Generalized Fuzzy c-means Algorithms

Nicolaos B. Karayiannis
Department of Electrical & Computer Engineering
University of Houston
Houston, Texas 77204-4793, USA
Karayiannis@UH.EDU

Abstract

This paper proposes Generalized fuzzy c-means (FCM) algorithms. The clustering problem is formulated as a constrained minimization problem, whose solution depends on the selection of a constraint function that satisfies certain conditions. If the constraint function is proportional to the generalized mean of the membership values, the solution of this minimization problem results in a broad family of Generalized FCM algorithms. The existing FCM algorithm can be obtained as a special case of the proposed formulation if the generalized mean coincides with the arithmetic mean. Other special cases include the Minimum FCM and the Geometric FCM. The proposed formulation also assigns to each feature vector a parameter that can be used to measure the certainty of its assignment into various clusters. The reliability of this certainty measure is verified by experiments involving an artificial data set containing outliers.

1. Introduction

Clustering refers to a broad spectrum of methods which subdivide a random data set into subsets, or clusters, which are pairwise disjoint, all nonempty, and reproduce the original data set via union.

Clustering is widely performed by an iterative algorithm which is known as the crisp c-means algorithm [1]. The crisp c-means algorithm assigns each feature vector to a single cluster and ignores the possibility that this feature vector may also belong to other clusters. As a result, for a given initial set of prototypes, the algorithm finds the nearest local minimum in the space of all possible c-partitions.

Fuzzy clustering algorithms consider each cluster as a fuzzy set, while a membership function measures the possibility that each feature vector belongs to a cluster. As a result, each feature vector may be assigned to multiple clusters with some degree of certainty measured by the membership function. Bezdek [1] used alternating optimization to develop the fuzzy c-means (FCM) algorithm, which became a powerful tool for clustering applications.

Krishnapuram and Keller [5] argued that if the feature space contains outliers the membership values produced by the FCM do not represent a "degree of belonging" and also fail to distinguish between a "typical" member of a cluster and an "extremely atypical" member of a cluster. They also developed the possibilistic c-means (PCM) algorithm by eliminating one of the constraints imposed on the search for c-partitions [5]. This paper presents a broad family of Generalized FCM algorithms, which include the existing FCM as a special case. The proposed approach formulates clustering as a constrained minimization problem but relaxes the constraints imposed on the search for c-partitions.

2. Fuzzy c-means Algorithms

Consider the set \( \mathcal{X} \) formed by \( M \) feature vectors from an \( n \)-dimensional Euclidean space, that is, \( \mathcal{X} = \{x_1, x_2, \ldots, x_M\} \), \( \mathcal{X} \subset \mathbb{R}^n \). The clustering process is based on the assignment of the feature vectors \( x_i \in \mathcal{X} \) into \( c \) clusters, which are represented by the prototypes \( v_j \in \mathcal{V} \subset \mathbb{R}^n \). The certainty of the assignment of the feature vector \( x_i \in \mathcal{X} \) to the \( j \)th cluster is measured by the membership function \( u_{ij} = u_j(x_i) \). In fuzzy c-partitions the membership \( u_{ij} \) is allowed to take values from the interval \([0, 1]\). The \( c \times n \) matrix \( U = [u_{ij}] \) is a fuzzy c-partition in the set \( \mathcal{U}_c \) defined as

\[
\mathcal{U}_c = \left\{ U \in \mathbb{R}^{cn} \mid u_{ij} \in [0, 1] \forall i, j; \sum_{j=1}^{c} u_{ij} = 1 \forall i \right\}
\]
be seen as the reference point used for computing the membership values \( u_{ij}, j = 1, 2, \ldots, c \).

The previous discussion indicates that the search for alternative clustering algorithms involves the search for points of reference other than the harmonic mean \( \langle d(x_i, \cdot) \rangle_H \). This search can be facilitated by the well-known inequality
\[
\langle d(x_i, \cdot) \rangle_{\min} \leq \langle d(x_i, \cdot) \rangle_H \leq \langle d(x_i, \cdot) \rangle_G \leq \langle d(x_i, \cdot) \rangle_A,
\]
where \( \langle d(x_i, \cdot) \rangle_{\min} \) is the minimum of \( d(x_i, v_j), j = 1, 2, \ldots, c \), defined as
\[
\langle d(x_i, \cdot) \rangle_{\min} = \min_{1 \leq j \leq c} d(x_i, v_j),
\]
\( \langle d(x_i, \cdot) \rangle_G \) is the geometric mean of \( d(x_i, v_j), j = 1, 2, \ldots, c \), defined as
\[
\langle d(x_i, \cdot) \rangle_G = \left( \prod_{j=1}^{c} d(x_i, v_j) \right)^{\frac{1}{c}},
\]
and \( \langle d(x_i, \cdot) \rangle_A \) is the arithmetic mean of \( d(x_i, v_j), j = 1, 2, \ldots, c \), defined as
\[
\langle d(x_i, \cdot) \rangle_A = \frac{1}{c} \sum_{j=1}^{c} d(x_i, v_j).
\]

