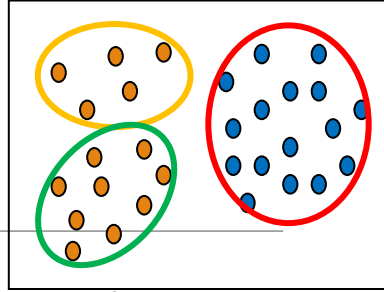


The background of the slide is a complex, abstract composition. It features a central white banner with a subtle, light gray geometric pattern. This banner is flanked by two large, semi-transparent panels. The left panel displays a grid of small, light gray plus signs. The right panel shows a network of thin, light gray lines connecting various points. Overlaid on these panels are two main visual elements: a large, semi-transparent red and orange geometric pattern on the left, and a large, semi-transparent green and blue geometric pattern on the right. The central white banner contains the text "Relative Measures" in a bold, black, sans-serif font. The overall aesthetic is modern and technical, suggesting a focus on data analysis or mathematics.

Relative Measures

Relative Measure



- Relative measure: Directly compare different clusterings, usually those obtained via different parameter settings for the same algorithm

- Silhouette coefficient as an internal measure:** Check cluster cohesion and separation

- For each point \mathbf{x}_i , its silhouette coefficient s_i is:
$$s_i = \frac{\mu_{out}^{\min}(\mathbf{x}_i) - \mu_{in}(\mathbf{x}_i)}{\max\{\mu_{out}^{\min}(\mathbf{x}_i), \mu_{in}(\mathbf{x}_i)\}}$$
 where $\mu_{in}(\mathbf{x}_i)$ is the mean distance from \mathbf{x}_i to points in its own cluster

$\mu_{out}^{\min}(\mathbf{x}_i)$ is the mean distance from \mathbf{x}_i to points in its closest cluster

- Silhouette coefficient (SC) is the mean values of s_i across all the points:
$$SC = \frac{1}{n} \sum_{i=1}^n s_i$$

- SC close to +1 implies good clustering

- Points are close to their own clusters but far from other clusters

- Silhouette coefficient as a relative measure:** Estimate the # of clusters in the data

$$SC_i = \frac{1}{n_i} \sum_{x_j \in C_i} s_j$$

Pick the k value that yields the best clustering, i.e., yielding high values for SC and SC_i ($1 \leq i \leq k$)