FACULTY OF EGINEERING AND TECHNOLOGY
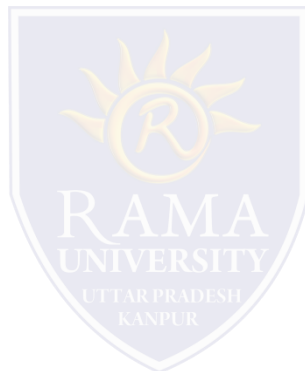
Soft Computing

LECTURE -19

Umesh Kumar Gera
Assistant Professor
Computer Science & Engineering

OUTLINE

- **Introduction of Data Clustering**

- **Classification tree of clustering algorithms**

- **Hierarchical Approach**

- **Partitioned Approach**

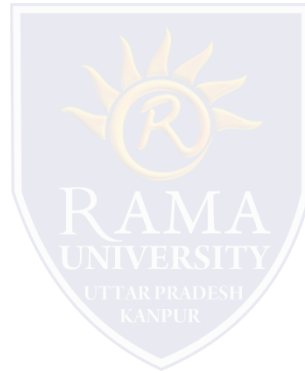- **Graph-Theoretic Approach**

- **References**

Introduction of data clustering

A cluster is made up of a number of similar objects grouped together. Everett [1974] defines a cluster as a set of entities which are alike and entities from different clusters are not alike. The clustering process has to find the hidden structure in the data.
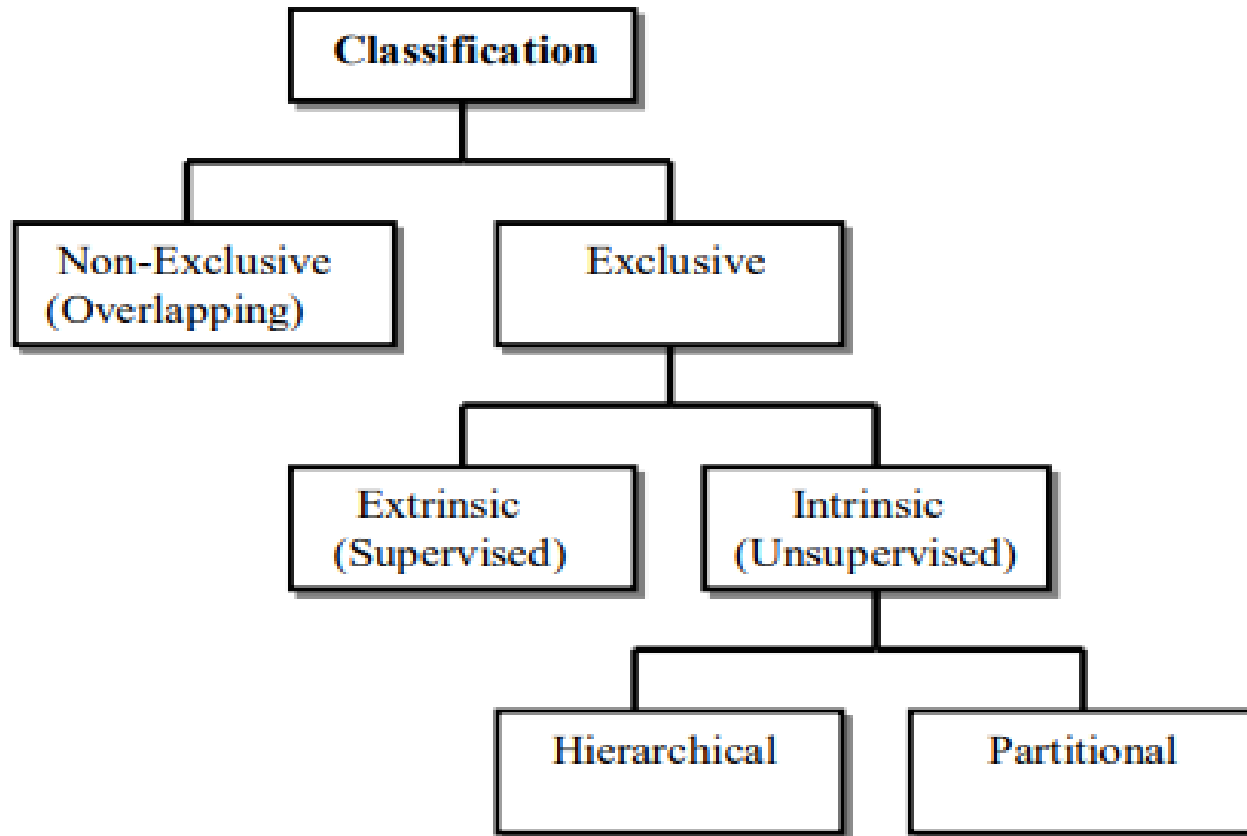
The typical components of a clustering task [Jain and Dubs 1988] are :

