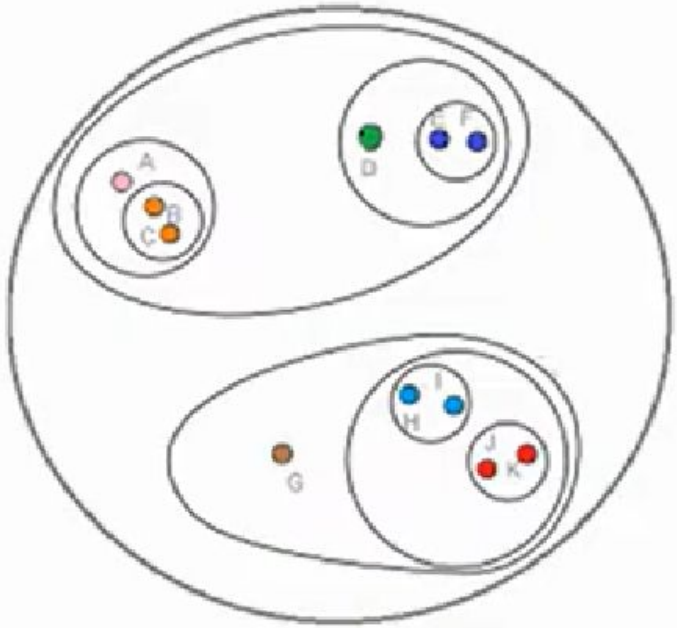
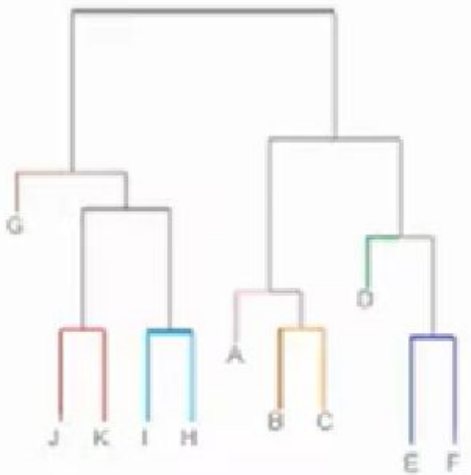
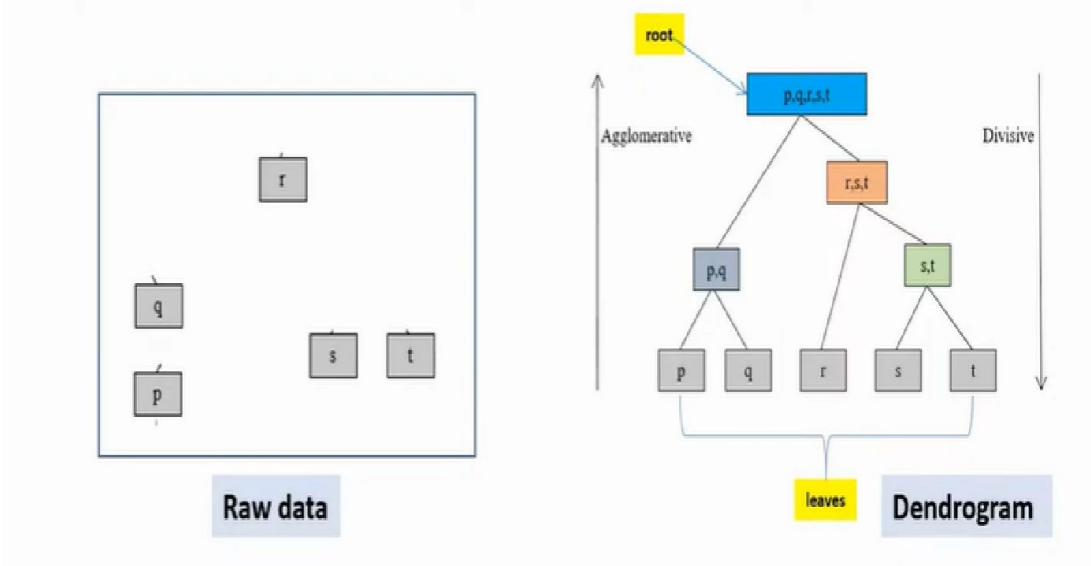
Hierarchical Clustering

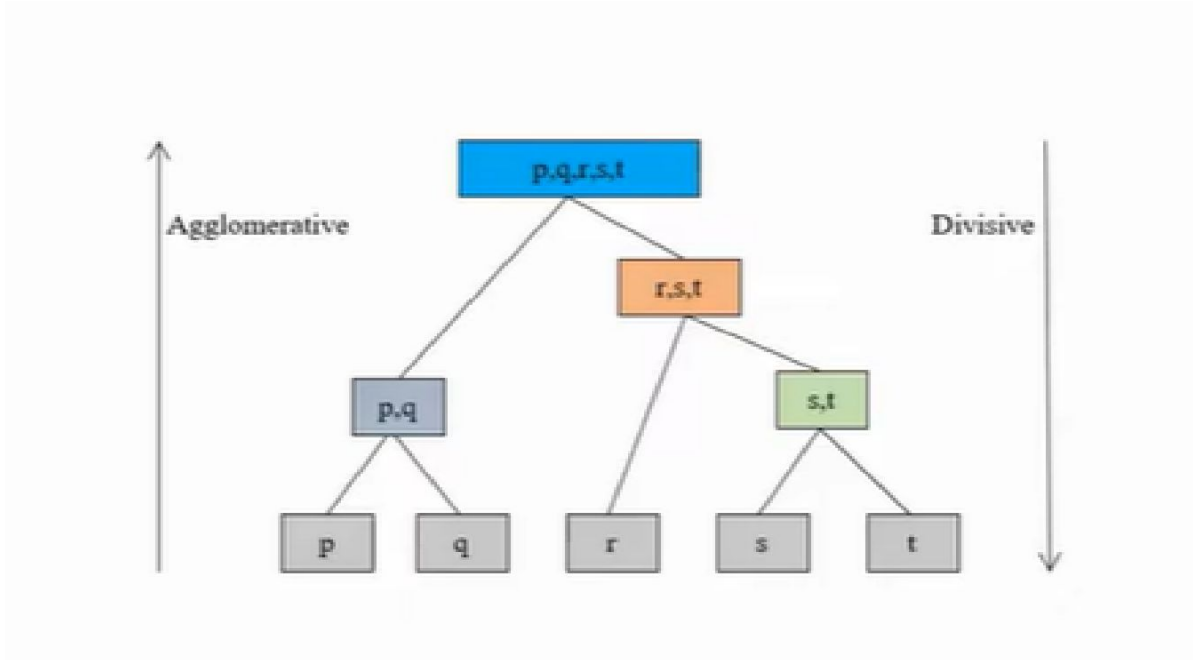
- Hierarchical clustering constructs **nested clusters** by successive merging or splitting of data.
- The hierarchy of clusters is represented as a tree.
- The tree is usually called a ***dendrogram.***







