

The background of the slide is a complex, abstract composition. It features a dark, muted purple or brownish background. Overlaid on this are several geometric and data-like elements: a network of thin, light-colored lines forming a mesh or web-like structure; numerous small, colored dots (green, blue, yellow) scattered across the field; and a series of faint, light-colored plus signs (+) arranged in a grid-like pattern. A prominent white, angular shape, resembling a stylized 'V' or a folded piece of paper, cuts across the center of the image, serving as a backdrop for the title. In the bottom left corner, there is a small, rectangular inset image showing a cluster of orange and red dots, possibly representing a specific data set or a biological structure like a galaxy or a cell cluster.

What Is Cluster Analysis?

What Is Cluster Analysis?

❑ What is a cluster?

- ❑ A cluster is a collection of data objects which are
 - ❑ Similar (or related) to one another within the same group (i.e., cluster)
 - ❑ Dissimilar (or unrelated) to the objects in other groups (i.e., clusters)

❑ Cluster analysis (or *clustering*, *data segmentation*, ...)

- ❑ Given a set of data points, partition them into a set of groups (i.e., clusters) which are as similar as possible
- ❑ Cluster analysis is **unsupervised learning** (i.e., no predefined classes)
 - ❑ This contrasts with *classification* (i.e., *supervised learning*)
- ❑ Typical ways to use/apply cluster analysis
 - ❑ As a stand-alone tool to get insight into data distribution, or
 - ❑ As a preprocessing (or intermediate) step for other algorithms