1.   pattern representation

2.   proximity/similarity measures

3.   clustering algorithms

4.   cluster evaluation/ validation

5.   interpretation of results

Classification tree of clustering algorithms

Hierarchical Approach

Hierarchical clustering builds a hierarchy of clusters.

The traditional representation of this hierarchy is a tree,

with individual elements at one end and a single cluster

with every element at the other. A hierarchical algorithm

yields a dendrogram representing the nested grouping

of patterns and similarity levels at which groupings

change.
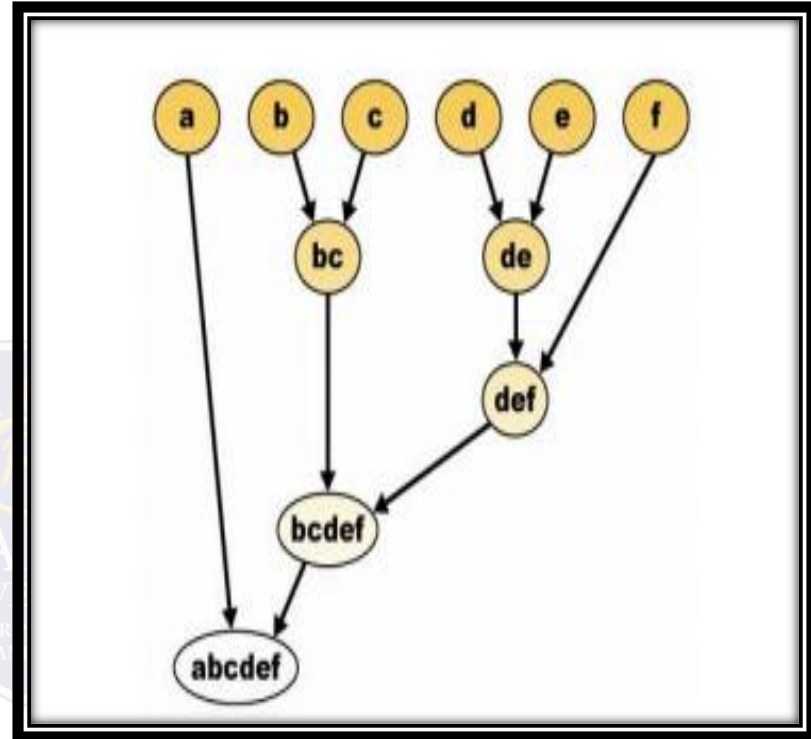
Partitioned Approach

A partitioned clustering algorithm obtains a single partition of the data instead of a clustering structure, such as the dendrogram produced by a hierarchical technique.

The k-means is the simplest and most commonly used partitioned algorithm employing a squared error criterion [McQueen 1967].