3. Searching for Alternative c-Partitions

Consider the FCM algorithm which corresponds to \( m = 2 \). In this case the membership function is given by
\[
 u_{ij} = \left( \sum_{i=1}^{c} \frac{d(x_i, v_j)}{d(x_i, v_j)} \right)^{-1} = \frac{1}{d(x_i, v_j)} \left( \sum_{i=1}^{c} \frac{1}{d(x_i, v_j)} \right)^{-1}
\]
Bezdek suggested in [1] that the membership function corresponding to \( m = 2 \) relates to the harmonic mean \( \langle d(x_i, \cdot) \rangle_H \) of the distances \( d(x_i, v_j), j = 1, 2, \ldots, c \), which is defined by
\[
\frac{1}{\langle d(x_i, \cdot) \rangle_H} = \frac{1}{c} \sum_{j=1}^{c} \frac{1}{d(x_i, v_j)}.
\]
Clearly, the comparison between (5) and (6) results in
\[
u_{ij} = \frac{1}{c} \langle d(x_i, \cdot) \rangle_H.
\]
According to (7), the membership value \( u_{ij} \) assigned to the feature vector \( x_i \) depends on its distance from the corresponding prototype \( v_j \) relative to the harmonic mean \( \langle d(x_i, \cdot) \rangle_H \) of its distances from all \( c \) prototypes. If \( d(x_i, v_j) \leq \langle d(x_i, \cdot) \rangle_H \), (7) indicates that \( u_{ij} \geq \frac{1}{c} \).
Conversely, if \( d(x_i, v_j) > \langle d(x_i, \cdot) \rangle_H \), then \( u_{ij} < \frac{1}{c} \). Note that \( \frac{1}{c} \) is the membership value corresponding to the maximally fuzzy partition, where \( u_{ij} = \frac{1}{c} \). Thus, the harmonic mean \( \langle d(x_i, \cdot) \rangle_H \), can be seen as the reference point used for computing the membership values \( u_{ij}, j = 1, 2, \ldots, c \).

4. Generalized Fuzzy c-means Algorithms

A family of Generalized FCM algorithms can be developed by considering the solution of the minimization
problem
\[ \min_{U \in \mathbb{R}^{c \times n}} \left\{ \sum_{i=1}^{M} \sum_{j=1}^{c} (u_{ij})^m d(x_i, v_j) \right\}, \] (14)
where \( 1 < m < \infty \) and the \( c \times n \) matrix \( U = [u_{ij}] \) is a \( c \)-partition in the set \( \mathcal{U}_0 \) defined as
\[ \mathcal{U}_0 = \left\{ U \in \mathbb{R}^{c \times n} \mid u_{ij} \geq 0 \forall i, j; \sum_{j=1}^{c} u_{ij} = \theta_i \forall i; 0 < \sum_{i=1}^{M} u_{ij} < \infty \forall j \right\} \] (15)
and \( \theta_i > 0, \forall i = 1, 2, \ldots, M \). For any set of parameters \( \theta_i, 1 \leq i \leq M \), the solution of the minimization problem (14) satisfies the conditions \( 0 < \sum_{j=1}^{c} u_{ij} = \theta_i \forall i \).
Instead of directly specifying the parameters \( \theta_i, 1 \leq i \leq M \), this approach searches for solutions in \( \mathcal{U}_0 \) which also satisfy an additional constraint of the form
\[ F(u_{ij}, \forall j \in \mathcal{C}) = 1 \forall i \leq i \leq M, \] (16)
where \( \mathcal{C} = \{1, 2, \ldots, c\} \) and \( F(.\) is a function which satisfies the condition:
\[ F(\alpha u_{ij}, \forall j \in \mathcal{C}) = \alpha F(u_{ij}, \forall j \in \mathcal{C}) \forall \alpha > 0. \] (17)
The solution of this minimization problem is given by the following Theorem [2]:

