

Fuzzy C-means and Fuzzy Linear Clustering

Fuzzy C-means:

Given:

$$D = \{p_1, p_2, \dots, p_n\} \text{ and } K \geq 2$$

$$p_i = \langle x_{i1}, x_{i2}, \dots, x_{is} \rangle : 1 \leq i \leq n$$

Find:

- * K fuzzy clusters A_j , $1 \leq j \leq K$,
Cluster center c_j for each cluster A_j ,
- * Membership values W_{ij} (≥ 0) for each point p_i and cluster A_j .

Minimize:

$$J_m = \sum_{j=1}^K \sum_{i=1}^n (W_{ij})^m \|p_i - c_j\|^2, \quad m > 1 \quad (1)$$

subject to $\sum_{j=1}^K W_{ij} = 1$, $1 \leq i \leq n$,

and $0 < \sum_{i=1}^n W_{ij} < n$ for $1 \leq j \leq K$.

Basic FCM Algorithm

Step 1.

Initialize weight matrix W (which is a n by K matrix).

Step 2.

Repeat

2.1

Compute centroids c_1, \dots, c_K from W and D .

2.2

Re-compute W from c_1, \dots, c_K and D .

Until convergence

The equations for determining W_{ij} and c_j :

$$W_{ij} = \frac{[\|p_i - c_j\|^2]^{-1/(m-1)}}{\sum_{l=1}^K [\|p_i - c_l\|^2]^{-1/(m-1)}} \quad (2)$$

$1 \leq j \leq K$ and $1 \leq i \leq n$.

$$c_j = \frac{\sum_{i=1}^n (W_{ij})^m p_i}{\sum_{i=1}^n (W_{ij})^m}. \quad (3)$$

Fuzzy Linear Clustering:

Given:

$$D = \{p_1, p_2, \dots, p_n\} \text{ and } K \geq 2$$

$$p_i = \langle x_i, y_i \rangle$$

$$x_i = \langle x_{i1}, x_{i2}, \dots, x_{is} \rangle$$

Find:

- * K fuzzy linear clusters A_j , $1 \leq j \leq K$,
Cluster center characterized by

$$g_j(x) = a_{j0} + a_{j1}x_1 + \dots + a_{js}x_s.$$
- * Membership values W_{ij} (≥ 0) for each point p_i and cluster A_j .

Minimize:

$$\Phi_m = \sum_{j=1}^K \sum_{i=1}^n (W_{ij})^m [y_i - g_j(x_i)]^2, \quad m > 1. \quad (4)$$

$$\text{subject to } \sum_{j=1}^K W_{ij} = 1, \quad 1 \leq i \leq n,$$

$$\text{and } 0 < \sum_{i=1}^n W_{ij} < n \text{ for } 1 \leq j \leq K.$$

Compared with Fuzzy C-means:

- Cluster center c_j for the cluster A_j in C-means
- Cluster center $g_j(x)$, a linear equation for cluster A_j in fuzzy linear clustering
- Distance $\|p_i - c_j\|$ in C-means
- Distance $|y_i - g_j(x_i)|$ in fuzzy linear clustering

- Computation of g_j coefficients:
Linear regression from current membership values W_{ij}
- Computation of membership values W_{ij} :

$$W_{ij} = \frac{[(y_i - g_j(x_i))^2]^{-1/(m-1)}}{\sum_{l=1}^K [(y_i - g_l(x_i))^2]^{-1/(m-1)}} \quad (5)$$

$1 \leq j \leq K$ and $1 \leq i \leq n.$

- Previous studies show that fuzzy linear clustering has strong generalization capability for function approximation.