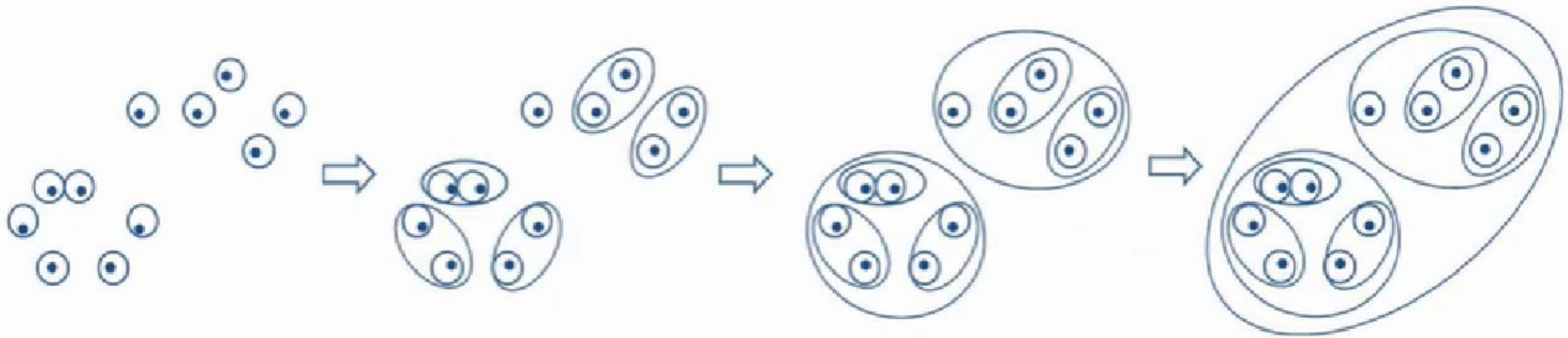
- The dendrogram tool explains, how clusters are formed and visualizes clusters at different scales.
- The tree that results from the technique shows the **similarity between the data points**.
- The root of the dendrogram is the single cluster that contains all the data points.
- The leaves are the clusters containing only one data point each.
- Cluster partitioning is computed by **selecting a cut** on the tree at a certain level.

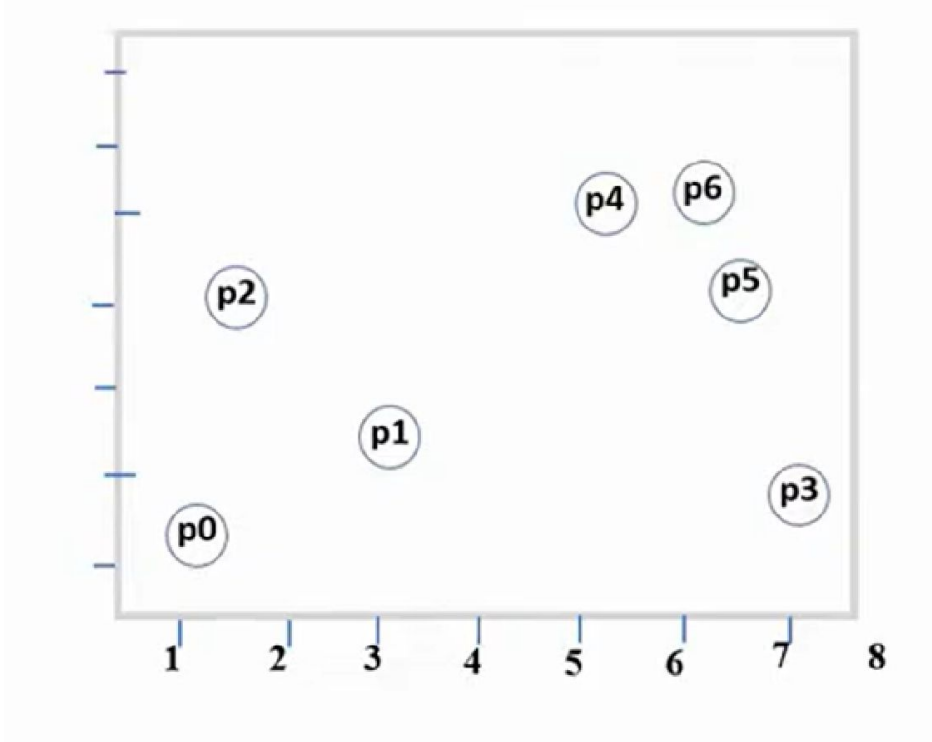


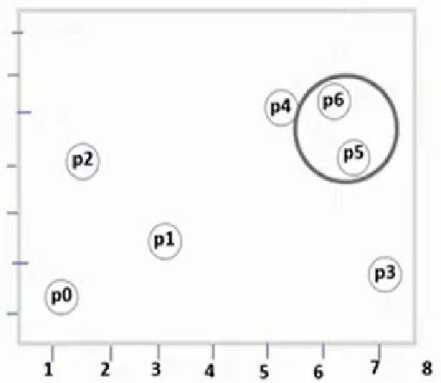
- Hierarchical clustering algorithms are performed in two approaches such as **top-down or bottom-up.**

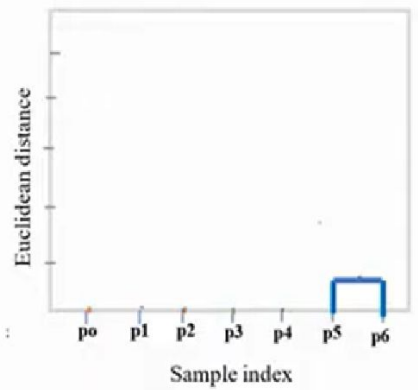
Bottom up Approach

- In **bottom-up algorithms**, each data point is treated as a **single cluster** and then iteratively merges (or agglomerate) with **pairs of clusters** until all clusters have been **merged into a single cluster** that holds all data points.
  - The bottom-up hierarchical clustering is also called as **Hierarchical Agglomerative Clustering (HAC).**

