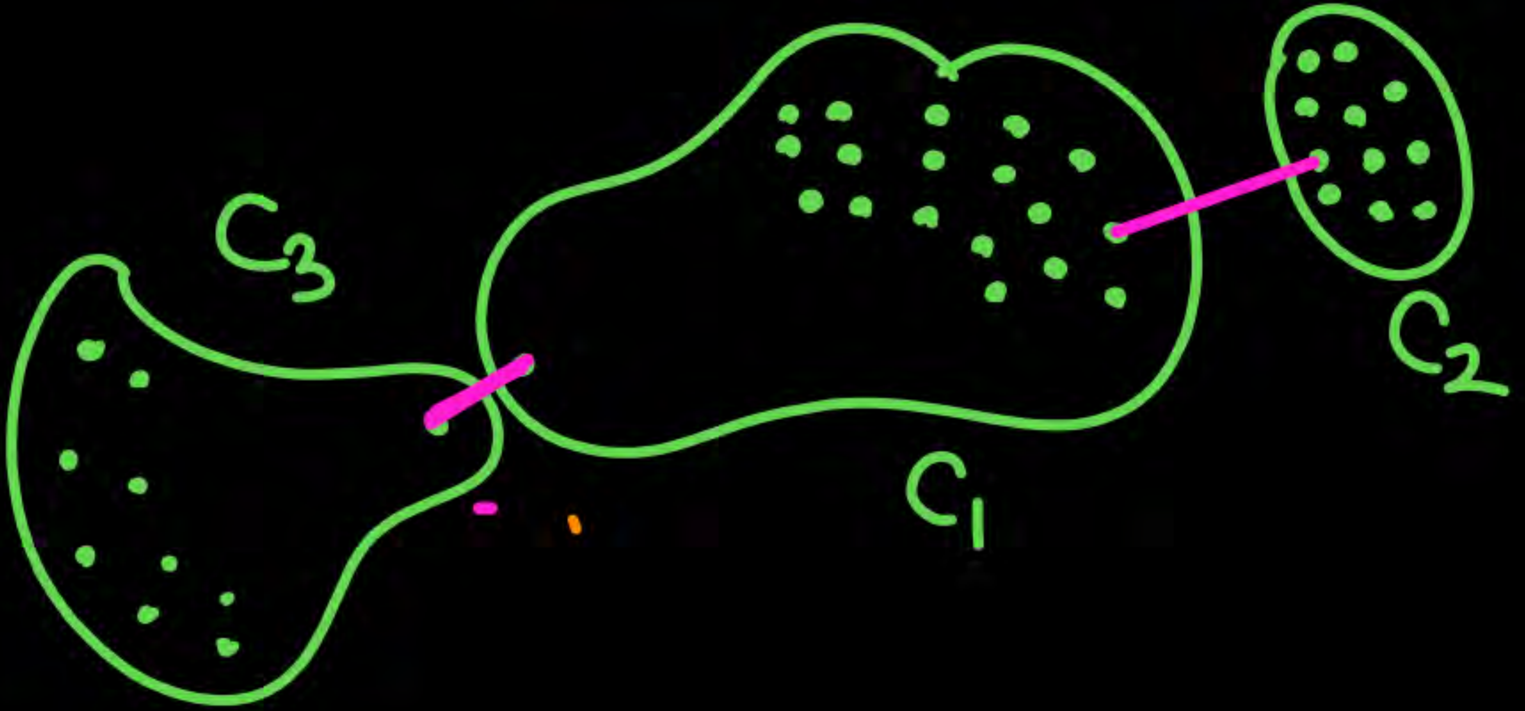
How we find the group linkage/ distance between groups ?

Single Linkage

(we have done)

$d_{c_1c_2} > d_{c_1c_3}$

- Chaining effect
- they try to create elongated clusters.



Clustering

How we find the group linkage/ distance between groups ?