**Theorem:** Consider the minimization problem (14) with \( d(x_i, v_j) = ||x_i - v_j||^2 \). Define the sets \( \mathcal{I}_i = \{j \mid 1 \leq j \leq c; d(x_i, v_j) = 0\} \) and \( \mathcal{I}_i = \{1, 2, \ldots, c\} - \mathcal{I}_i \). The necessary conditions for solutions (U, V) \( \in \mathcal{U}_0 \times \mathbb{R}^{c \times n} \) of (14) under the constraint (16) are:
- if \( \mathcal{I}_i \neq \emptyset \), then \( u_{ij} = 0 \forall j \in \mathcal{I}_i \) and \( F(u_{ij}, \forall j \in \mathcal{I}_i) = 1 \forall i \);
- or if \( \mathcal{I}_i = \emptyset \), then
\[ u_{ij} = \frac{d(x_i, v_j)^{1/m}}{F(d(x_i, v_j)^{1/m}, \forall j \in \mathcal{C})}, 1 \leq i \leq M; 1 \leq j \leq c, \] (18)
and
\[ v_j = \frac{\sum_{i=1}^{M} (u_{ij})^m x_i}{\sum_{i=1}^{M} (u_{ij})^m}, 1 \leq j \leq c. \] (19)
If \( \mathcal{I}_i = \emptyset \), then
\[ \theta_i = \frac{1}{c} \sum_{j=1}^{c} d(x_i, v_j)^{1/m}, \forall j \in \mathcal{C}, \] (20)
The FCM can be obtained from the above Theorem in the special case where
\[ F(u_{ij}, \forall j \in \mathcal{C}) = \frac{1}{c} \sum_{j=1}^{c} (c u_{ij}). \] (21)
For this constraint function, the membership can be obtained from (18) as
\[ u_{ij} = \frac{1}{d(x_i, v_j)^{1/m}} \left( \sum_{j=1}^{c} \frac{1}{d(x_i, v_j)^{1/m}} \right)^{-1} \]
\[ = \frac{1}{c} \frac{d(x_i, \cdot)^{1/m}}{d(x_i, v_j)^{1/m}}, \] (22)
where \( d(x_i, \cdot)^{1/m} \) denotes the harmonic mean of \( d(x_i, v_j)^{1/m}, \forall j \in \mathcal{C} \), defined as
\[ \frac{1}{d(x_i, \cdot)^{1/m}} = \frac{1}{c} \sum_{j=1}^{c} \frac{1}{d(x_i, v_j)^{1/m}}. \] (23)
In this case, \( (d(x_i, \cdot)^{1/m})_{H} \) plays the role of the point of reference used for computing the membership.

Section 3 motivates the search for a variety of clustering algorithms whose point of reference lies in the interval \([d(x_i, \cdot)^{1/m}]_{\min}, d(x_i, \cdot)^{1/m}]_{\max}\), where \( (d(x_i, \cdot)^{1/m}]_{\min} \) and \( (d(x_i, \cdot)^{1/m}]_{\max} \) denote the minimum and maximum of \( d(x_i, v_j)^{1/m}, \forall j \in \mathcal{C} \), respectively. Thus, the development of Generalized FCM algorithms can be accomplished by computing the membership as
\[ u_{ij} = \frac{1}{c} \frac{d(x_i, v_j)^{1/m}}{d(x_i, \cdot)^{1/m}}, \] (24)
where \( d(x_i, \cdot)^{1/m} \) is the point of reference, determined in terms of \( d(x_i, v_j)^{1/m}, \forall j \in \mathcal{C} \). The constraint function (21) that resulted in the FCM is the arithmetic mean of \( c u_{ij}, v_j \in \mathcal{C} \). Thus, the selection of points of reference in the interval \([d(x_i, \cdot)^{1/m}]_{\min}, d(x_i, \cdot)^{1/m}]_{\max}\) can be facilitated by interpreting the arithmetic mean (21) as a special case of the generalized mean, which is often used to form aggregation operations on fuzzy sets [4].
Consider the generalized mean of \( c u_{ij}, \forall j \in \mathcal{C} \), defined \( \forall R \in (\infty, 0) \cup (0, \infty) \) as [4]
\[ F_R(u_{ij}, \forall j \in \mathcal{C}) = \left( \frac{1}{c} \sum_{j=1}^{c} (c u_{ij})^R \right)^{1/R}, \] (25)
As $R \to \infty$, (25) approaches the maximum of $c u_{ij}, \forall j \in C$ [4], that is,

$$\lim_{R \to \infty} F_R(u_{ij}, \forall j \in C) = \max_{\forall j \in C} (c u_{ij}) = c \max_{\forall j \in C} u_{ij}. \quad (26)$$