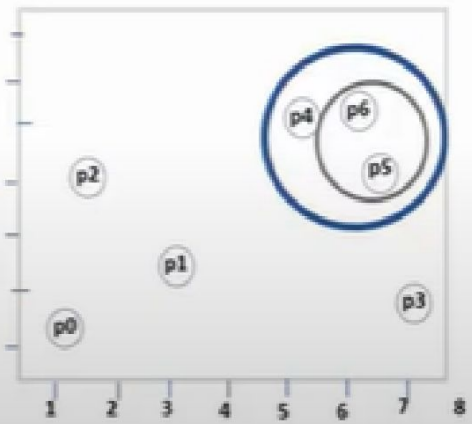


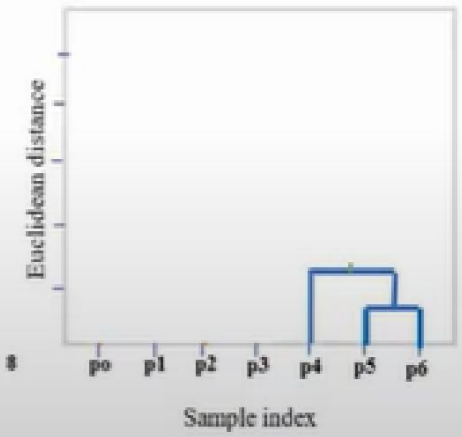

p4
p6
p2
p5
p1
p3
p0
1
2
3
4
5
6
7
8


p4
p6
p2
p5
p1
p3
p0
1
2
3
4
5
6
7
8

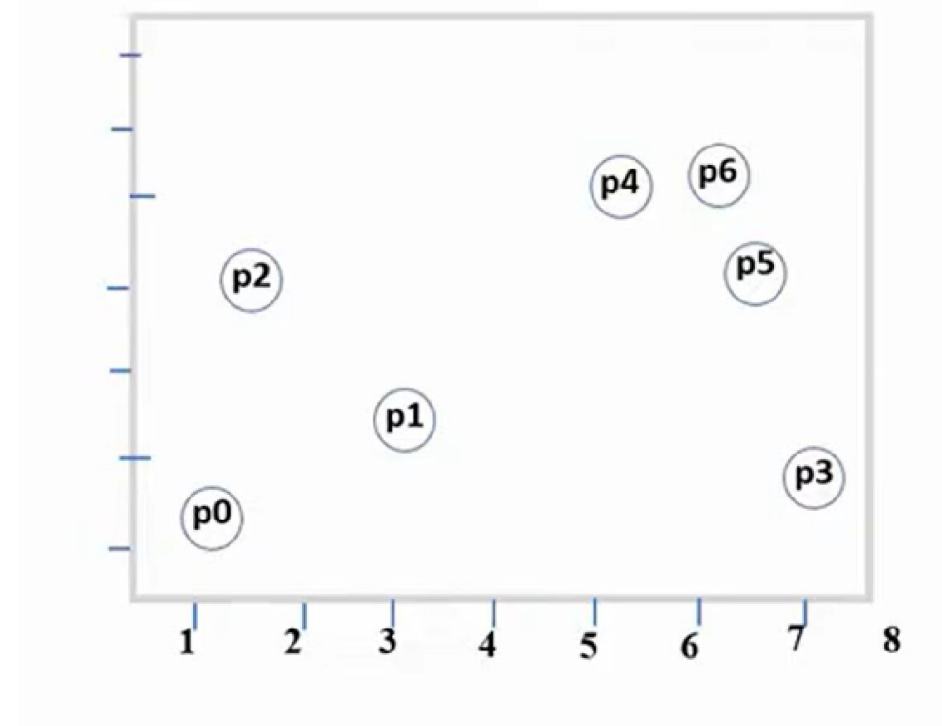

Euclidean distance
p0
p1
p2
p3
p4
p5
p6
Sample index

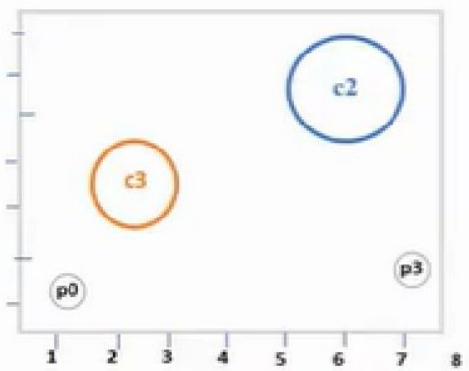
Step 1

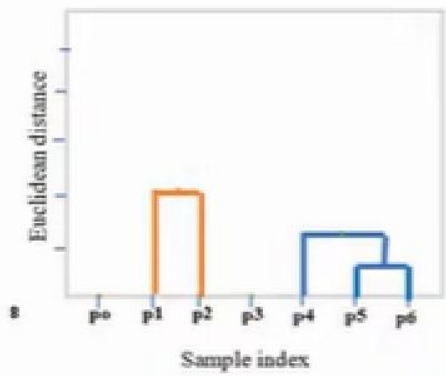

p4
p6
p2
p5
p1
p3
p0
1
2
3
4
5
6
7
8


Euclidean distance
p0
p1
p2
p3
p4
p5
p6
Sample index

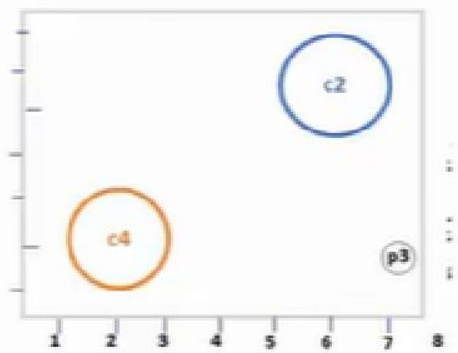
Step 2

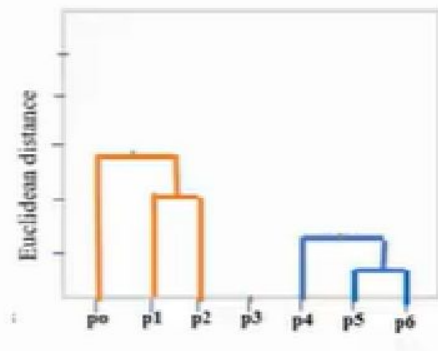

p4
p6
p2
p5
p1
p3
p0
1
2
3
4
5
6
7
8


c2
c3
p0
p3
1
2
3
4
5
6
7
8

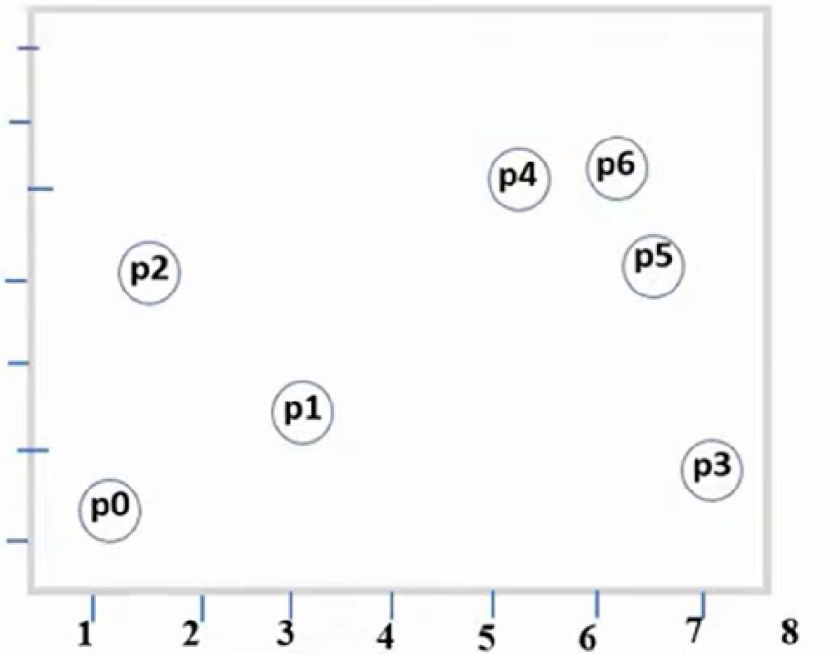

Euclidean distance
p0
p1
p2
p3
p4
p5
p6
Sample index

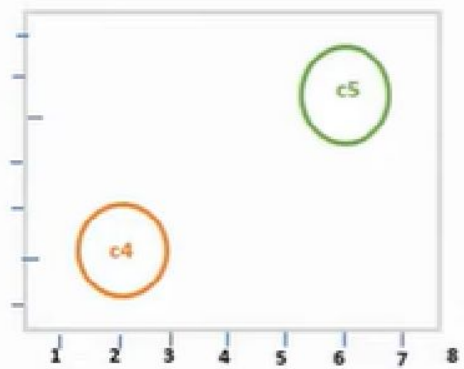
Step 3

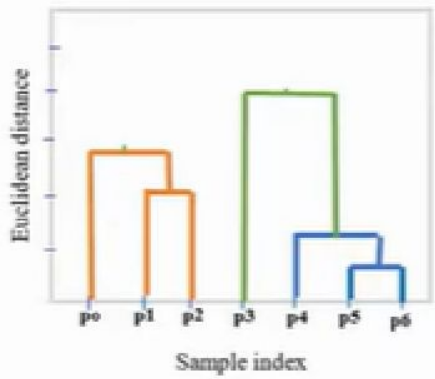

c2
c4
p3
1
2
3
4
5
6
7
8


Euclidean distance
p0
p1
p2
p3
p4
p5
p6

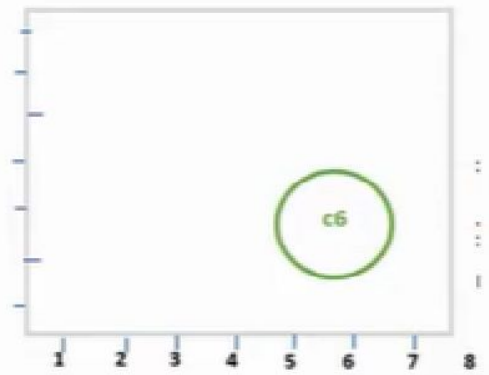
Step 4

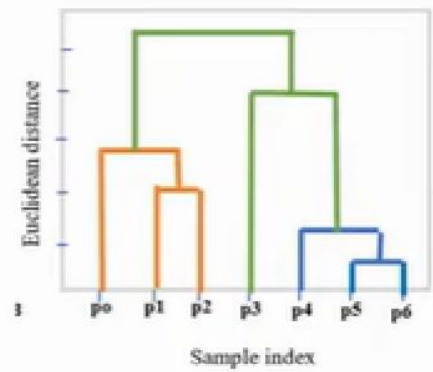

p4
p6