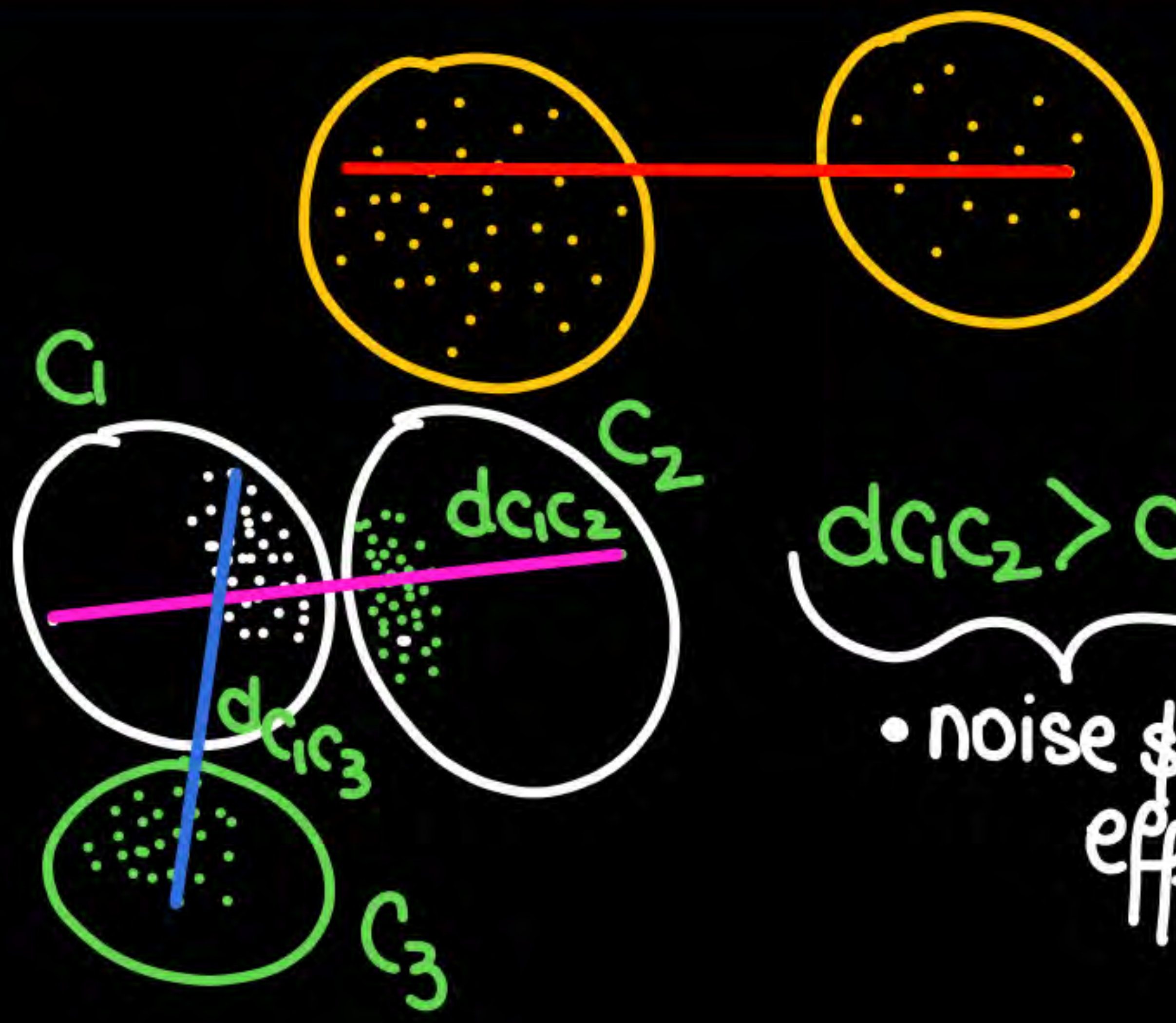
Problem in Single Linkage : Chaining...

- **Characteristics**
- **Tends to produce long, "loose" clusters** that may be less compact. ← less Compact.
- **Sensitive to noise** and outliers.
- Can create chaining effects, where clusters are elongated.
- **Chain Effect**: ~~Complete~~ Single linkage can suffer from the chaining phenomenon, where clusters that are close together are merged, even if they should not be, resulting in elongated and less meaningful clusters.

How we find the group linkage/ distance between groups ?

Complete Linkage

(• It tends to create Small Clusters)



$\underbrace{d_{c_1c_2} > d_{c_1c_3}}$

• noise & outlier can effect ✓

Complete linkage

- While complete linkage has its advantages, such as producing compact clusters, it also has several potential issues:
- **Sensitivity to Outliers**: Since complete linkage uses the maximum distance between points, it is highly sensitive to outliers. A single outlier can significantly affect the distance calculation and, consequently, the clustering results.
- Cluster Shape: Complete linkage tends to produce clusters of roughly equal size and shape, which may not be appropriate for all datasets. If the data has clusters of varying shapes and sizes, complete linkage might not capture the true structure of the data. ⇒ bcoz of limitation of keeping small clusters.
- **Computational Complexity**: Hierarchical clustering, in general, has high computational complexity. For large datasets, the distance calculations in complete linkage can be particularly time-consuming.
- **Scalability**: As the dataset grows, the memory and computational requirements increase significantly, making complete linkage less suitable for large datasets.

- Single linkage can result in long stringy clusters and "chaining" while complete linkage tends to make highly compact clusters.

