

The background of the slide is a complex, abstract composition. It features a dark, reddish-brown base color. Overlaid on this are several geometric and data-related elements: a network of thin, light-colored lines forming a mesh or web; numerous small, green and blue dots scattered across the field; and a series of vertical, slightly wavy lines on the right side. In the center, there is a large, white, angular shape that resembles a stylized letter 'A' or a large arrow pointing downwards, which serves as a backdrop for the title. To the left of this central shape, there is a smaller, rectangular inset image showing a cluster of orange and red dots on a light blue background, with a grid of small black crosses overlaid. The overall aesthetic is technical and data-driven.

# **Basic Concepts of Partitioning Algorithms**

# Partitioning Algorithms: Basic Concepts

---

- ❑ Partitioning method: Discovering the groupings in the data by optimizing a specific objective function and iteratively improving the quality of partitions
- ❑ *K*-partitioning method: Partitioning a dataset ***D*** of ***n*** objects into a set of ***K*** clusters so that an objective function is optimized (e.g., the sum of squared distances is minimized, where  $c_k$  is the centroid or medoid of cluster  $C_k$ )

❑ A typical objective function: **Sum of Squared Errors (SSE)**

$$SSE(C) = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - c_k\|^2$$

- ❑ Problem definition: Given *K*, find a partition of *K clusters* that optimizes the chosen partitioning criterion
  - ❑ Global optimal: Needs to exhaustively enumerate all partitions
  - ❑ Heuristic methods (i.e., greedy algorithms): *K-Means*, *K-Medians*, *K-Medoids*, etc.