

The background features a complex geometric pattern of thin, light-colored lines forming a network of triangles and polygons. Overlaid on this are several semi-transparent elements: a horizontal band of purple and blue wavy patterns at the top, a vertical band of orange and red wavy patterns on the left, and a large white banner with a grey border that contains the title. Small grey plus signs are scattered throughout the background.

# **Requirements and Challenges**

# Considerations for Cluster Analysis

---

## ❑ Partitioning criteria

- ❑ Single level vs. hierarchical partitioning (often, multi-level hierarchical partitioning is desirable, e.g., grouping topical terms)

## ❑ Separation of clusters

- ❑ Exclusive (e.g., one customer belongs to only one region) vs. non-exclusive (e.g., one document may belong to more than one class)

## ❑ Similarity measure

- ❑ Distance-based (e.g., Euclidean, road network, vector) vs. connectivity-based (e.g., density or contiguity)

## ❑ Clustering space

- ❑ Full space (often when low dimensional) vs. subspaces (often in high-dimensional clustering)

# Requirements and Challenges

---

## □ Quality

- Ability to deal with different types of attributes: Numerical, categorical, text, multimedia, networks, and mixture of multiple types
- Discovery of clusters with arbitrary shape
- Ability to deal with noisy data

## □ Scalability

- Clustering all the data instead of only on samples
- High dimensionality
- Incremental or stream clustering and insensitivity to input order

## □ Constraint-based clustering

- User-given preferences or constraints; domain knowledge; user queries

## □ Interpretability and usability