Single
Complete
Avg

} Linkage v-high Computation bcoz we have to find distance b/w all points of one cluster with points of other cluster

Clustering

How we find the group linkage/ distance between groups ?

Average Linkage

Solve the problem

- Solve problem of both single & complete linkage
- distance b/w cluster ⇒ avg(·)

How we find the group linkage/ distance between groups ?

Centroid Linkage

- → less Computation
- → noise/ outlier

Point	Coordinates (x, y)
A	(1,1)
B	(2, 2)
C	(5, 5)
D	(6, 6)
E	(8, 8)

	A	B	C	D	E
A	0	$\sqrt{2}$	$\sqrt{32}$	$\sqrt{50}$	$\sqrt{98}$
B	$\sqrt{2}$	0	$\sqrt{18}$	$\sqrt{32}$	$\sqrt{72}$
C	$\sqrt{32}$	$\sqrt{18}$	0	$\sqrt{2}$	$\sqrt{18}$
D	$\sqrt{50}$	$\sqrt{32}$	$\sqrt{2}$	0	$\sqrt{8}$
E	$\sqrt{98}$	$\sqrt{72}$	$\sqrt{18}$	$\sqrt{8}$	0

	AB	C	D	E
AB	0	$\sqrt{32}$	$\sqrt{50}$	$\sqrt{98}$
C	$\sqrt{32}$	0	$\sqrt{2}$	$\sqrt{18}$
D	$\sqrt{50}$	$\sqrt{2}$	0	$\sqrt{8}$
E	$\sqrt{98}$	$\sqrt{18}$	$\sqrt{8}$	0

Complete linkage

	AB	CD	E
AB	0	$\sqrt{50}$	$\sqrt{98}$
CD	$\sqrt{50}$	0	$\sqrt{18}$
E	$\sqrt{98}$	$\sqrt{18}$	0

(AB)(CDE)

Point	Coordinates (x, y)
A	(1, 1)
B	(2, 2)
C	(5, 5)
D	(6, 6)
E	(8, 8)

	A	B	C	D	E
A	0	$\sqrt{2}$	$\sqrt{32}$	$\sqrt{50}$	$\sqrt{98}$
B	$\sqrt{2}$	0	$\sqrt{18}$	$\sqrt{32}$	$\sqrt{72}$
C	$\sqrt{32}$	$\sqrt{18}$	0	$\sqrt{2}$	$\sqrt{18}$
D	$\sqrt{50}$	$\sqrt{32}$	$\sqrt{2}$	0	$\sqrt{8}$
E	$\sqrt{98}$	$\sqrt{72}$	$\sqrt{18}$	$\sqrt{8}$	0

	AB	C	D	E
AB	0	$\sqrt{18}$	$\sqrt{32}$	$\sqrt{72}$
C	$\sqrt{18}$	0	$\sqrt{2}$	$\sqrt{18}$
D	$\sqrt{32}$	$\sqrt{2}$	0	$\sqrt{8}$
E	$\sqrt{72}$	$\sqrt{18}$	$\sqrt{8}$	0

Single linkage

	AB	CD	E
AB	0	$\sqrt{18}$	$\sqrt{72}$
CD	$\sqrt{18}$	0	$\sqrt{8}$
E	$\sqrt{72}$	$\sqrt{8}$	0

AB

$\underbrace{\text{CDE}}$.

Point	Coordinates (x, y)
A	(1,1)
B	(2, 2)
C	(5, 5)
D	(6, 6)
E	(8, 8)

	A	B	C	D	E
A	0	$\sqrt{2}$	$\sqrt{32}$	$\sqrt{50}$	$\sqrt{98}$
B	$\sqrt{2}$	0	$\sqrt{18}$	$\sqrt{32}$	$\sqrt{72}$
C	$\sqrt{32}$	$\sqrt{18}$	0	$\sqrt{2}$	$\sqrt{18}$
D	$\sqrt{50}$	$\sqrt{32}$	$\sqrt{2}$	0	$\sqrt{8}$
E	$\sqrt{98}$	$\sqrt{72}$	$\sqrt{18}$	$\sqrt{8}$	0