p2
p5
p1
p3
p0
1
2
3
4
5
6
7
8


c5
c4
1
2
3
4
5
6
7
8


Euclidean distance
p0
p1
p2
p3
p4
p5
p6
Sample index

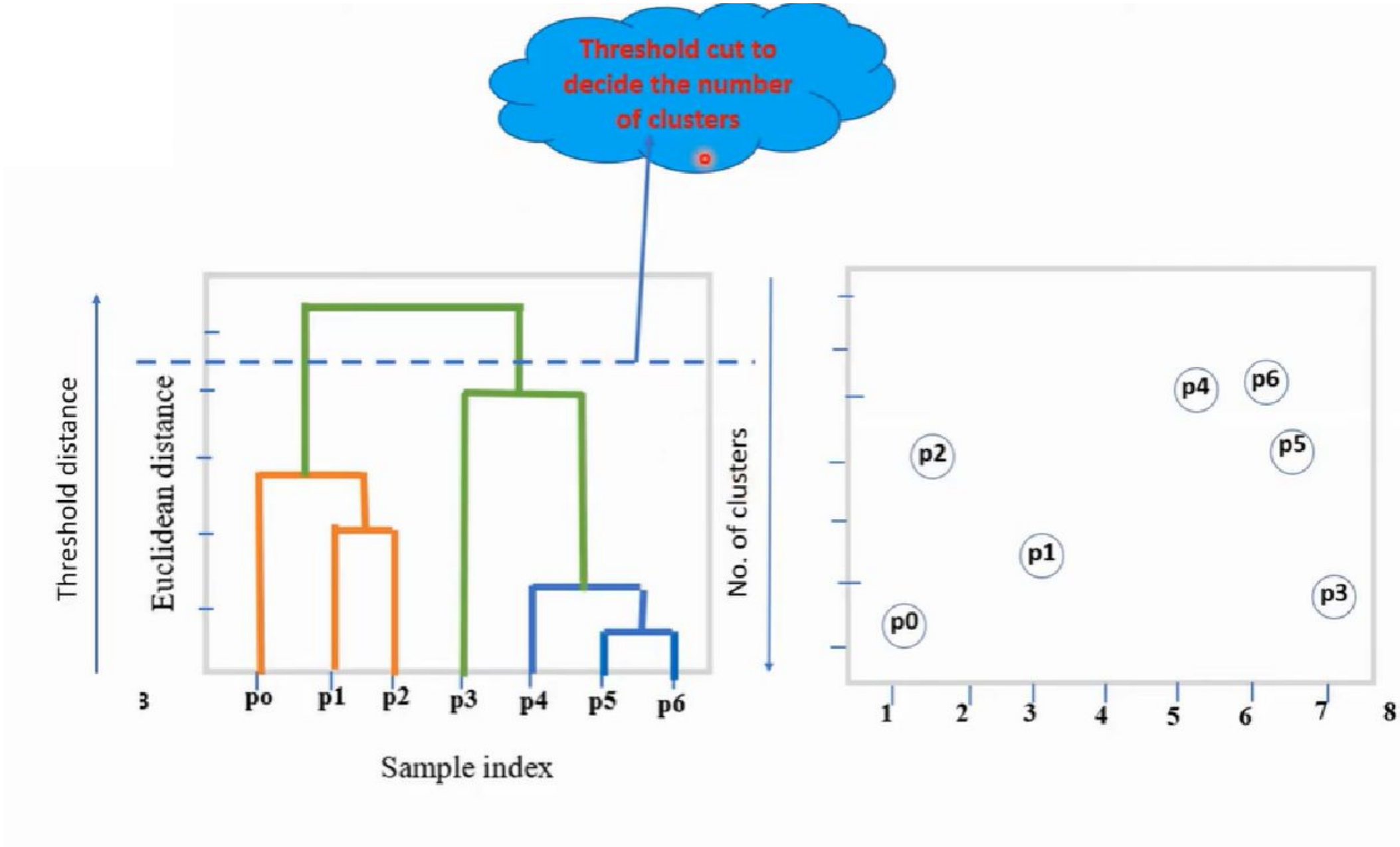
Step 5

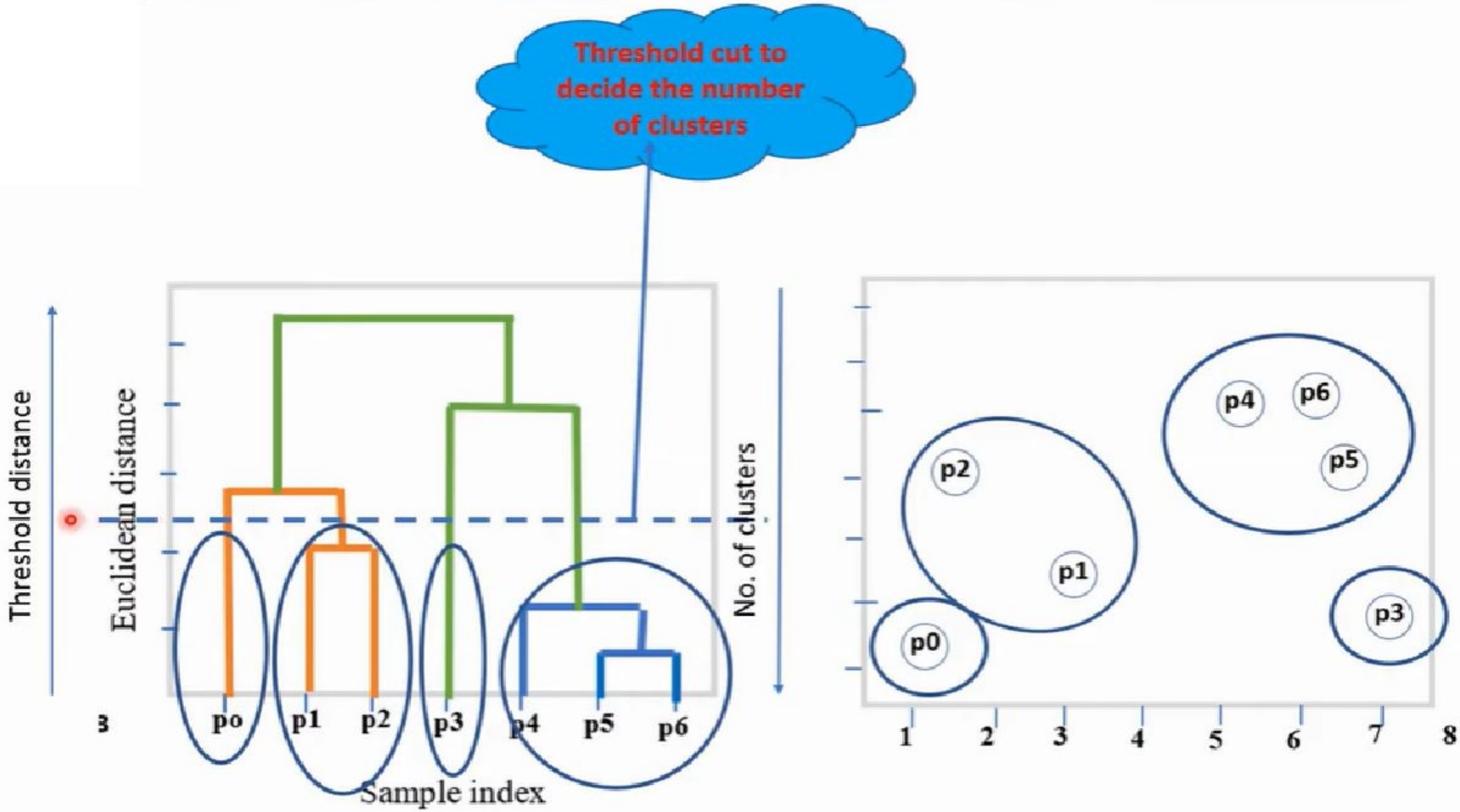

c6
1
2
3
4
5
6
7
8


Euclidean distance
p0
p1
p2
p3
p4
p5
p6
Sample index

Step 6

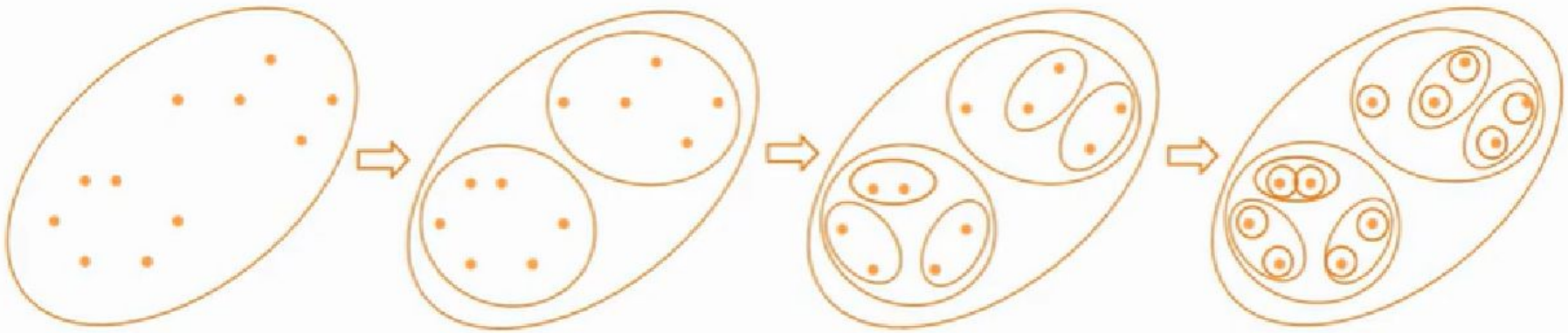
- ***Bottom-up* agglomerative clustering** applies the following algorithm:
  - Start with each data point in a separate cluster.
  - Repeatedly join the closest pair of clusters.
  - At each step, a stopping criterion is checked:
    - there is only one cluster; a predetermined number of clusters has been reached; the distance between the closest clusters is greater than a predetermined threshold; etc.


Threshold cut to decide the number of clusters
Threshold distance
Euclidean distance
No. of clusters
3
po
p1
p2
p3
p4
p5
p6
Sample index
p0
p1
p2
p3
p4
p5
p6
1
2
3
4
5
6
7
8


Threshold cut to decide the number of clusters