For $R = 1$, (25) gives the arithmetic mean of $c u_{ij}, \forall j \in C$, that is,

$$F_1(u_{ij}, \forall j \in C) = \frac{1}{c} \sum_{j=1}^{c} (c u_{ij}) = \sum_{j=1}^{c} u_{ij}. \quad (27)$$

As $R \to 0$, (25) approaches the geometric mean of $c u_{ij}, \forall j \in C$. In mathematical terms,

$$\lim_{R \to 0} F_R(u_{ij}, \forall j \in C) = \left( \prod_{j=1}^{c} (c u_{ij}) \right)^{\frac{1}{c}} = c \left( \prod_{j=1}^{c} u_{ij} \right)^{\frac{1}{c}} \quad (28)$$

For $R = -1$, $F_R(\cdot)$ is the harmonic mean of $c u_{ij}, \forall j \in C$, that is,

$$F_{-1}(u_{ij}, \forall j \in C) = c \left( \sum_{j=1}^{c} \frac{1}{c u_{ij}} \right)^{-1} = c^2 \left( \sum_{j=1}^{c} \frac{1}{u_{ij}} \right)^{-1} \quad (29)$$

where $u_{ij} \neq 0 \forall i, j$.

It can easily be verified that for any $\alpha > 0$ and $R \in (-\infty, 0) \cup (0, \infty)$, $F_R(\alpha u_{ij}, \forall j \in C) = \alpha F_R(u_{ij}, \forall j \in C)$. Thus, (25) is an admissible constraint function. According to the Theorem, the membership can be obtained in this case as

$$u_{ij}(R) = \frac{d(x_i, v_j)^{1-m}}{c \left( \frac{1}{c} \sum_{l=1}^{c} d(x_i, v_l)^{1-m} \right)^{\frac{1}{m}}}$$

$$= \frac{1}{c} \left( \frac{1}{c} \sum_{l=1}^{c} \left( \frac{d(x_i, v_j)}{d(x_i, v_l)} \right)^{\frac{R}{R-1}} \right)^{-\frac{1}{R}}. \quad (30)$$

For a fixed value of $m \in (1, \infty)$, the performance of the family of Generalized FCM algorithms determined by (19) and (30) depends on the value of $R \in (-\infty, 0) \cup (0, \infty)$.

The properties of the parameters $\theta_i, 1 \leq i \leq M$, resulting from the constraint function defined in (25) are described by the following proposition [2]:

**Proposition:** Suppose the constraint function $F_R(\cdot)$ is defined in (25).

If $F_R(u_{ij}, \forall j \in C) = 1$, then $\theta_i(R) = \sum_{j=1}^{c} u_{ij} \geq 1 \forall R \in [-1, 1]$, with the equality holding if $R = 1$ or if $R \neq 1$ and $u_{ij} = \frac{1}{c} \forall j \in C$.

If $F_R(u_{ij}, \forall j \in C) = 1$, then $\lim_{R \to 0} \theta_i(R) = \sum_{j=1}^{c} u_{ij} \geq 1 \forall R \in (0, 1)$, with the equality holding if $u_{ij} = \frac{1}{c} \forall j \in C$.