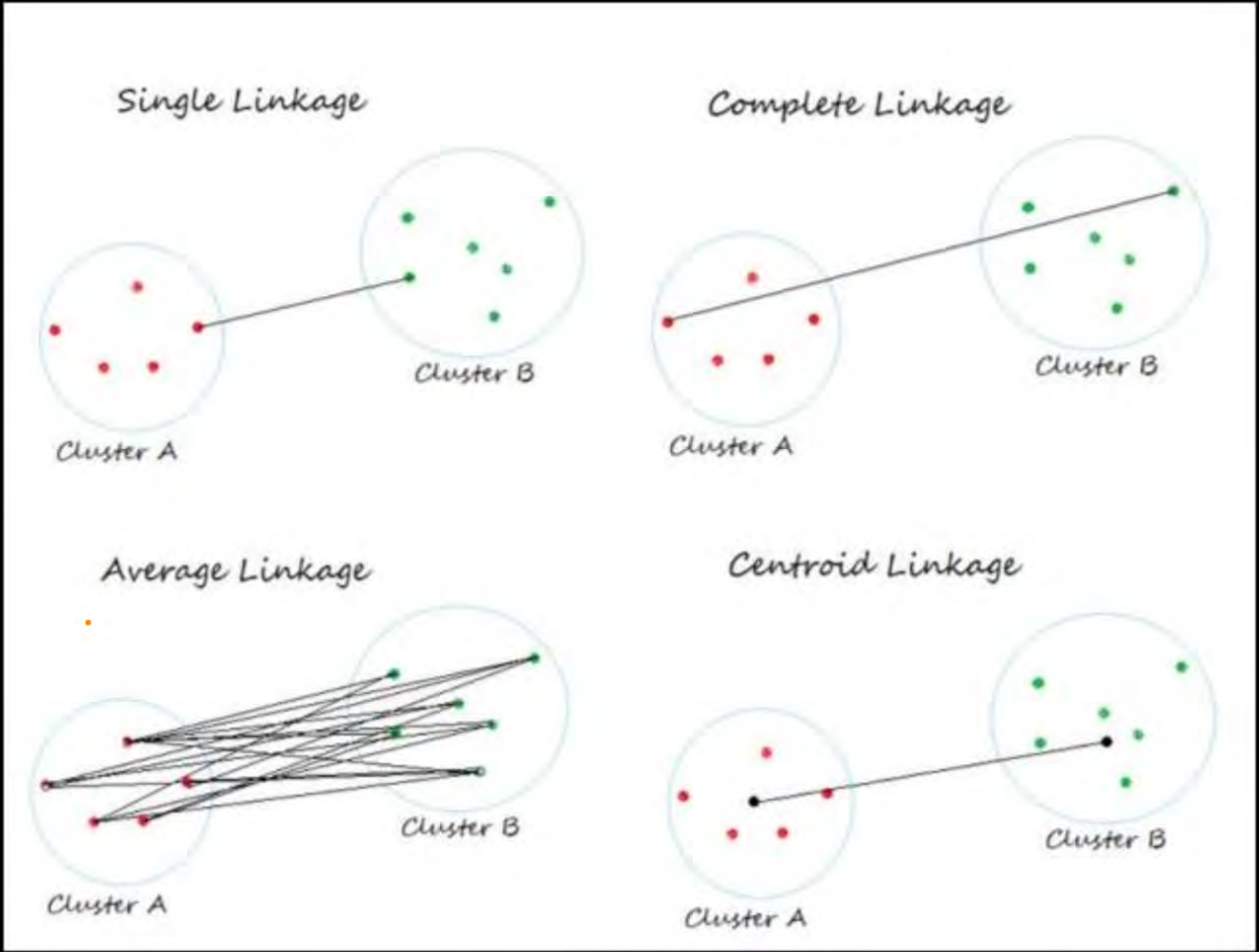
	A	B	CD	E
A	0	$\sqrt{2}$	$\sqrt{32}$	$\sqrt{98}$
B	$\sqrt{2}$	0	$\sqrt{18}$	$\sqrt{72}$
CD	$\sqrt{32}$	$\sqrt{18}$	0	$\sqrt{8}$
E	$\sqrt{98}$	$\sqrt{72}$	$\sqrt{8}$	0

Single linkage

	AB	CD	E
AB	0	$\sqrt{18}$	$\sqrt{72}$
CD	$\sqrt{18}$	0	$\sqrt{8}$
E	$\sqrt{72}$	$\sqrt{8}$	0

AB

$\underbrace{\text{CDE}}$.


Single Linkage
Cluster A
Cluster B
Complete Linkage
Cluster A
Cluster B
Average Linkage
Cluster A
Cluster B
Centroid Linkage
Cluster A
Cluster B

Flat Clustering ⇒ Non deterministic → Kmeans or Kmedoid → final result are never same

advantage

→ deterministic

→ No need to specific 'K' we find best K from dendogram.

disadvantage

* Computationally extensive

* final Result depend on type of linkage.