Threshold distance
Euclidean distance
No. of clusters
p0
p1
p2
p3
p4
p5
p6
Sample index
1
2
3
4
5
6
7
8

Top Down Approach

- The top-down approach starts with a set of data points as a **single cluster** and splits (or divisive) the cluster into further clusters until the preferred **number of clusters** is formed.
  - The top-down approach is also called as **Hierarchical Divisive Clustering (HDC).**



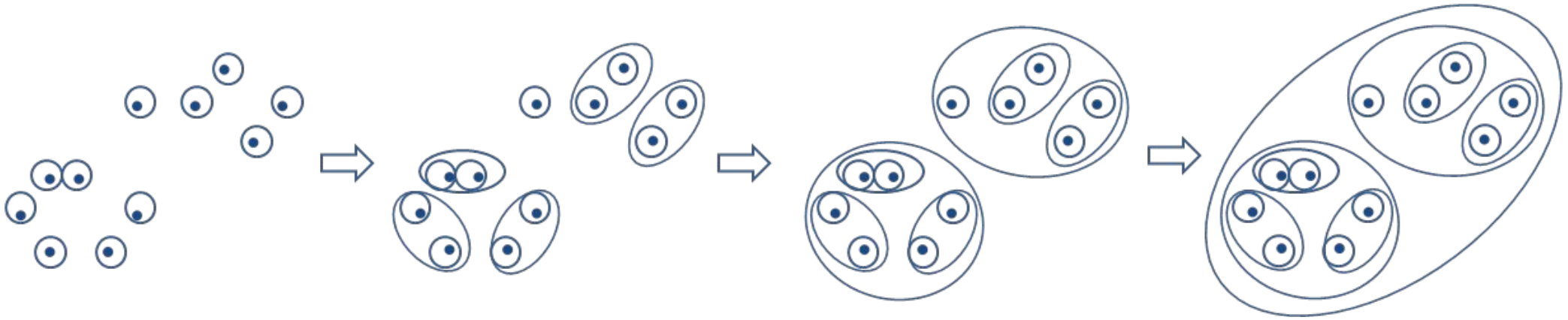
- ***Top-down* divisive clustering** applies the following algorithm:
  - Start with all the data in a single cluster.
  - Consider every possible way to divide the cluster into two.
  - Choose the best division.
  - Recursively, it operates on both sides until a stopping criterion is met.
    - there are as much clusters as data; the predetermined number of clusters has been reached; the maximum distance between all possible partition divisions is smaller than a predetermined threshold; etc.

