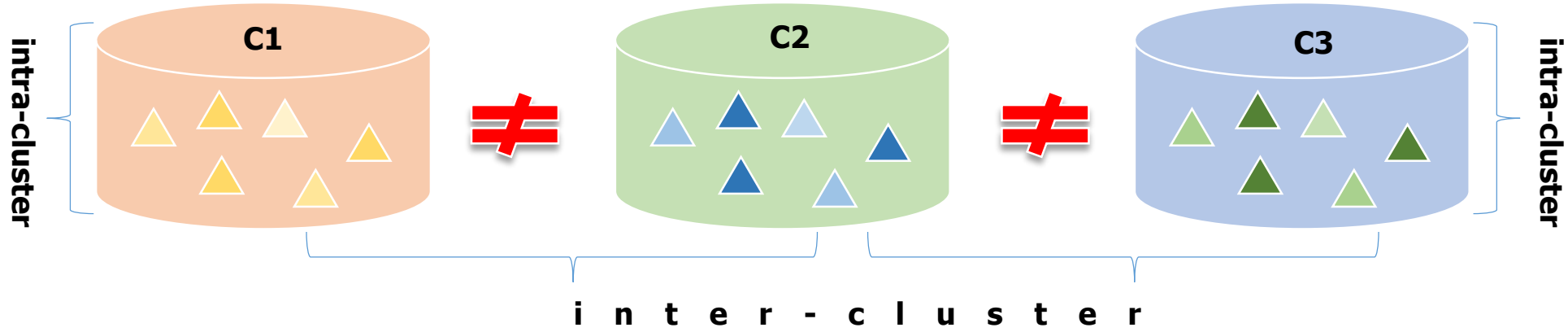
Unsupervised Learning

Clustering

- A technique to group similar observations into one type
- Use of clustering
  - ✓ To understand certain patterns. Eg:
    - ➢ Shopping pattern of buyers
    - ➢ Performance patterns of schools / colleges
    - ➢ Can use this to identify missing class

Cluster properties

- High **intra-cluster** similarity (within cluster)
- Low **inter-cluster** similarity (across clusters)



Types of Clustering

- ❑ **Exclusive Clustering** – data belongs to only 1 cluster (k-means)
- ❑ **Overlapping Clustering** – data can belong to multiple clusters (Fuzzy C-means)
- ❑ **Hierarchical Clustering** – union between 2 nearest clusters (Eg: Animal kingdom)

Some prominent clustering algorithms are

- **Connectivity-based clustering**: Clusters formed according to their distances (Hierarchical)
- **Centroid-based clustering:** Iterative clustering algorithm where a similarity of an observation is determined by the closeness with the cluster's centroid
- **Distribution-based clustering:** Probability of a cluster belonging to a certain type of statistical distribution
- **Density-based clustering:** Observations are classified as Higher area and lower area densities before grouping them. Requires density guidelines

- **Clustering**
  - **k-Means**
  - **Hierarchical**

- **Dimensional Reduction**
  - **PCA (Principal Component Analysis)**
  - **LDA (Linear Discriminant Analysis)**
  - **Kernel PCA**

- **Market-Basket Analysis**
  - **Apriori**