

Cluster Validity

August 25, 2014

1 指標 1

$$WCD = \frac{\sum_k^K \frac{\sum_i^{I_k} \min_j(d(x_{ki}, x_{kj}))}{I_k}}{K} \quad (1)$$

$$BCD = \frac{\sum_k^K \min_{ilj}(d(x_{ki}, x_{lj}))}{K} \quad (2)$$

$$\min1 = \frac{WCD}{BCD} = \frac{\sum_k^K \frac{\sum_i^{I_k} \min_j(d(x_{ki}, x_{kj}))}{I_k}}{\sum_k^K \min_{ilj}(d(x_{ki}, x_{lj}))} \quad (3)$$

ただし、クラス数: K 、クラス ID: k または l 、サンプル ID: x_{ki} または x_{kj} 、クラス k のサンプル数: I_k 、 $f(j)$ に関する最小値: $\min_j(f(j))$ 、距離関数: $d(x, y)$ 。
。

2 指標 2

$$WCD = \frac{\sum_k^K \frac{1}{I_k} \sum_i^{I_k} \min_j(d(x_{ki}, x_{kj}))}{K} \quad (4)$$

$$WCS = \frac{1}{K} \frac{\sum_k^K \frac{1}{I_k} \sum_i^{I_k} \min_j(d(x_{ki}, x_{kj}))}{K} \quad (5)$$

$$BCD = \frac{\sum_k^K \min_{ilj}(d(x_{ki}, x_{lj}))}{K} \quad (6)$$

$$BCS = \frac{1}{K} \frac{\sum_k^K \min_{ilj}(d(x_{ki}, x_{lj}))}{K} \quad (7)$$

$$\min2 = WCS.BCS = \frac{1}{K} \frac{\sum_k^K \frac{1}{I_k} \sum_i^{I_k} \min_j(d(x_{ki}, x_{kj}))}{K} + \frac{\sum_k^K \min_{ilj}(d(x_{ki}, x_{lj}))}{K} \quad (8)$$

ただし、クラス数: K 、クラス ID: k または l 、サンプル ID: x_{ki} または x_{kj} 、クラス k のサンプル数: I_k 、 $f(j)$ に関する最小値: $\min_j(f(i))$ 、距離関数: $d(x, y)$ 。
。

3 指標 3

$$\min 3 = \sum_k^K \frac{\sum_i^{I_k} \min_j(d(x_{ki}, x_{kj}))}{I_k} \quad (9)$$

ただし、クラスタ数: K 、クラスタ ID: k または l 、サンプル ID: x_{ki} または x_{kj} 、クラスタ k のサンプル数: I_k 、 $f(j)$ に関する最小値: $\min_j(f(j))$ 、距離関数: $d(x, y)$ 。

4 指標 4

$$\min 4 = \frac{\sum_k^K \frac{1}{I_k} \sum_i^{I_k} \min_j(d(x_{ki}, x_{kj}))}{K} \quad (10)$$

ただし、クラスタ数: K 、クラスタ ID: k または l 、サンプル ID: x_{ki} または x_{kj} 、クラスタ k のサンプル数: I_k 、 $f(j)$ に関する最小値: $\min_j(f(j))$ 、距離関数: $d(x, y)$ 。

5 指標 5

$$\min 5 = \frac{\sum_k^K \sum_i^{I_k} \min_j(d(x_{ki}, x_{kj}))}{K} \quad (11)$$

ただし、クラスタ数: K 、クラスタ ID: k または l 、サンプル ID: x_{ki} または x_{kj} 、クラスタ k のサンプル数: I_k 、 $f(j)$ に関する最小値: $\min_j(f(j))$ 、距離関数: $d(x, y)$ 。

6 指標 5(mst1)

$$\text{mst1} = \sum_{j=1}^k \frac{\sum_{i=1}^{n_j} p_{ij}}{n_j} \quad (12)$$

ただし、 p_{ij} : クラスタ j の i 番目のパスの長さ (MST による)、 n_j : クラスタ j のパス数。

7 指標 6(mst2)

$$B = \frac{\sum_{m,n}^k \text{dmst}(z_m, z_n)}{P} \quad (13)$$

$$\text{mst2} = \frac{\text{mst1}}{B} \quad (14)$$

ただし、 z_i : クラスタ i の重心にもっとも近いデータポイント、 P : z により構成される MST の総パス数。

8 指標 7(mst3)

$$mst3 = \frac{\sum_{i,j; i \neq j; i > j}^k \frac{\sum_s^{mst(C_i \cup C_j).PATH} mst(C_i \cup C_j).Ps}{mst(C_i \cup C_j).PATH}}{{}_k C_2} \quad (15)$$

ただし、 k : クラスタ数、 C_i : クラスタ i 、 $C_i \cup C_j$: クラスタ i と クラスタ j の要素の和集合、 $mst(C_i \cup C_j)$: 集合全体による MST、 $mst().PATH$: パスの総数、 $mst().Ps$: MST における s 番目のパス、 ${}_k C_2$: コンビネーション $\{k, 2\}$ 。