Basic Structure of Fuzzy Cluster

Fuzzy cluster was introduced by Prof. L.A. Zadeh from Berkeley in 1965. The theory of fuzzy cluster defines the degree in which the elements x of the cluster X is in the fuzzy cluster A or in another formula.

$$A = \{(x, \mu(x)); x \in X\}$$

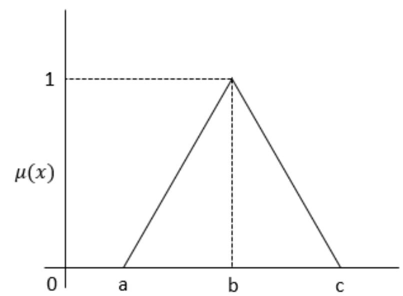
Above formula represents the fuzzy cluster X where μ_A is the membership function, $\mu_A(x)$ is the membership function of x in fuzzy cluster A (Kirschfink 1999).

There are several types of membership function used, they are:

1. Trimf

   Trimf function is used to create a membership function with triangle curve. There are three parameters that can be used on Trimf curve, namely $a, b,$ and c.

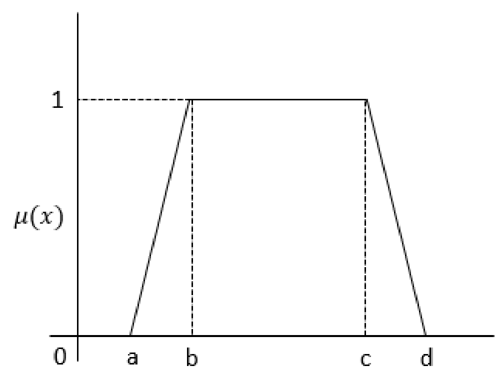
$$f(a,b,c) = \begin{cases} 0, & x \leq a \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ \dfrac{c-x}{x-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases}$$



2. Trapmf

   Trapmf function is used to create a membership function with trapezoid curve. There are four parameters that can be used on Trapmf curve, namely $a, b, c,$ and d.

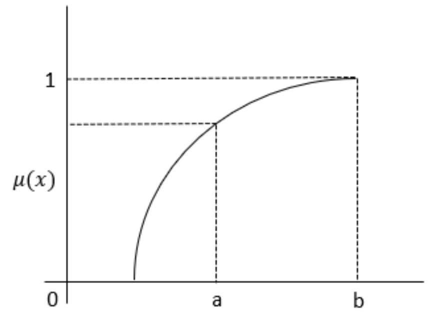
$$f(a,b,c,d) = \begin{cases} 0, & x \leq a \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \dfrac{d-x}{x-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases}$$



3. Smf

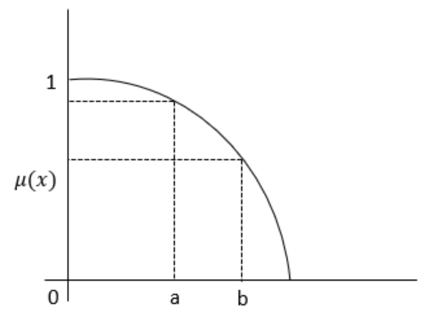
   Smf function is used to create a membership function with S curve. There are two parameters that can be used on S curve, namely a and b.

$$
f(a,b) = \begin{cases} 0, & x \leq a \\ 2\left(\dfrac{x-a}{b-a}\right)^2, & a \leq x \leq \dfrac{(a+b)}{2} \\ 1 - 2\left(\dfrac{b-x}{b-a}\right)^2, & \dfrac{(a+b)}{2} \leq x \leq b \\ 1, & x \geq b \end{cases}
$$



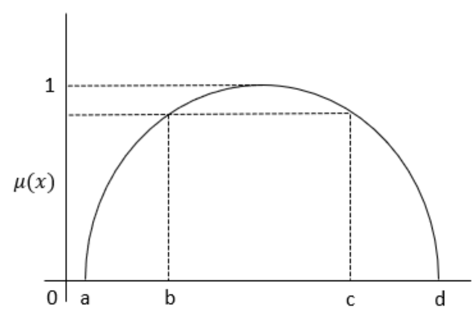
4. Zmf
   Zmf function is used to create a membership function with Z curve. There are two parameters that can be used on Z curve, namely a and b.

