

Silhouette
Coefficient

Cluster Validity

- Extrinsic - ground truth is available

Examples

- F-Score

- Precision

- Recall

Cluster Validity

Extrinsic - ground truth is available

Examples: Precision, Recall, F-Score

Intrinsic - ground truth is not available

- Comparing Inter/Intra clust dist (BetaCU, C-index)
- Compactness (Davies-Bouldin Idx)
- Sim Pts in diff clust (Normalized cuts, modularity, Silhouette Coeff, Dunn Idx)

Silhouette Coeff

Given $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ w/ $\text{label}(x_i) = \hat{y}_i$

$\{\hat{y}_1, \hat{y}_2, \hat{y}_3, \dots, \hat{y}_n\}$

$$\text{Let } \mu_{in}(x_i) = \frac{\sum_{j \in C_{\hat{y}_i}} d(x_i, x_j)}{|C_{\hat{y}_i}| - 1}$$

avg dist to
points in
my cluster

$$\mu_{out}^{min}(x_i) = \min_{j \neq \hat{y}_i} \left\{ \frac{\sum_{p \in C_j} d(x_i, p)}{|C_j|} \right\}$$

avg dist
to "2nd
Closest"
cluster.

diff between
dist in cluster and
dist in cluster and
"2nd closest cluster"

$$s_i = \frac{\mu_{out}^{min}(x_i) - \mu_{in}(x_i)}{\max \{\mu_{out}^{min}(x_i), \mu_{in}(x_i)\}}$$

rescale
[-1, 1]

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

avg of
 s_i 's

SC Props & Meaning

- Compares mean dist of points to other points in cluster
- $[-1, 1]$
- $+1$ indicates points are generally "closer" to their cluster than other cluster

SC Example

	X ₁
X ₁	4
X ₂	1.1
X ₃	12
X ₄	16.4
X ₅	2.3
X ₆	5
X ₇	15
X ₈	13.7
X ₉	3.5



$$C = \{ C_1 = \{x_2, x_5\},$$

$$C_2 = \{x_1, x_3, x_4, x_6, x_7, x_8, x_9\} \}$$

$$s_i = \frac{\mu_{out}^{\min}(x_i) - \mu_{in}(x_i)}{\max\{\mu_{out}^{\min}(x_i), \mu_{in}(x_i)\}}$$

$$s_5 = \frac{\mu_{out}^{\min}(x_5) - \mu_{in}(x_5)}{\max\{\mu_{out}^{\min}(x_5), \mu_{in}(x_5)\}}$$

$$\mu_{out}^{\min}(x_5) = \min_{j \neq i, j \in C_1} \left\{ \frac{\sum_{p \in C_j} \delta(x_5, p)}{|C_j|} \right\}$$

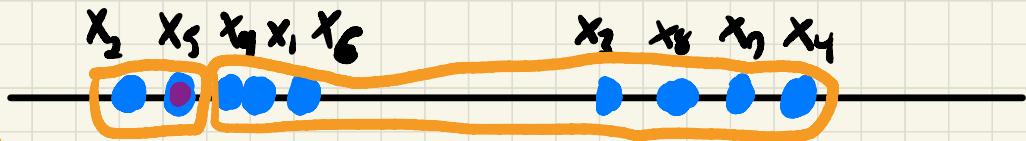
$$= \sum_{p \in C_2} \delta(x_5, p) / |C_2|$$

$$= \underbrace{\delta(x_5, x_1) + \delta(x_5, x_3) + \dots + \delta(x_5, x_9)}_7$$

$$= 7.64$$

SC Example

	X_1
X_1	4
X_2	1.1
X_3	12
X_4	16.4
X_5	2.3
X_6	5
X_7	15
X_8	13.7
X_9	3.5



$$C = \{ C_1 = \{x_2, x_5\},$$

$$C_2 = \{x_1, x_3, x_4, x_6, x_7, x_8, x_9\}$$

$$\mu_{in}(x_5) = \frac{\sum_{x_i \in C_1} \delta(x_5, x_i)}{|C_1| - 1}$$

$$= \frac{\delta(x_5, x_2)}{2 - 1} = 1.2$$

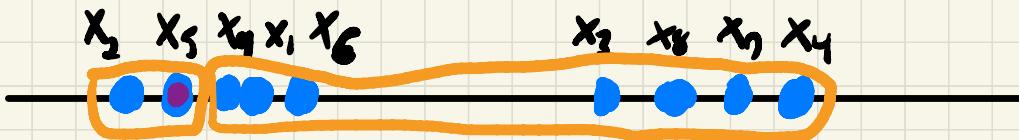
$$s_i = \frac{\mu_{out}^{\min}(x_i) - \mu_{in}(x_i)}{\max\{\mu_{out}^{\min}(x_i), \mu_{in}(x_i)\}}$$

$$s_5 = \frac{\mu_{out}^{\min}(x_5) - \mu_{in}(x_5)}{\max\{\mu_{out}^{\min}(x_5), \mu_{in}(x_5)\}}$$

1.64 ↗

SC Example

	X ₁
X ₁	4
X ₂	1.1
X ₃	12
X ₄	16.4
X ₅	2.3
X ₆	5
X ₇	15
X ₈	13.7
X ₉	3.5



$$C = \{ C_1 = \{x_2, x_5\},$$

$$C_2 = \{x_1, x_3, x_4, x_6, x_7, x_8, x_9\} \}$$

$$s_i = \frac{\mu_{out}^{\min}(x_i) - \mu_{in}^{\min}(x_i)}{\max\{\mu_{out}^{\min}(x_i), \mu_{in}^{\min}(x_i)\}}$$

$$s_5 = \frac{7.4 - 1.2}{\max(7.4, 1.2)} = .84$$

$$s_5 = \frac{\mu_{out}^{\min}(x_5) - \mu_{in}^{\min}(x_5)}{\max\{\mu_{out}^{\min}(x_5), \mu_{in}^{\min}(x_5)\}}$$