- **Advantages of Agglomerative Clustering**
- **Versatility**: Can be used with various types of distance metrics and linkage criteria, making it adaptable to different types of data and clustering goals.
- **Hierarchy**: Produces a hierarchy of clusters, allowing the examination of data at different levels of granularity.
- **Intuitive Visualization**: The dendrogram provides a clear and interpretable visualization of the clustering process.
- **Disadvantages of Agglomerative Clustering**
- **Computational Complexity**: The algorithm can be computationally intensive, especially for large datasets, as it requires calculating and updating a distance matrix.
- **Sensitivity to Noise and Outliers**: Can be affected by noise and outliers, which may lead to less meaningful clusters.
- **Choice of Linkage and Distance Metric**: The results can vary significantly depending on the chosen linkage criteria and distance metric, which may require experimentation and domain knowledge to select appropriately.

Linkage Method	Description	Advantages	Disadvantages	Best Used For
Single Linkage	Minimum distance between points in the clusters	Tends to find long, chain-like clusters	Sensitive to noise and outliers, can produce chaining effect	Clusters with elongated shapes
Complete Linkage	Maximum distance between points in the clusters	Produces compact, spherical clusters	Sensitive to outliers, can create tightly packed clusters regardless of actual data structure	Clusters of similar size and shape, when compact clusters are desired
Average Linkage	Average distance between all points in the clusters	Balances between single and complete linkage	May not perform well if clusters are of different sizes or densities noise & outlier	Clusters with moderate structure, balance between compactness and separation
Centroid Linkage	Distance between centroids of the clusters	Takes into account the overall geometry of the cluster	Can produce clusters with centroids that are not part of the original data	Clusters where centroids are meaningful

1. **What is a dendrogram primarily used for in hierarchical clustering?**

- A) To determine the optimal number of clusters

Pimary

- B) To visualize the hierarchical relationships between clusters
- C) To calculate the distance between data points
- D) To reduce the dimensionality of the data

2. **In a dendrogram, what does the height of the linkage indicate?**

- A) The distance between individual data points
- B) The similarity between clusters
- C) The dissimilarity between clusters
- D) The size of each cluster

Which of the following linkage criteria can be used in hierarchical clustering to create a dendrogram?

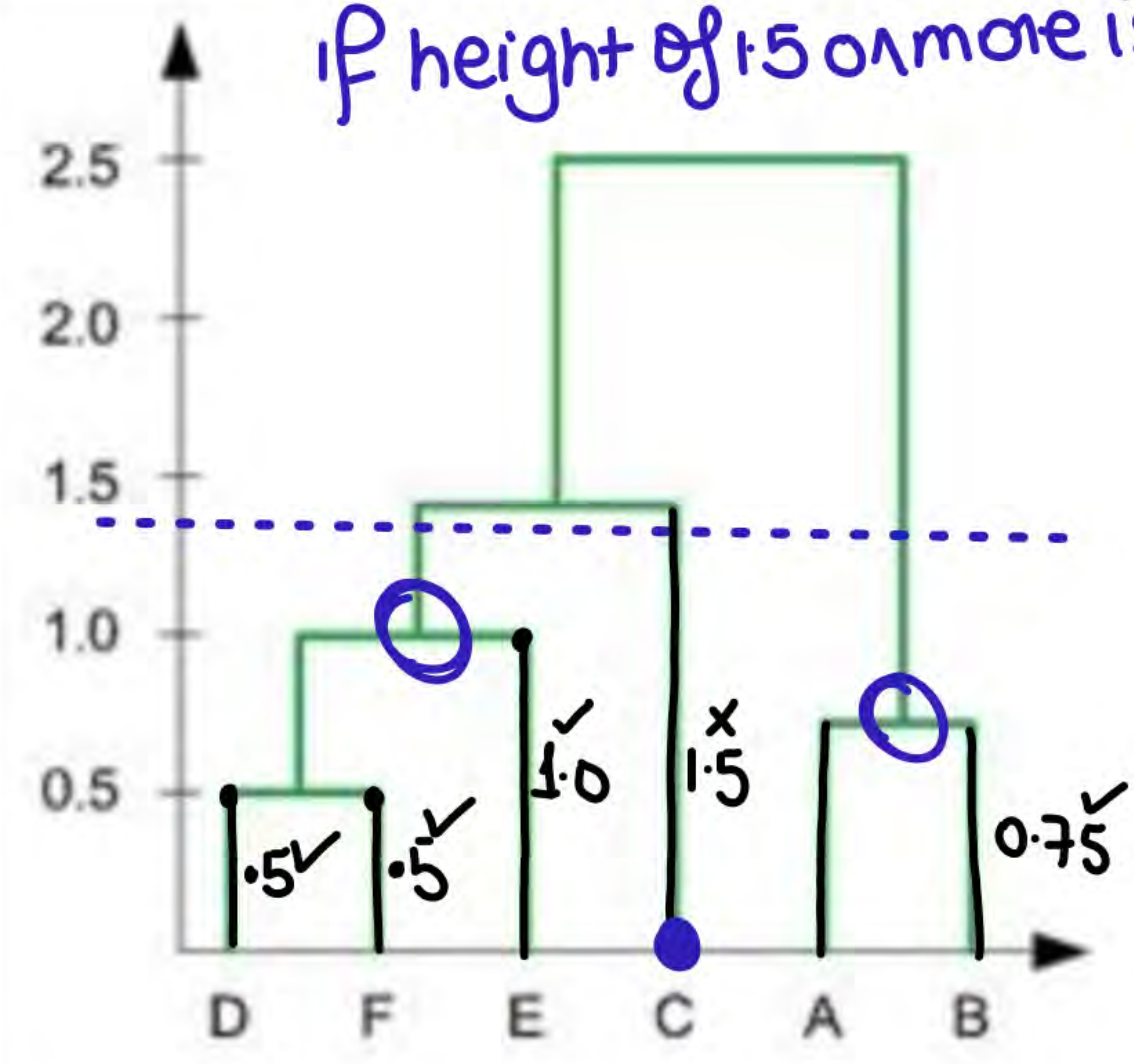
- A) Single linkage
- B) Complete linkage
- C) Average linkage
- D) All of the above

