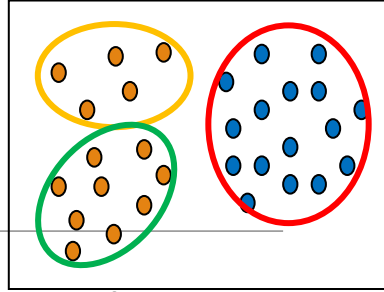# Relative Measure



❑ 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 $\boldsymbol{x}_i$, its silhouette coefficient $s_i$ is: $s_i = \dfrac{\mu_{out}^{\min}(\mathbf{x}_i) - \mu_{in}(\mathbf{x}_i)}{\max\{\mu_{out}^{\min}(\mathbf{x}_i), \mu_{in}(\mathbf{x}_i)\}}$

  where $\mu_{in}(\mathbf{x}_i)$ is the mean distance from $\boldsymbol{x}_i$ to points in its own cluster

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

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

  ❑ *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

$SC_i = \dfrac{1}{n_i}\sum_{x_j \in C_i} s_j$    Pick the *k* value that yields the best clustering, i.e., yielding high values for *SC* and *SC*$_i$ (1 ≤ *i* ≤ *k*)

2