If $R = 1$, then $\theta_i = 1 \forall i = 1, 2, \ldots, M$. If $R \neq 1$, the parameters $\theta_i, 1 \leq i \leq M$, carry significant information regarding the certainty of the assignment of the corresponding vectors $x_i, 1 \leq i \leq M$ into $c$ clusters. In this case, $\theta_i = 1$ implies that the assignment of $x_i$ into $c$ clusters is maximally fuzzy, that is, $u_{ij} = \frac{1}{c} \forall j \in C$. Moreover, the membership values $u_{ij}, \forall j \in C$ are concentrated in a small neighborhood of $\frac{1}{c}$ if $\theta_i$ approaches 1. The membership $u_{ij}$ approaches $\frac{1}{c}$ for all $j \in C$ if $m$ approaches infinity. For the range of finite values of $m$ typically used in practical applications, the condition $u_{ij} = \frac{1}{c} \forall j \in C$ is satisfied if $d(x_i, v_j) = d(x_i, v_2) = \ldots = d(x_i, v_c)$. A feature vector $x_i$ that is equidistant from all cluster centers is frequently an outlier. In this particular case, $\theta_i(R) = 1 \forall R \in (-\infty, 0) \cup (0, \infty)$. If $R \neq 1$, the value of the parameter $\theta_i$ relative to 1 can be used to measure the certainty of the assignment of the corresponding feature vector $x_i$ into $c$ clusters. More specifically, the certainty of assignment of $x_i$ decreases as $\theta_i$ approaches 1 from the left if $R \in (1, \infty)$ or as $\theta_i$ approaches 1 from the right if $R \in (-\infty, 0) \cup (0, 1)$.

The GFCM algorithms can be summarized as follows:

1. Select $c, m, R$ and $\varepsilon$; fix $N$; set $\nu = 0$.

2. Generate an initial set of prototypes $V_0 = \{v_{1,0}, v_{2,0}, \ldots, v_{c,0}\}$.

3. Set $\nu = \nu + 1$.

   - $u_{ij,\nu} = \frac{1}{\varepsilon} \left( \frac{1}{\varepsilon} \sum_{l=1}^{c} \left( \frac{d(x_i, v_{j,\nu-1})}{d(x_i, v_{l,\nu-1})} \right)^{\frac{\nu-1}{\nu}} \right)^{-\frac{1}{\nu}}$
     \forall i = 1, 2, \ldots, M; \ j = 1, 2, \ldots, c.

   - $v_{j,\nu} = \left( \sum_{i=1}^{M} (u_{ij,\nu})^m x_i \right) / \left( \sum_{i=1}^{M} (u_{ij,\nu})^m \right), \ \forall j = 1, 2, \ldots, c.$

   - $E_{\nu} = \sum_{j=1}^{c} d(v_{j,\nu}, v_{j,\nu-1})$

4. If $\nu < N$ and $E_{\nu} > \varepsilon$, then go to step 3.

5. Special Cases

   This section presents specific algorithms resulting from the above formulation for some characteristic values of $R \in (-\infty, 0) \cup (0, \infty)$. 

1039
5.1 Minimum Fuzzy c-means Algorithms

This algorithm can be obtained as a special case of the Generalized FCM in the case where \( R \to \infty \). The membership can be evaluated in this case from (30) using that

\[
\lim_{R \to \infty} \left( \frac{1}{c} \sum_{\ell=1}^{c} \left( \frac{d(x_i, v_j)}{d(x_i, v_{\ell})} \right)^{\frac{R}{R-1}} \right)^h = \max_{v_{\ell} \in C} \left( \frac{d(x_i, v_j)}{d(x_i, v_{\ell})} \right)^{\frac{1}{R-1}}. \tag{31}
\]

Combining (30) and (31) gives

\[
\lim_{R \to \infty} u_{ij}(R) = \frac{1}{c} \left( \max_{v_{\ell} \in C} \left( \frac{d(x_i, v_j)}{d(x_i, v_{\ell})} \right)^{\frac{1}{R-1}} \right)^{-1} = \frac{1}{c} \left( \min_{v_{\ell} \in C} d(x_i, v_j) \right)^{\frac{1}{R-1}} \leq \frac{1}{c} \left( \frac{d(x_i, \cdot)}{d(x_i, v_j)} \right)^{\frac{1}{R-1}} \tag{32}
\]

where \( \langle d(x_i, \cdot)^{\frac{1}{R-1}} \rangle_{\min} \) is the minimum of \( d(x_i, v_j)^{\frac{1}{R-1}} \), \( \forall j \in C \). If the membership is given in (32),

\[
\theta_i = \langle d(x_i, \cdot)^{\frac{1}{R-1}} \rangle_{\min} \frac{1}{c} \sum_{j=1}^{c} \frac{1}{d(x_i, v_j)^{\frac{1}{R-1}}} \leq \frac{\langle d(x_i, \cdot)^{\frac{1}{R-1}} \rangle_{\min}}{\langle d(x_i, \cdot)^{\frac{1}{R-1}} \rangle_{H}} \leq 1 \forall i = 1, 2, \ldots, M, \tag{33}
\]

where \( \langle d(x_i, \cdot)^{\frac{1}{R-1}} \rangle_{H} \) is the harmonic mean of \( d(x_i, v_j)^{\frac{1}{R-1}} \), \( \forall j \in C \). The equality in (33) holds if \( x_i \) is an outlier, that is, if \( d(x_i, v_1) = d(x_i, v_2) = \ldots = d(x_i, v_c) \). Thus, the certainty of the assignment of \( x_i \in \mathcal{X} \) into \( c \) clusters decreases as \( \theta_i \) approaches 1 from the left.