SC Example: S_3

	X_1
x_1	4
x_2	1.1
x_3	12
x_4	16.4
x_5	2.3
x_6	5
x_7	15
x_8	13.7
x_9	3.5



$$C = \{ C_1 = \{x_2, x_5\},$$

$$C_2 = \{x_1, x_3, x_4, x_6, x_7, x_8, x_9\} \}$$

$$S_i = \frac{\mu_{out}^{\min}(x_i) - \mu_{in}(x_i)}{\max\{\mu_{out}^{\min}(x_i), \mu_{in}(x_i)\}}$$

$$S_3 = \frac{\mu_{out}^{\min}(x_3) - \mu_{in}(x_3)}{\max\{\mu_{out}^{\min}(x_3), \mu_{in}(x_3)\}}$$

$$\mu_{out}^{\min}(x_3) = \min_{j \neq i} \left\{ \frac{\sum_{p \in C_j} S(x_3, p)}{|C_j|} \right\}$$

$$= \frac{\sum_{p \in C_1} S(x_3, p)}{|C_1|} = \frac{S(x_3, x_2) + S(x_3, x_5)}{2} = 10.3$$

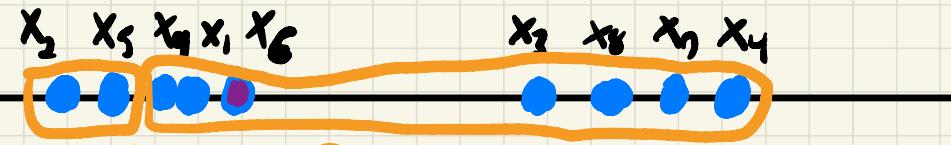
$$\mu_{in}(x_3) = \frac{\sum_{x_i \in C_2} S(x_3, x_i)}{|C_2| - 1}$$

$$= \frac{S(x_3, x_9) + S(x_3, x_1) + \dots + S(x_3, x_7)}{6} = 5.43$$

$$S_3 = \frac{10.3 - 5.43}{\max(10.3, 5.43)} = .47$$

SC Example: S_6

X ₁	4
X ₂	1.1
X ₃	12
X ₄	16.4
X ₅	2.3
X ₆	5
X ₇	15
X ₈	13.7
X ₉	3.5



$$C = \{ C_1 = \{x_2, x_5\},$$

$$C_2 = \{x_1, x_3, x_4, x_6, x_7, x_8, x_9\} \}$$

$$S_i = \frac{\mu_{out}(x_i) - \mu_{in}(x_i)}{\max\{\mu_{out}(x_i), \mu_{in}(x_i)\}}$$

$$S_6 = \frac{\mu_{out}(x_6) - \mu_{in}(x_6)}{\max\{\mu_{out}(x_6), \mu_{in}(x_6)\}}$$

$$\begin{aligned} \mu_{out}(x_6) &= \min_{j \neq 6} \left\{ \frac{\sum_{p \in C_j} \delta(x_6, p)}{|C_j|} \right\} \\ &= \frac{\sum_{p \in C_1} \delta(x_6, p)}{|C_1|} = \frac{\delta(x_6, x_2) + \delta(x_6, x_5)}{2} = 3.3 \end{aligned}$$

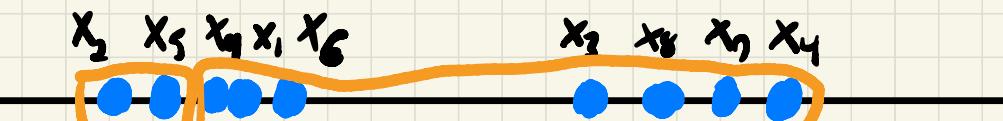
$$\mu_{in}(x_6) = \frac{\sum_{x_i \in C_2} \delta(x_6, x_i)}{|C_2| - 1}$$

$$= \frac{\delta(x_6, x_9) + \delta(x_6, x_1) + \dots + \delta(x_6, x_4)}{6} = 6.6$$

$$S_6 = \frac{3.3 - 6.6}{\max(3.3, 6.6)} = -0.5$$

SC Example

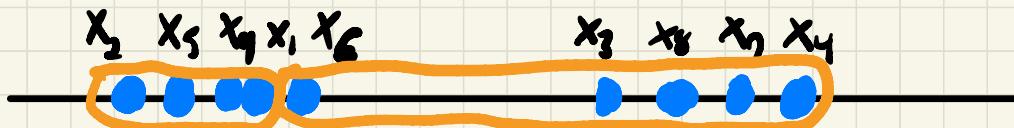
	X ₁
X ₁	4
X ₂	1.1
X ₃	12
X ₄	16.4
X ₅	2.3
X ₆	5
X ₇	15
X ₈	13.7
X ₉	3.5



$$\mathcal{C} = \left\{ C_1 = \{x_2, x_5\}, C_2 = \{x_1, x_3, x_4, x_6, x_7, x_8, x_9\} \right\}$$

$$SC = .2$$

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



$$SC = .57$$



$$SC = .8$$