$$
f(a,b) = \begin{cases} 1, & x \leq a \\ 2\left(\dfrac{x-a}{b-a}\right)^2, & a \leq x \leq \dfrac{(a+b)}{2} \\ 1 - 2\left(\dfrac{b-x}{b-a}\right)^2, & \dfrac{(a+b)}{2} \leq x \leq b \\ 0, & x \geq b \end{cases}
$$



5. Pimf
   Pimf function is used to create a membership function with Pi curve. There are four parameters that can be used on Trapmf curve, namely $a, b, c,$ and d.

$$
f(a,b,c,d) = smf(a,b) \cdot zmf(c,d)
$$



Fuzzy c-means

According to Jang *et al.* (1997), fuzzy c-means is the clustering algorithm where each data points in clusters are marked by the degree of membership. It was modified by Jim Bezdek in 1973 in hard clustering techniques. The fuzzy c-means distributes the n data vertices of x_j $(i = 1, 2, 3, \dots, m)$ to c clusters and find the cluster centroids for each cluster which minimizes the dissimilarity of objective function. Step by step of fuzzy c-means clustering algorithm is as follows.

1. Create a matrix nxm where n is the number of data points and m is the number of attributes or columns. Thus, x_{kj} is the data point for k and j in which $k = 1, 2, 3, \dots, n$ and $j = 1, 2, 3, \dots, m$.

2. Determine parameters, they are c as the number of clusters, m as the *fuzzifier*, *maxIter* as the maximum iteration, ξ as the smallest tolerance error expected, $P_0 = 0$ as the initial objective function, and $t = 1$ as the initial iteration

3. Generate random numbers μ_{ik} $(i = 1, 2, 3, \dots, c \ and \ k = 1, 2, 3, \dots, n)$ as the elements of initial partition matrix U

$$U_0 = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \cdots & \mu_{1c}(x_c) \\ \vdots & \vdots & & \vdots \\ \mu_{n1}(x_1) & \mu_{n2}(x_2) & \cdots & \mu_{nc}(x_c) \end{bmatrix}$$

The partition matrix on fuzzy clustering must be fulfilled the following conditions

$$\mu_{ik} = [0,1], \qquad (1 \leq i \leq c; 1 \leq k \leq n)$$

$$\sum_{i=1}^{n} \mu_{ik} = 1\,, \qquad 1 \leq i \leq c$$

$$0 < \sum_{i=1}^{c} \mu_{ik} < c\,, \qquad 1 \leq k \leq n$$

Calculate the total of each attribute or columns

$$Q_j = \sum_{i=1}^{c} \mu_{ik}$$

where $j = 1, 2, 3, \dots, m$

Afterwards, calculate

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j}$$

4. Calculate k cluster centroids $\left(V_{ij}\right)$ where $i \ = 1, 2, 3, \dots, c$ and $j \ = 1, 2, 3, \dots, m$

$$V_{ij} = \frac{\sum_{k=1}^{n}\left((\mu_{ik})^m \cdot X_{kj}\right)}{\sum_{k=1}^{n}(\mu_{ik})^m} \text{ where } V = \begin{bmatrix} V_{11} & \cdots & V_{1m} \\ \vdots & \ddots & \vdots \\ V_{c1} & \cdots & V_{cm} \end{bmatrix}$$

5. Calculate the objective function on iteration t as $J(P, U, X, c, m)$ using the following formula

$$P_t = \sum_{k=1}^{n} \sum_{i=1}^{c} \left(\left[\sum_{j=1}^{m} \left(X_{kj} - V_{ij} \right)^2 \right] (\mu_{ik})^m \right)$$

6. Recalculate the partition matrix U using the following formula

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{p} (X_{kj} - V_{ij})^2 \right]^{\frac{-1}{p-1}}}{\sum_{i=1}^{c} \left[\sum_{j=1}^{p} (X_{kj} - V_{ij})^2 \right]^{\frac{-1}{p-1}}}$$

7. Evaluate the stop condition
   a. If $P_t < \xi$ or $t > maxIter$, then the iteration stops
   b. If condition in a) is not fulfilled, then step back to 6)