5.2 Fuzzy c-means Algorithms

Consider the proposed formulation with \( R = 1 \). In this case, the resulting algorithm is the solution of the minimization problem (14) under the constraint that

\[
F_1(u_{ij}, \forall j \in C) = \frac{1}{c} \sum_{j=1}^{c} (c u_{ij}) = \sum_{j=1}^{c} u_{ij} = 1 \forall i. \tag{34}
\]

In this particular case, (34) specifies the parameters \( \theta_i, 1 \leq i \leq M \), which are involved in the definition of \( U_{\theta} \). The membership that corresponds to this algorithm can be obtained from (30) with \( R = 1 \) as

\[
u_{ij}(1) = \left( \sum_{\ell=1}^{c} \left( \frac{d(x_i, v_j)}{d(x_i, v_{\ell})} \right)^{\frac{1}{R-1}} \right)^{\frac{1}{{\frac{1}{R-1}}}}. \tag{35}
\]

5.3 Geometric Fuzzy c-means Algorithms

This algorithm can be obtained as a special case of the Generalized FCM in the case where \( R \to 0 \). Using the properties of the generalized mean,

\[
\lim_{R \to 0} \left( \frac{1}{c} \sum_{\ell=1}^{c} \left( \frac{d(x_i, v_j)}{d(x_i, v_{\ell})} \right)^{\frac{R}{R-1}} \right)^h = \left( \prod_{\ell=1}^{c} \left( \frac{d(x_i, v_j)}{d(x_i, v_{\ell})} \right)^{\frac{1}{R-1}} \right)^{\frac{1}{h}}. \tag{36}
\]

Substituting (36) in (30) gives

\[
\lim_{R \to 0} u_{ij}(R) = \frac{1}{c} \left( \prod_{\ell=1}^{c} \left( \frac{d(x_i, v_j)}{d(x_i, v_{\ell})} \right)^{\frac{1}{R-1}} \right)^{\frac{1}{h}} = \frac{1}{c} \left( \frac{d(x_i, \cdot)^{\frac{1}{R-1}}}_{G} \right) \right) \leq 1 \forall i = 1, 2, \ldots, M, \tag{37}
\]

where \( \langle d(x_i, \cdot)^{\frac{1}{R-1}} \rangle_{G} \) is the geometric mean of \( d(x_i, v_j)^{\frac{1}{R-1}} \), \( \forall j \in C \), defined as

\[
\langle d(x_i, \cdot)^{\frac{1}{R-1}} \rangle_{G} = \left( \prod_{j=1}^{c} d(x_i, v_j)^{\frac{1}{R-1}} \right)^{\frac{1}{h}} = \left( \frac{1}{c} \sum_{j=1}^{c} d(x_i, v_j)^{\frac{1}{R-1}} \right)^{\frac{1}{h}}. \tag{38}
\]

For the membership given in (37), \( \theta_i = \sum_{j=1}^{c} u_{ij} \) can be determined as

\[
\lim_{R \to 0} \theta_i(R) = \langle d(x_i, \cdot)^{\frac{1}{R-1}} \rangle_{G} \frac{1}{c} \sum_{j=1}^{c} \frac{1}{d(x_i, v_j)^{\frac{1}{R-1}}} \leq \frac{\langle d(x_i, \cdot)^{\frac{1}{R-1}} \rangle_{G}}{\langle d(x_i, \cdot)^{\frac{1}{R-1}} \rangle_{H}} \geq 1 \forall i = 1, 2, \ldots, M. \tag{39}
\]

The equality in (39) holds if \( x_i \) is an outlier, that is, if \( d(x_i, v_1) = d(x_i, v_2) = \ldots = d(x_i, v_c) \). Thus, the assignment of \( x_i \) into \( c \) clusters becomes increasingly uncertain as \( \theta_i \) approaches 1 from the right.
6. Experimental Results

