Fuzzy c-means clustering

Recall the membership function $c(i,j)$:

$$c(i,j) = \begin{cases} 1 & x_i \text{ belongs to class } j \\ 0 & x_i \text{ does not belong to class } j \end{cases}$$

Observe that it allows cases in which x_i may belong to more than one class. It also allows for a "fuzzy" description, where $0 \leq c(i,j) \leq 1$, and for a probabilistic description. We impose the conditions: $c(i,j) \geq 0$, and for all i, $\sum_j c(i,j) = 1$. Here the value of $c(i,j)$ can be interpreted as the likelihood that x_i belongs to class j.
The following error criterion generalizes the k-means error criterion:

$$J = \sum_{i=1}^{m} \sum_{j=1}^{k} c(i,j)^p \|x_i - u_j\|^2 \tag{0}$$

where $c(i,j)$ is the fuzzy degree of membership of x_i in cluster j and p is a non-negative constant. (Note: the algorithm presented here fails for the case of $p = 1$.) Similar to the case of k-means the explicit minimization of the J is not known, but it can be minimized iteratively. Given all the values of $c(i,j)$ we can compute the vectors u_j, and from the vectors u_j we can compute all the $c(i,j)$. The explicit formulas are given below. The proofs are in the next page.

Given the $c(i,j)$ the u_j are computed as follows:

$$\text{for all } j: \quad u_j = \frac{\sum_{i=1}^{m} c(i,j)^p x_i}{\sum_{i=1}^{m} c(i,j)^p} \tag{1}$$

Observe that u_j is a weighted mean of the x_i.
Given all the u_j the values of $c(i,j)$ are computed as follows:

$$\begin{aligned} & d(i,j) = \|x_i - u_j\|^2 \quad \text{for } i = 1,\ldots,m, \quad j = 1,\ldots,k \\ & \text{Special cases: if } d(i^*,j^*) = 0 \text{ then } c(i^*,j^*) = 1, \text{ and } c(i^*,t) = 0 \text{ for } t \neq j^* \\ & \text{Otherwise:} \quad c(i,j) = \frac{1/d_{ij}^{1/(p-1)}}{\sum_{t=1}^{k} 1/d_{it}^{1/(p-1)}} \end{aligned} \tag{2}$$

Just as in the case of k-means the fuzzy c-means is applied to the data by repeating steps 1,2 (or 2,1). It is guaranteed to converge to a local minimum of J.
The value of p must be determined experimentally. A commonly used value is $p = 2$.

Proofs

The formula (1) is obtained by computing the u_j as the minimizers of (0). Taking the derivative (gradient) of (1) with respect to u_j and equating to 0 we get:

$$\frac{\partial J}{\partial u_i} = \frac{\partial(\sum_{i=1}^{m} c(i,j)^p \|x_i - u_j\|^2)}{\partial u_i} = \sum_{i=1}^{m} c(i,j)^p(-2x_i + 2u_j) = 0$$

This gives:

$$\sum_{i=1}^{m} c(i,j)^p x_i = u_j \sum_{i=1}^{m} c(i,j)^p$$

which is equivalent to (1). To simplify the notation in the proof of (2) we write: $c_{ij} = c(i,j)$, $d_{ij} = \|x_i - u_j|^2$. Solving for c_{ij} requires solving the following:

$$\text{minimize } J = \sum_{i=1}^{m}\sum_{j=1}^{k} c_{ij}^p d_{ij} \quad \text{subject to} \quad \sum_{t=1}^{k} c_{it} = 1, \quad \text{for } i = 1,\ldots,m$$

Observe that there are no constraints that relate c_{i_1,j_1} to c_{i_2,j_2} if $i_1 \neq i_2$. This means that we can solve for c_{ij}, $j = 1\ldots k$ by minimizing separately for each i the following expression:

$$\text{minimize } J_i = \sum_{j=1}^{k} c_{ij}^p d_{ij} \quad \text{subject to} \quad \sum_{t=1}^{k} c_{it} = 1$$

For the special where $d(i^*, j^*) = 0$ the solution $c(i^*, j^*) = 1$, and $c(i^*, t) = 0$ for $t \neq j^*$ gives $J_{i^*} = 0$, which is clearly optimal. For the non special cases we use the method of Lagrange multipliers. The Lagrangian is:

$$L_i = J_i - \lambda_i(\sum_{t=1}^{k} c_{it} - 1)$$

We have:

$$\frac{\partial L_i}{\partial c_{ij}} = d_{ij} \cdot p \cdot c_{ij}^{p-1} - \lambda_i$$

Equating to 0 and solving for c_{ij} in terms of the Lagrange multiplier λ_i we get:

$$c_{ij} = \left(\frac{\lambda_i}{p}\right)^{1/(p-1)} \cdot \frac{1}{d_{ij}^{1/(p-1)}} \tag{3}$$

Substituting this value into the ith constraint $\sum_{t=1}^{k} c_{it} = 1$ we get:

$$\left(\frac{\lambda_i}{p}\right)^{1/(p-1)} = \frac{1}{\sum_{t=1}^{k}\left(\frac{1}{d_{it}}\right)^{1/(p-1)}}$$

Substituting this in (3) gives the desired result.