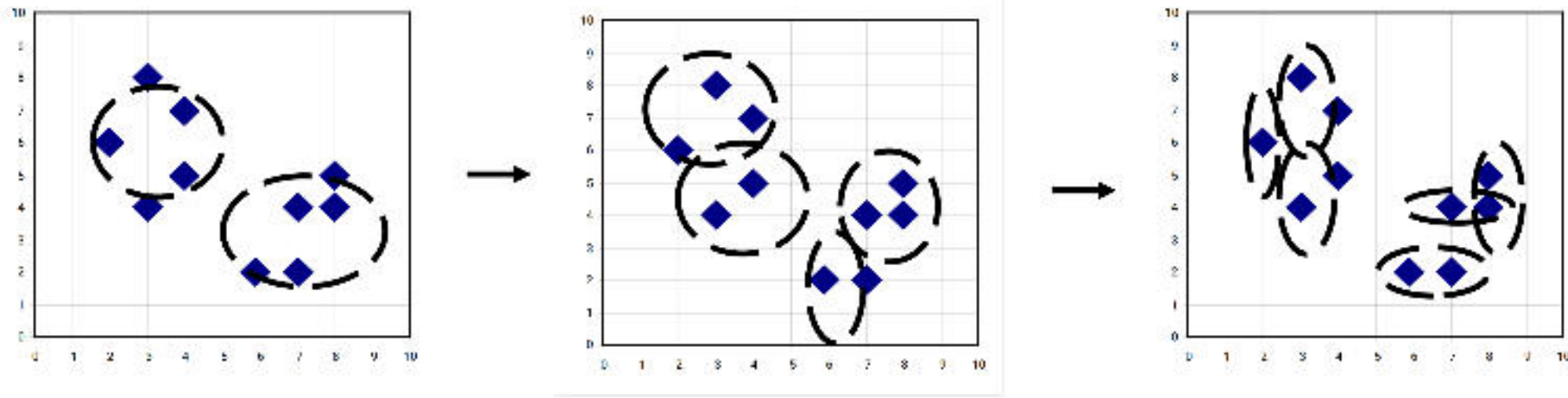


The background features a complex, abstract design. It includes a network of thin, reddish-brown lines connecting various points, some of which are green dots. There are also larger, fainter geometric shapes and patterns in shades of brown and grey. A white, angular shape, resembling a stylized 'V' or a folded piece of paper, is positioned behind the title text. On the left side, there is a small, rectangular inset showing a cluster of orange and red dots on a light background, with a grid of small '+' symbols overlaid.

Divisive Clustering Algorithms

Divisive Clustering

- DIANA (Divisive Analysis) (Kaufmann and Rousseeuw, 1990)
 - Implemented in some statistical analysis packages, e.g., Splus
- Inverse order of AGNES: Eventually each node forms a cluster on its own



- Divisive clustering is a top-down approach
 - The process starts at the root with all the points as one cluster
 - It recursively splits the higher level clusters to build the dendrogram
 - Can be considered as a global approach
 - More efficient when compared with agglomerative clustering

More on Algorithm Design for Divisive Clustering

- Choosing which cluster to split
 - Check the sums of squared errors of the clusters and choose the one with the largest value
- Splitting criterion: Determining how to split
 - One may use Ward's criterion to chase for greater reduction in the difference in the SSE criterion as a result of a split
 - For categorical data, Gini-index can be used
- Handling the noise
 - Use a threshold to determine the termination criterion (do not generate clusters that are too small because they contain mainly noises)