Module - 5

Clustering Analysis

Nithin Kumar
Asst professor
Dept of CSE
VVCE, Mysuru

Cluster Analysis groups data objects based only on information found in the data that describes the objects & their relationships.

→ The goal is that the objects within a group be similar to one another & different from the objects in other groups.

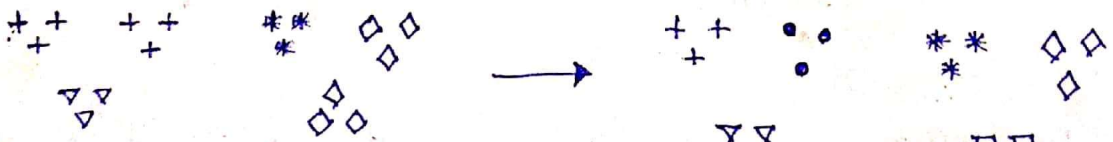
The greater the similarity within a group & the greater the difference b/w groups the better or more distinct the clustering.

* Different types of clusterings :- The various types of clusterings are

* Hierarchical v/s partitional :- A "partitional clustering" is simply a division of the set of data objects into non-overlapping subsets (clusters) such that each data object is in exactly one subset.

(a) original points

(b) 2 - clusters



(c) 4 - clusters

(d) 6 - clusters

→ "If

are organized as a tree.
Each node (cluster) in the tree is a union of its children (sub clusters) & root of the tree is the cluster containing all objects but leafs of tree are single clusters of individual data objects.

* Exclusive V/s overlapping V/s Fuzzy :- In "Exclusive clustering", we assign each object to a single cluster. There are many situations in which a point could reasonably be placed in more than one cluster & these situations are better addressed by Non-Exclusive clustering.

→ "Non-Exclusive" or "overlapping" clustering is used to reflect the fact than an object can simultaneously belong to more than one group (class)
For Ex, A person at a University can be both an enrolled student & Employee of University.

→ "Fuzzy" clustering, where every objects belongs to every cluster with membership weight that is between "0" & "1"
"0" (Absolutely doesn't belong) & "1" (Absolutely belongs).
clusters are treated as fuzzy sets i.e one in which an object belongs to any set with a weight that is between 0 & 1.

* Complete V/s partial :- A complete clustering assigns every object to a cluster, where as a partial clustering does not.

→ The motivation for a partial clustering is that some objects in a data set may not belong to well-defined

→ For Example, Some Newspaper Stories May have a Common theme, while other Stories are More Generic or "one-of-a-kind".

* Different Types of clusters :-

Clustering Aims to find useful Groups of objects (clusters), where usefulness is defined by the Goals of Data Analysis.

→ There are Several different Notations of clusters that prove -ve useful in practice

* Well-Separated :- The data objects with a Cluster Must have Small distance & distance b/w two clusters Must be high.

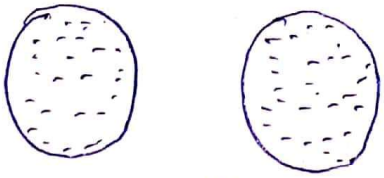


fig:- Well-Separated Clusters

Sometimes a threshold is used to Specify that all the objects in a cluster Must be Sufficiently close (Similar) to one Another.

* prototype-Based clusters (Center-Based) :- A cluster is a Set of objects in which Each object is closer (Similar) to the prototype that defines the cluster than to the prototype of Any other clusters.