threshold ⇒ hyperParameter.

What is the purpose of cutting a dendrogram at a certain height?

- A) To determine the number of clusters
- B) To determine the size of each cluster
- C) To identify the most similar data points
- D) To visualize the data in two dimensions

If height of 1.5 or more is not allowed to form cluster.



3 Clusters

DFE
AB
C

Clustering

How to find the best k

avg linkage compared to K-medoid → more Robust.

The distance in the dendogram show the dissimilarity between the clusters ...

⇒ Computational Complexity of Aglomerative ⇒ • naive aglo- -merative ⇓ $O(N^3)$

→ Efficient → $O(N^2 + N.N\log N)$

N times Sorting

Step1 · N Clusters → N^2.

Step2 · N-1 " → N^2.

Step3 · N-2 → N^2.

⋮

StepN 1

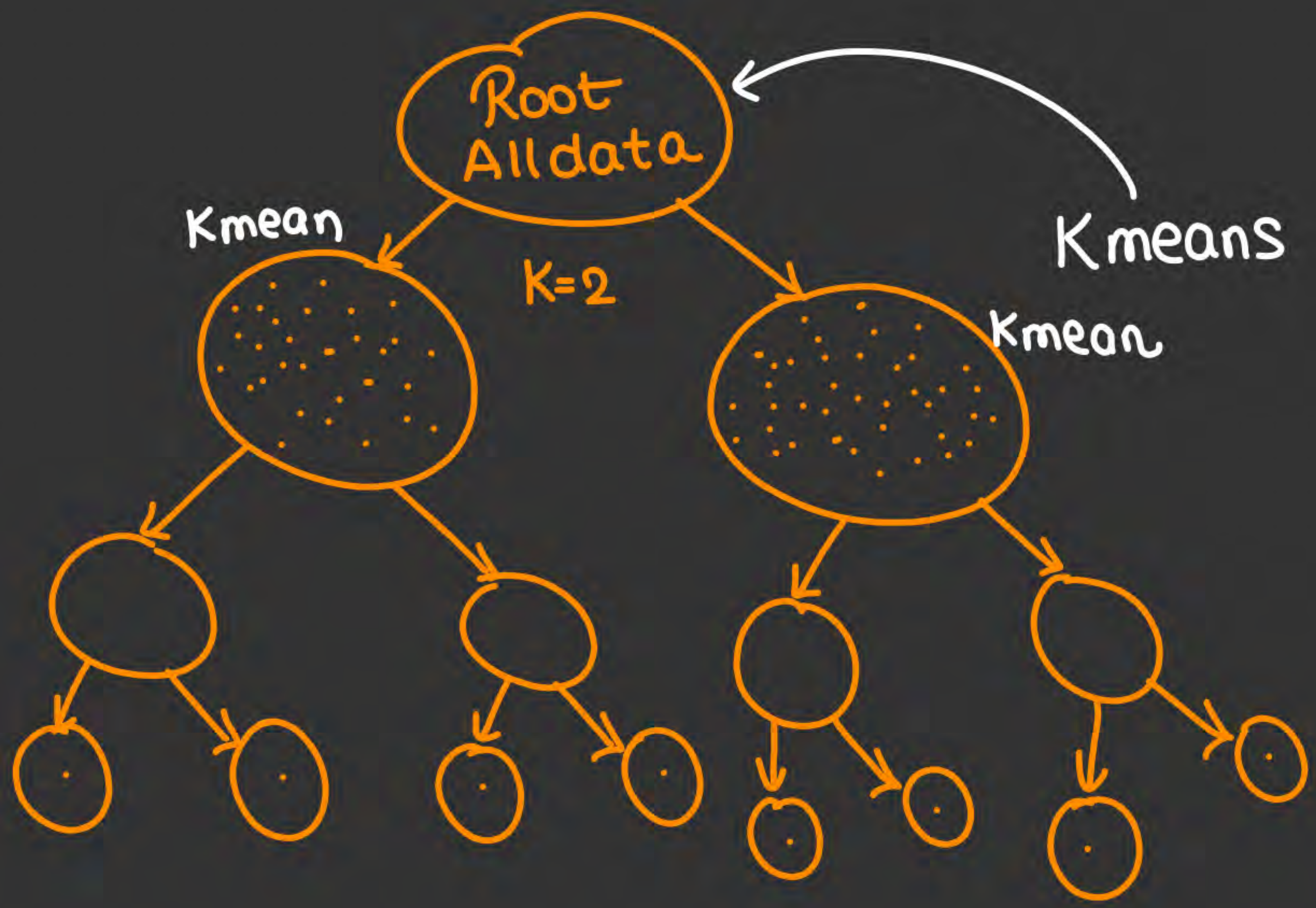
Top-down ⇒.

