

Assessment Pt 2

10/5

BetaCU measure: ratio of mean intracluster dist to the mean intercluster dist

$$\text{BetaCU} = \frac{w_{in}/N_{in}}{w_{out}/N_{out}} = \frac{N_{out}}{N_{in}} \cdot \frac{w_{in}}{w_{out}} = \frac{N_{out} \sum_{i=1}^k w(C_i, C_i)}{N_{in} \sum_{i=1}^k w(C_i, \bar{C}_i)}$$

Small BetaCU \Rightarrow on avg intracluster dist is smaller than intercluster dist

Silhouette Coeff: measure cohesion & separation of clusters

$$\mu_{in}(\vec{x}_i) = \frac{\sum_{x_j \in C_{\hat{y}_i} \setminus \{\vec{x}_i\}} \|\vec{x}_j - \vec{x}_i\|}{n_{\hat{y}_i} - 1}$$

mean dist from \vec{x}_i to all other pts in cluster

$$\mu_{out}(\vec{x}_i) = \min_{j \neq \hat{y}_i} \left\{ \frac{\sum_{\vec{z} \in C_j} \|\vec{z} - \vec{x}_i\|}{n_j} \right\}$$

mean dist to all pts in closest cluster

Silhouette coeff for a pt

$$\forall \vec{x}_i \in D$$

$$s_i = \frac{\mu_{out}(\vec{x}_i) - \mu_{in}(\vec{x}_i)}{\max\{\mu_{out}(\vec{x}_i), \mu_{in}(\vec{x}_i)\}} \in [-1, 1]$$

close to -1
 \Rightarrow pt is closer to another cluster than its own cluster

close to 0
 \Rightarrow pt is close to bdd of cluster

close to +1
 \Rightarrow pt is closer to pts in its own cluster

Silhouette Coeff of clustering? mean s_i across all pts

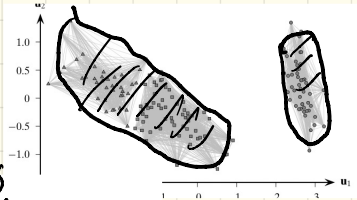
$$SC = \frac{1}{n} \sum_{i=1}^n s_i$$

a value closer to +1 is a better clustering

Relative measures - diff parms of same algo

vary k for k -means

$k=3$



lots of
High vals

