Silhouette Coefficient

Cluster Validity

- Extrinsic – ground truth is available

    Examples

Cluster Validity

Extrinsic - ground truth is available

Examples: Precision, Recall, F-Score

Intrinsic - ground truth is not available

Silhouette Coeff

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

Let

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

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

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

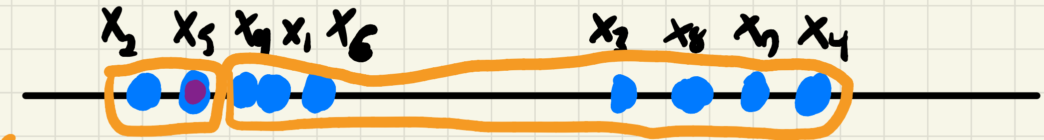
$$SC = \frac{1}{n} \sum_{i=1}^{n} s_i$$

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

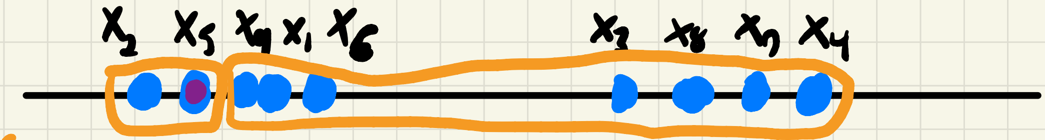


$$\mathcal{C} = \{ C_1 = \{x_2, x_5\},$$
$$C_2 = \{x_1, x_3, x_4, x_6, x_7, x_8, x_9\}\}$$

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

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



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

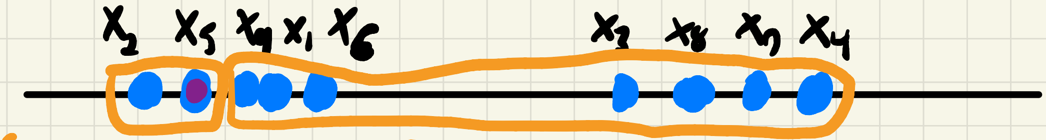
$$s_i = \frac{\mu_{out}^{min}(x_i) - \mu_{in}(x_i)}{\max\{\mu_{out}^{min}(x_i), \mu_{in}(x_i)\}}$$

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

$$\mu_{in}(x_5) =$$

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



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

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

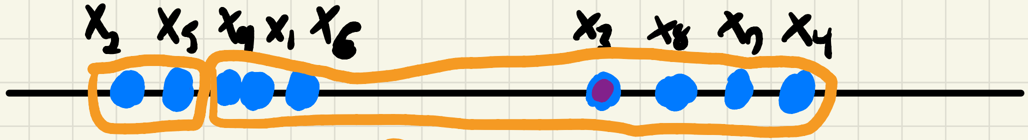
7.64

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

s_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



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

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

$$s_3 = \frac{\mu_{out}^{min}(x_3) - \mu_{in}(x_3)}{\max\{\mu_{out}^{min}(x_3), \mu_{in}(x_3)\}}$$

$$\mu_{out}^{min}(x_3) = \min_{j \neq \hat{y}_3} \left\{ \frac{\sum_{p \in C_j} \delta(x_3, p)}{|C_j|} \right\}$$

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

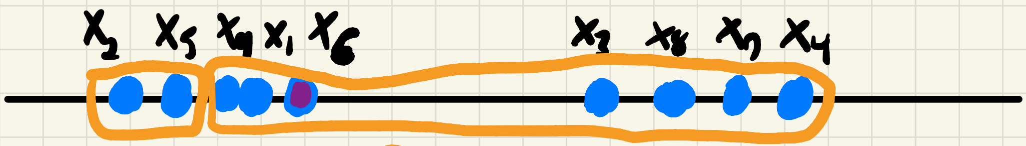
$$\mu_{in}(x_3) = \frac{\sum_{x_j \in C_2} \delta(x_3, x_j)}{|C_2| - 1}$$

$$= \frac{\delta(x_3, x_9) + \delta(x_3, x_1) + \cdots + \delta(x_3, x_4)}{6} = 5.43$$

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

SC Example: s_6

	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



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

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

$$s_6 = \frac{\mu_{out}^{min}(x_6) - \mu_{in}(x_6)}{\max\{\mu_{out}^{min}(x_6), \mu_{in}(x_6)\}}$$

$$\mu_{out}^{min}(x_6) = \min_{j \neq \hat{y}_6} \left\{ \frac{\sum_{p \in C_j} \delta(x_6, p)}{|C_j|} \right\}$$

$$= \frac{\sum_{p \in C_1} \delta(x_6, p)}{|C_1|} = \frac{\delta(x_6, x_2) + \delta(x_6, x_5)}{2} = 3.3$$

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

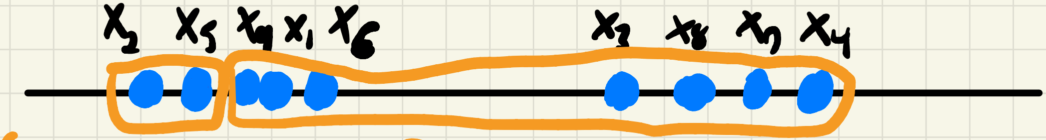
$$= \frac{\delta(x_6, x_9) + \delta(x_6, x_1) + \cdots + \delta(x_6, x_4)}{6} = 6.6$$

$$S_3 = \frac{3.3 - 6.6}{\max(3.3, 6.6)} = -0.5$$

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

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



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

$SC = .2$

$SC = .57$

$SC = .8$