Divisive Clustering

- In this algo we apply the flat Clustering in iterative manner.

The main idea behind this is ....

It is simply the iterative application of flat clustering


Root
Alldata
Kmeans
Kmean
K=2
Kmean

Divisive Clustering

Divisive clustering, also known as "top-down" clustering, is a type of hierarchical clustering that starts with all data points in a single cluster and iteratively splits them into smaller clusters until each data point is its own cluster or until a stopping criterion is met. This approach is the opposite of agglomerative clustering, which starts with each data point as its own cluster and then merges them.

Divisive Clustering

- **Initial State**: All data points are grouped into a single cluster.
- **Process**: Iteratively splits the clusters into smaller clusters.
- **Stopping Criteria**: The process continues until each data point is in its own cluster or a predefined number of clusters is reached.

If the variance of cluster

$$\Rightarrow \frac{1}{N_i} \sum_{l=1}^{N_i} \| x_i - \mu_i \|^2 < \delta \text{ then do not split.}$$

Divisive Clustering

- **Steps in Divisive Clustering**
- **Start with a Single Cluster:**
  - Begin with all data points in one large cluster.
- **Choose a Cluster to Split:**
  - Select the cluster that needs to be split. This could be based on various criteria such as the largest cluster or the cluster with the highest variance.
- **Split the Cluster:**
  - Use a clustering algorithm (such as K-means) to divide the chosen cluster into two smaller clusters. This is the core step where a decision on how to split the data is made.
- **Repeat:**
  - Continue the process of choosing and splitting clusters until the stopping criterion is met.

Comparison

Feature	Agglomerative Clustering	Divisive Clustering
Approach	Bottom-up	Top-down
Initial State	Each data point is its own cluster	All data points are in a single cluster
Process	Merges the closest pairs of clusters iteratively	Splits the clusters iteratively
Termination Condition	Until all points are merged into one cluster or a specified number of clusters is achieved	Until each point is its own cluster or a specified number of clusters is reached
Complexity	Typically more computationally expensive for large datasets due to repeated merging steps	Can be more efficient for large datasets as it avoids repeated merging
Example Algorithms	Single linkage, complete linkage, average linkage, Ward's method	Recursive application of clustering algorithms like K-means or spectral clustering
Dendrogram	Built from the bottom up, starting with individual points	Built from the top down, starting with all points
Usage	Commonly used due to simplicity and easy implementation	Less commonly used due to complexity in deciding optimal splits

Clustering

Comparison

Flexibility	Generally more flexible and easier to implement with different linkage criteria	Requires an effective strategy for splitting clusters
Sensitivity to Noise	Less sensitive to noise, as noise points are merged into clusters gradually	More sensitive to noise, as initial splits can be affected by outliers
Example Use Cases	Hierarchical document clustering, gene expression data analysis, image segmentation	Rarely used, but can be applied in specific scenarios needing top-down clustering