- When merging two clusters in **agglomerative clustering**, how to measure the similarity of two clusters?
- The **linkage criterion** determines the metric used for the cluster merging strategy:
  - ***Maximum* or *complete* linkage**
    - minimizes the maximum distance between observations of pairs of clusters. Based on the similarity of the two least similar members of the clusters, this clustering tends to give tight spherical clusters as a final result.
  - ***Average* linkage**
    - averages similarity between members, i.e., minimizes the average of the distances between all observations of pairs of clusters.
  - ***Ward* linkage**
    - minimizes the sum of squared differences within all clusters. It is thus a variance-minimizing approach and in this sense is similar to the K-means objective function, but tackled with an agglomerative hierarchical approach.

Hierarchical Agglomerative Clustering

Hierarchical Agglomerative Clustering

- Each data point is treated as a **single cluster** and then iteratively merges with **pairs of clusters** until all clusters have been **merged into a single cluster** that holds all data points.
  - Hierarchical Agglomerative Clustering (HAC) is also called as bottom-up approach

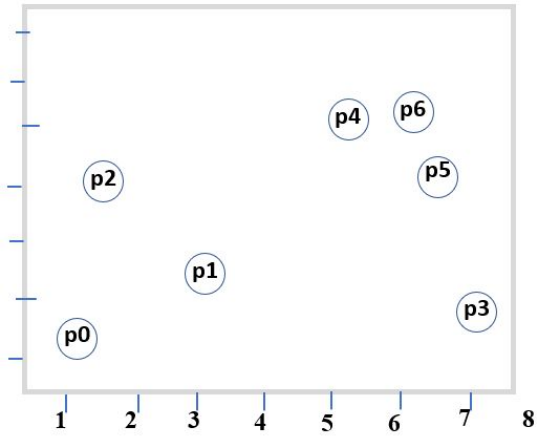


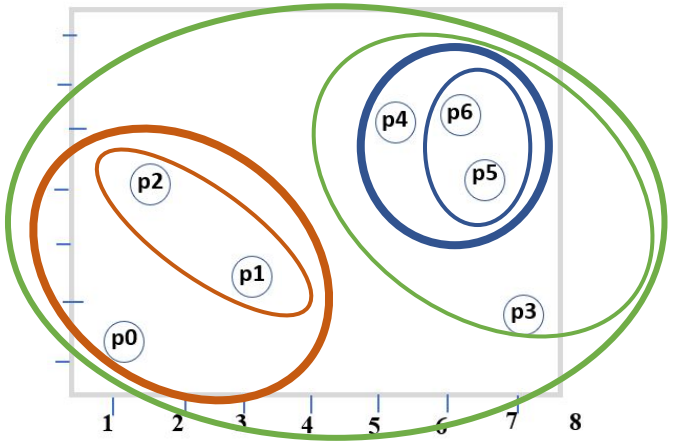
- Bottom-up clustering constructs a **dendrogram tree**.
- ***Bottom-up* agglomerative clustering** applies the following steps:
  - Start with each data point (P_1, P_2 ... P_n) in a separate cluster.
  - Compute a distance matrix between the clusters.
    - Distance metrics used is Euclidean metrics

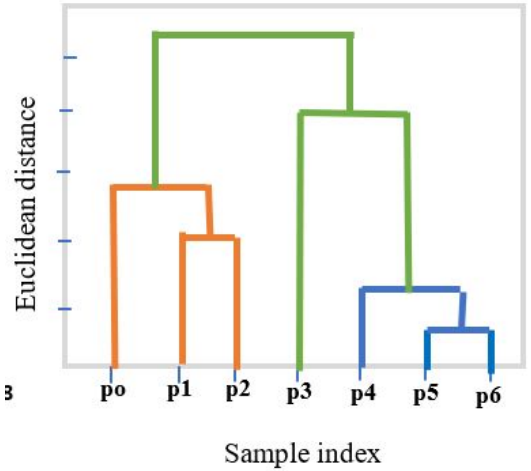
	P_1	P_2	...	P_n
P_1				
P_2				
...				
P_n				

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

  - Repeatedly join the closest pair of clusters and update the distance matrix.
  - Stop the algorithm when all data points are merged into the single cluster.

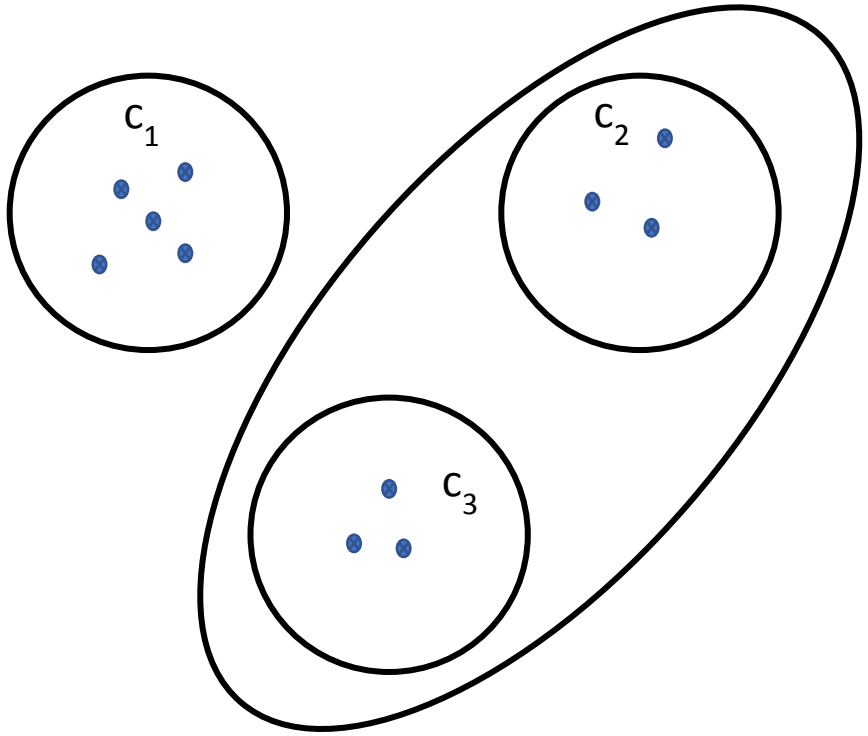






- How to measure the similarity of two clusters?
  - Linkage criteria

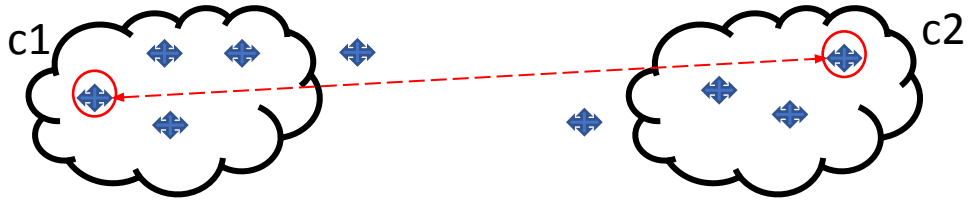
$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$



