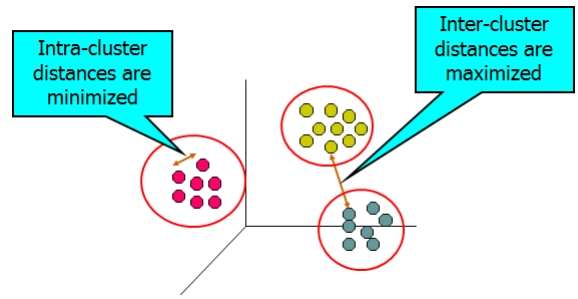
Cluster Analysis

- What is Cluster Analysis?
- Types of Clustering

What is Clustering?

- Finding groups of objects in data such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups.



What is Clustering?

Applications of cluster analysis:

- Group diseases in the field of medicine
- Group related documents for browsing
- Group genes and proteins that have similar functionality
- Group stocks with similar price fluctuations

3

Types of Clustering

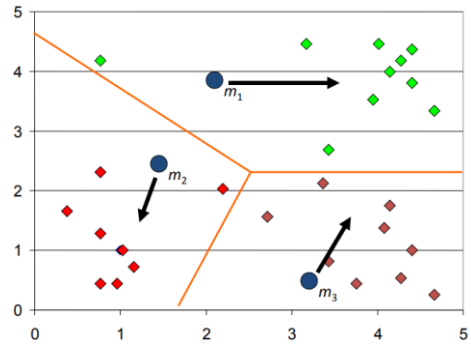
There are different **types of clustering** methods, including:

- Partitioning Clustering
- Hierarchical Clustering
- Density-based Clustering
- Grid-based Clustering

4

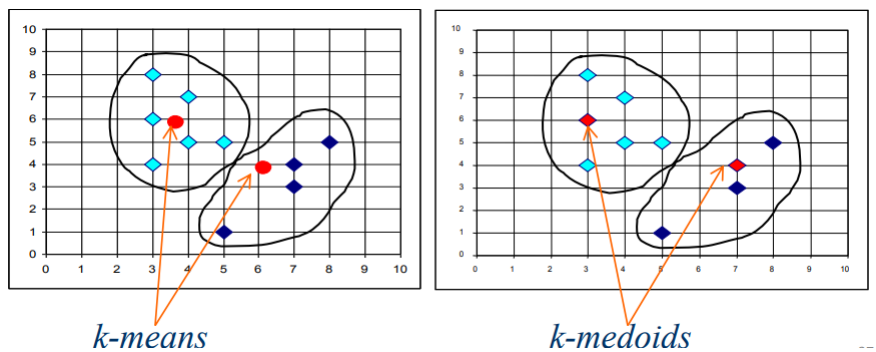
Partitioning Clustering

- Effective methods: K-means and K-medoids algorithms.
- Each point is assigned to the cluster with the closest centroid.



5

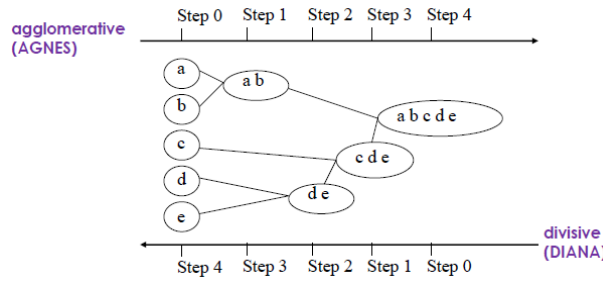
Partitioning Clustering



6

Hierarchical Clustering

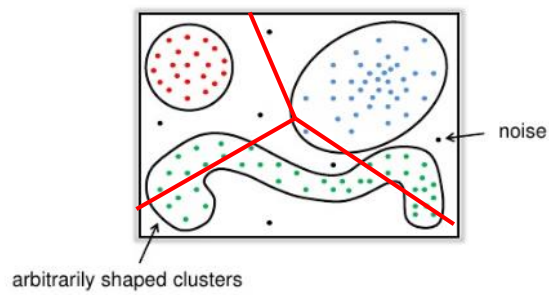
- Hierarchical methods can be
  Agglomerative: bottom-up approach
  Divisive: top-down approach



7

Density Based Clustering

- A cluster is defined as a connected dense component which can grow in any direction that density leads.
- Arbitrary shaped cluster.



8

Density Based Clustering

- **DBSCAN** Basic Idea:
  - Group together points in high-density
  - Mark as outliers points that lie alone in low-density regions
- **OPTICS:** Ordering points to identify the clustering structure is an algorithm for finding density-based clusters.

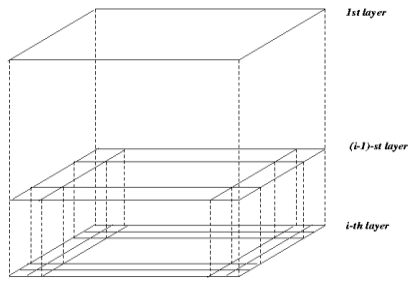
9

Grid Based Clustering

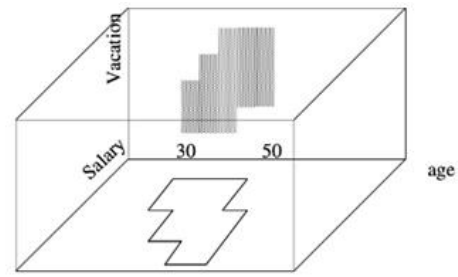
- Quantizes space into a finite number of cells that form a grid structure.
- Grid based methods can be:
  - Sting: A statistical information grid approach
  - Clique: A dimension growth subspace clustering method

10

Grid Based Clustering



Sting



Clique

11

Cluster Validation

- The procedure of evaluating the "goodness" of the resulting clusters is known as a cluster validity.
- We have to pick the number of clusters.

12