1. Randomly generate k clusters and determine the cluster centers or directly generate k seed points as cluster centers.

2. Assign each point to the nearest cluster center.

3. Recomputed the new cluster centers.

4. Repeat until the convergence criterion is met
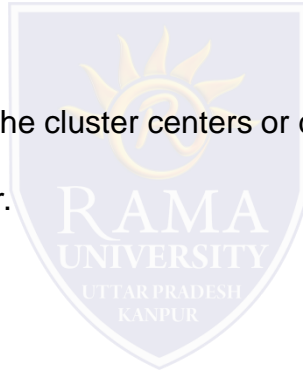
Graph-Theoretic Approach
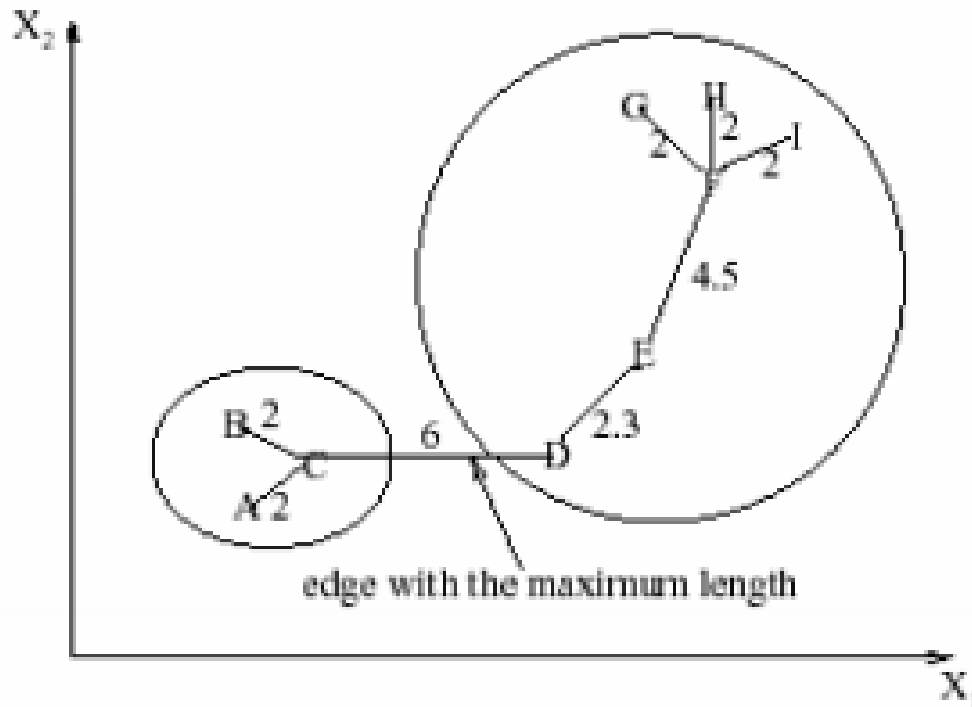
Grid-based Algorithms

A number of clustering algorithms have been presented for spatial data, known as grid-based algorithms. These algorithms quantize the space into a finite number of cells and then do all operations on the quantized space. Then it generates a hierarchical structure of the grid cells so as to represent the clustering information at different levels. STING (Statistical Information Grid-based method) is representative of this category. It divides the spatial area into rectangular cells using a hierarchical structure.
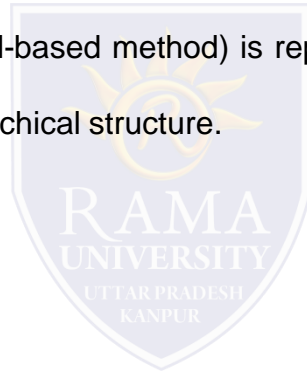
MULTIPLE CHOICE QUESTION

1. The process by which you become aware of messages through your sense is called _____

a) Organization

b) Sensation

c) Interpretation-Evaluation

d) Perception

2. Susan is so beautiful; I bet she is smart too. This is an example of _____

a) The halo effect

b) The primary effect

c) A self-fulfilling prophecy

d) The recency effect

3. _____ prevents you from seeing an individual as an individual rather than as a member of a group.

a) Cultural mores

b) Stereotypes

c) Schematas

d) Attributions

4. When you get fired from your job and you determine it is because your boss dislikes you, you are most likely exhibiting?

a) Self-promotion

b) Fundamental attribution error

c) Over-attribution

d) Self-serving bias

5. What is Mindless processing?

a) careful, critical thinking

b) inaccurate and faulty processing

c) information processing that relies heavily on familiar schemata

d) processing that focuses on unusual or novel events
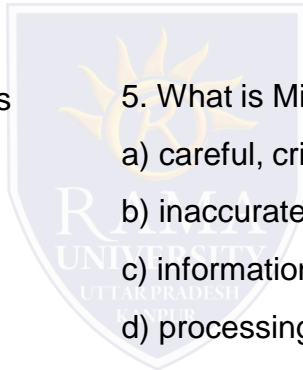
REFERENCES

❏ https://notendur.hi.is/benedikt/Courses/ch14_2005.pdf