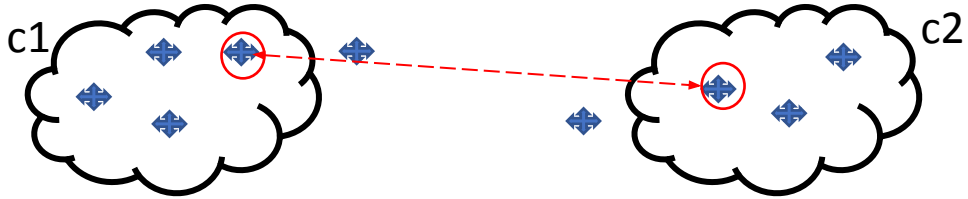
- **Complete linkage**
- **Single linkage**
- **Average linkage**

Cluster similarity measure

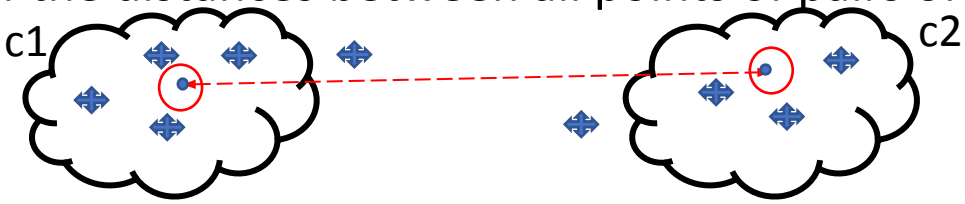
- **Complete linkage**
  - The similarity of two clusters is the similarity of their most **dissimilar** members (maximum distance)



- **Single linkage**
  - The similarity of two clusters is the similarity of their most **similar** members (minimum distance)



- **Average linkage**
  - The similarity of two clusters is the average of the distances between all points of pairs of clusters. (minimum distance between centroids)



Data point	x	y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

Problem statement:
Cluster the given dataset using hierarchical agglomerative clustering.
Also construct the dendrogram tree.
Note:
Single linkage criteria
Euclidean distance

Step 1:
Start with each data point (P_1, P_2, P_3, P_4, P_5, P_6) in a separate cluster.

Step 2:
Compute a distance matrix between the clusters.

$$d(P1, P2) = d(P2, P1)$$

$$d(P1, P2) = \sqrt{(0.40 - 0.22)^2 + (0.53 - 0.38)^2}$$
$$= \sqrt{(0.18)^2 + (0.15)^2}$$
$$= \sqrt{0.0324 + 0.0225} = \sqrt{0.0549}$$
$$d(P1,P2) = 0.23$$

	P1	P2	P3	P4	P5	P6
P1	**0**	0.23	0.22	0.37	0.34	0.23
P2	0.23	**0**	0.15	0.20	0.14	0.25
P3	0.22	0.15	**0**	0.15	0.28	0.11
P4	0.37	0.20	0.15	**0**	0.29	0.22
P5	0.34	0.14	0.28	0.29	**0**	0.39
P6	0.23	0.25	0.11	0.22	0.39	**0**

	P1	P2	P3	P4	P5	P6
P1	**0**	0.23	0.22	0.37	0.34	0.23
P2	0.23	**0**	0.15	0.20	0.14	0.25
P3	0.22	0.15	**0**	0.15	0.28	0.11
P4	0.37	0.20	0.15	**0**	0.29	0.22
P5	0.34	0.14	0.28	0.29	**0**	0.39
P6	0.23	0.25	0.11	0.22	0.39	**0**

Step 3:

Repeatedly join the closest pair of clusters and update the distance matrix.

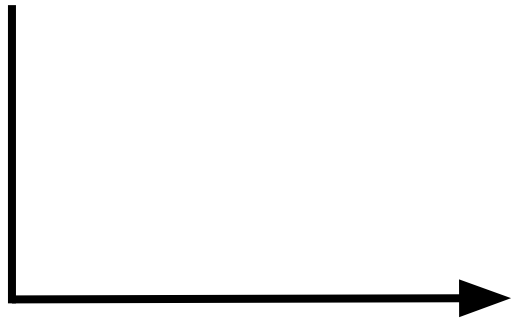
Merge
P3,P6

	P1	P2	P36	P4	P5
P1	**0**	0.23	0.22	0.37	0.34
P2	0.23	**0**	0.15	0.20	0.14
P36	0.22	0.15	**0**	0.15	0.28
P4	0.37	0.20	0.15	**0**	0.29
P5	0.34	0.14	0.28	0.29	**0**

	P1	P2	P36	P4	P5
P1	**0**	0.23	0.22	0.37	0.34
P2	0.23	**0**	0.15	0.20	0.14
P36	0.22	0.15	**0**	0.15	0.28
P4	0.37	0.20	0.15	**0**	0.29
P5	0.34	0.14	0.28	0.29	**0**

Step 3:
Repeatedly join the closest pair of clusters and update the distance matrix.

Merge
P2,P5



	P1	P25	P36	P4
P1	**0**	0.23	0.22	0.37
P25	0.23	**0**	0.15	0.20
P36	0.22	0.15	**0**	0.15
P4	0.37	0.20	0.15	**0**

	P1	P25	P36	P4
P1	**0**	0.23	0.22	0.37
P25	0.23	**0**	0.15	0.20
P36	0.22	0.15	**0**	0.15
P4	0.37	0.20	0.15	**0**

Step 3:

Repeatedly join the closest pair of clusters and update the distance matrix.

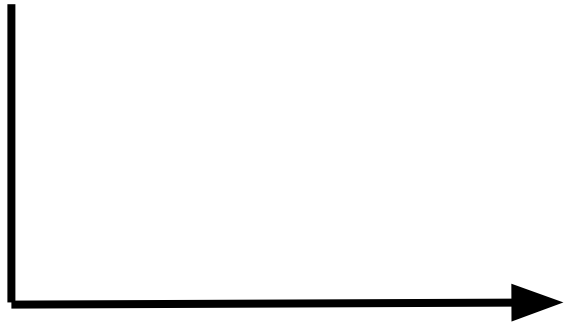
Merge P25,P36

	P1	P25 36	P4
P1	**0**	0.22	0.37
P25 36	0.22	**0**	0.15
P4	0.37	0.15	**0**

	P1	P2536	P4
P1	**0**	0.22	0.37
P2536	0.22	**0**	0.15
P4	0.37	0.15	**0**

Step 3:
Repeatedly join the closest pair of clusters and update the distance matrix.