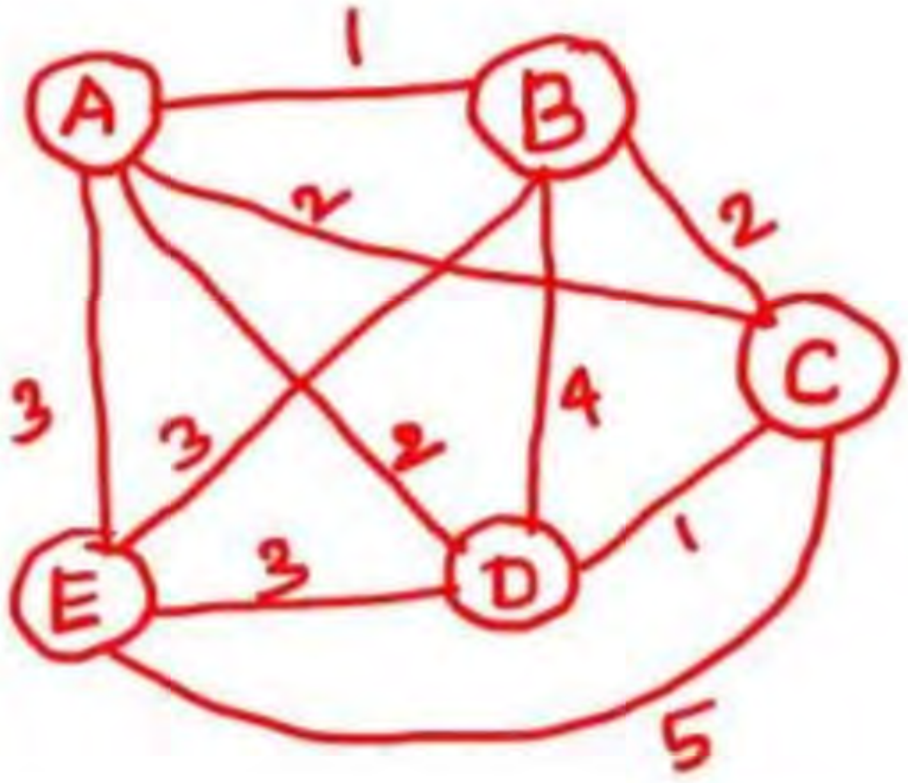
Divisive Clustering using MST

A Minimum Spanning Tree (MST) is a subset of the edges of a connected, undirected graph that connects all the vertices together, without any cycles, and with the minimum possible total edge weight. In other words, it is a tree that spans all the vertices in the graph and has the smallest sum of edge weights among all possible spanning trees.

Divisive Clustering using MST

- MST starts with a tree that consists of a point p.
  - Then check for the closest pair of points (p, q) such that p is in the current tree but q is not in the tree.
  - With this closest pair of points (p, q), add q to the tree and create an edge between p and q.
- Remove the edges from MST graph from largest to smallest repeatedly.
  - All the items are in one cluster {A, B, C, D, E}
  - Largest edge is between D and E. so remove it, and make as 2 clusters- {E}, {A., B, C, D}
  - Next, remove the edge between B and C, which results in {E}, {A, B} {C, D}
  - Finally, remove the edges between A and B (also between C and D), that results {E}, {4}, {B}, {C} and {D}

Divisive Clustering using MST



	A	B	C	D	E
A	0	1	2	2	3
B	1	0	2	4	3
C	2	2	0	1	5
D	2	4	1	0	3
E	3	3	5	3	0

Hierarchical Vs Flat

Aspect	Hierarchical Clustering	Flat Clustering
Methodology	Builds a hierarchy of clusters	Partitions data into a set number of clusters
Types	Agglomerative (bottom-up), Divisive (top-down)	K-means, K-medoids, etc.
Cluster Number	Does not require a predefined number of clusters	Requires a predefined number of clusters
Complexity	Typically more computationally intensive	Generally less computationally intensive
Flexibility	Can provide more flexible clustering Versatile	Less flexible due to predefined number of clusters → no flexibility
Visualization	Produces a dendrogram to visualize the clustering process	No hierarchical structure visualization; can use scatter plots
Merge/Split Criteria	Uses linkage criteria for merging/splitting clusters	Uses centroid or medoid to define cluster centers

Clustering

Hierarchical Vs Flat

	Hierarchical	Flat
Optimal Number of Clusters	Can determine optimal number of clusters using dendrogram.	Requires methods like Elbow or Silhouette for optimal number.
Data Size Suitability	Suitable for smaller datasets due to computational demands	Suitable for larger datasets.
Handling Noise and Outliers	Can be sensitive to noise and outliers → linkage	Can be more robust depending on the algorithm used
Result Interpretability	→ Easier to interpret hierarchical relationships	Interpretation depends on cluster centroids/medoids
Examples	Agglomerative Clustering, Divisive Clustering	K-means, K-medoids, DBSCAN
Initialization Dependence	Not dependent on initial cluster centers deterministic	Can be sensitive to initial cluster center selection Nondeterministic.

THANK - YOU