Merge P2536,P4



	P1	P2536P4
P1	**0**	0.22
P2536P4	0.22	**0**

	P1	P2	P3	P4	P5	P6
P1	**0**	0.23	0.22	0.37	0.34	0.23
P2	0.23	**0**	0.15	0.20	0.14	0.25
P3	0.22	0.15	**0**	0.15	0.28	0.11
P4	0.37	0.20	0.15	**0**	0.29	0.22
P5	0.34	0.14	0.28	0.29	**0**	0.39
P6	0.23	0.25	0.11	0.22	0.39	**0**

Merge 3,6

	P1	P2	P36	P4	P5
P1	**0**	0.23	0.22	0.37	0.34
P2	0.23	**0**	0.15	0.20	0.14
P36	0.22	0.15	**0**	0.15	0.28
P4	0.37	0.20	0.15	**0**	0.29
P5	0.34	0.14	0.28	0.29	**0**

Merge 2,5

	P1	P25	P36	P4
P1	**0**	0.23	0.22	0.37
P25	0.23	**0**	0.15	0.20
P36	0.22	0.15	**0**	0.15
P4	0.37	0.20	0.15	**0**

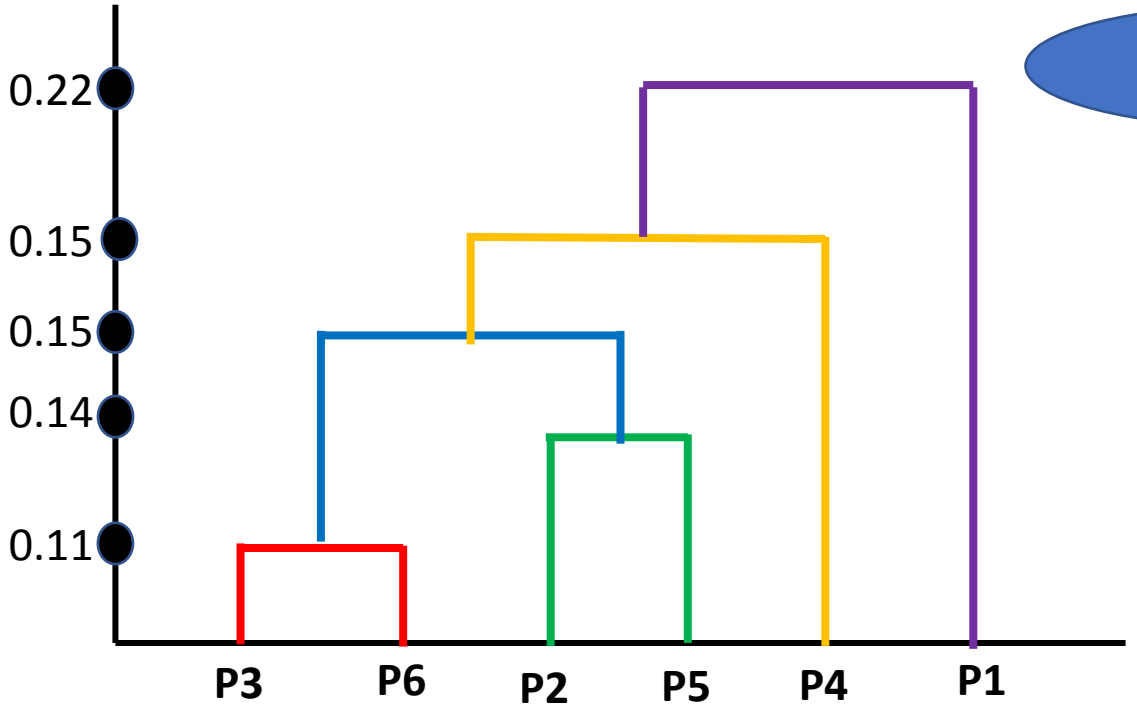
Merge 25,36

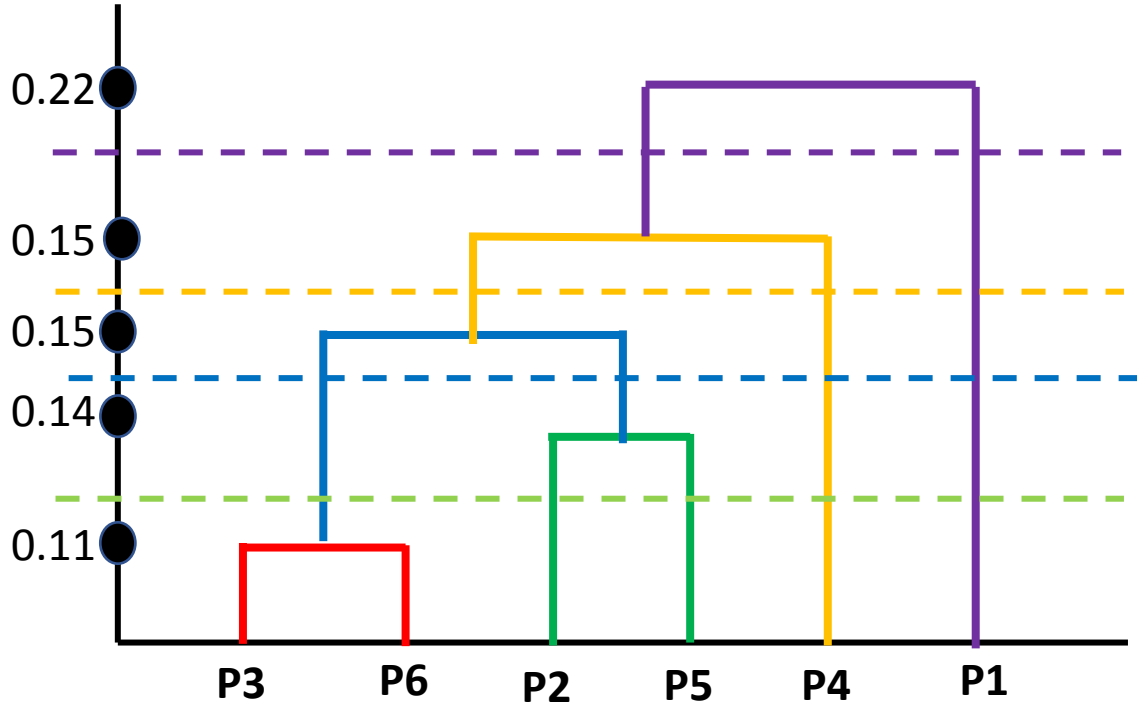
	P1	P25 36	P4
P1	**0**	0.22	0.37
P25 36	0.22	**0**	0.15
P4	0.37	0.15	**0**

Merge 2536,4

	P1	P2536P4
P1	**0**	0.22
P2536P4	0.22	**0**

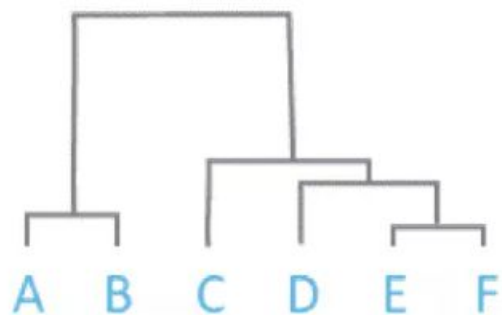
Dendrogram




0.22
0.15
0.15
0.14
0.11
P3
P6
P2
P5
P4
P1

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.70	0.45	0	
F	0.34	0.61	0.93	0.20	0.67	0

Dendrogram



Advantages

- The agglomerative technique is easy to implement.
- It can produce an ordering of objects, which may be informative for the display.
- In agglomerative Clustering, there is no need to pre-specify the number of clusters.
- By the Agglomerative Clustering approach, smaller clusters will be created, which may discover similarities in data

Disadvantages

- The agglomerative technique gives the best result in some cases only.
- The algorithm can never undo what was done previously, which means if the objects may have been incorrectly grouped at an earlier stage, and the same result should be close to ensure it.
- The usage of various distance metrics for measuring distances between the clusters may produce different results. So performing multiple experiments and then comparing the result is recommended to help the actual results' veracity.


THANK
YOU