The proposed Generalized FCM algorithms were used in the 2-partition problem considered by Krishnapuram and Keller [5]. Figure 1 shows 16 feature vectors, 14 of which form two well-separated physical clusters each containing 7 feature vectors. This set of feature vectors also contains two outliers located at equal distance from the two physical cluster centers. Tables 1 and 2 summarize the 2-partitions produced by Generalized FCM algorithms with \( m = 2 \) and \( R = 10 \), \( R = 1 \) (FCM), \( R = 0.5 \), and \( R \rightarrow 0 \) (Geometric FCM), respectively. As in [5], the feature vectors are numbered in Tables 1 and 2 in the order they are encountered by scanning the image shown in Figure 1 from the top to bottom and from the left to right. For each value of \( R \), Tables 1 and 2 show the membership values that assign each feature vector to the first and second physical clusters, respectively, and the values of \( \theta_i, i = 1, 2, \ldots, 16 \) assigned to each feature vector.

All the algorithms tested in these experiments resulted in the same crisp partition of the feature vectors that belong to the two physical clusters. Tables 1 and 2 can be used to evaluate various Generalized FCM algorithms in terms of their ability to: (i) identify outliers in the feature space, (ii) produce membership values that represent degree of belonging, and (iii) generate cluster centers close to the actual ones. For \( R > 1 \), it is not clear from the membership values that the feature vectors 1 and 2 are outliers. In this case, the membership values do not really represent “degree of belonging”. Nevertheless, the two outliers can be identified in this case by the corresponding values of \( \theta_1 \) and \( \theta_2 \), which are considerably closer to 1 than those of \( \theta_i, i = 3, 4, \ldots, 16 \). For \( R = 10, \theta_1 = \theta_2 = 1 \) while \( \theta_i, i = 3, 4, \ldots, 16 \) are slightly higher than \( \frac{1}{2} \). The FCM, which corresponds to \( R = 1 \), produced membership values that do not represent “degree of belonging”. Moreover, the two outliers cannot be identified in this case from the values of \( \theta_i, i = 1, 2, \ldots, 16 \), which are all equal to 1. As the value of \( R \) decreases below 1 and approaches 0, the membership values assigned to the feature vectors belonging to one of the two physical clusters become significantly higher than those assigned to the two outliers. For all values of \( R < 1 \), \( \theta_1 \) and \( \theta_2 \) are either equal to 1 \((R = 0.5)\) or slightly higher than 1 (Geometric FCM). In contrast, the values of \( \theta_i, i = 3, 4, \ldots, 16 \) are higher than 1. Thus, for \( R \in (0, 1) \) the membership values assigned to the feature vectors represent a “degree of belonging” to the clusters, which can also be accomplished by the FCM. Moreover, the values of \( \theta_i, i = 1, 2, \ldots, 16 \) identify clearly the outliers among the feature vectors. The cluster centers produced by the algorithms with \( R > 1 \) are far away from the actual centers (60, 150) and (140, 150) of the two physical clusters. In this case, the outliers attract the cluster centers because of the relatively high membership values assigned to them by the algorithm. The effect of the outliers in the calculation of the cluster.

| Table 1. Membership values and parameters \( \theta_i \) generated by the Generalized FCM algorithm with \( R = 10 \) and \( R = 1 \) (FCM). |
|-----------------|-----|-----|------|-----|-----|------|
| \( R = 10 \)    | \( R = 1 \) |
| \( i \) | \( u_{i1} \) | \( u_{i2} \) | \( \theta_i \) | \( u_{i1} \) | \( u_{i2} \) | \( \theta_i \) |
| 1    | 0.501 | 0.499 | 1.000 | 0.499 | 0.501 | 1.000 |
| 2    | 0.501 | 0.499 | 1.000 | 0.499 | 0.501 | 1.000 |
| 3    | 0.536 | 0.010 | 0.546 | 0.997 | 0.003 | 1.000 |
| 4    | 0.010 | 0.536 | 0.546 | 0.003 | 0.997 | 1.000 |
| 5    | 0.536 | 0.034 | 0.570 | 0.972 | 0.028 | 1.000 |
| 6    | 0.536 | 0.025 | 0.561 | 0.986 | 0.014 | 1.000 |
| 7    | 0.536 | 0.018 | 0.554 | 0.994 | 0.006 | 1.000 |
| 8    | 0.536 | 0.014 | 0.550 | 0.997 | 0.003 | 1.000 |
| 9    | 0.536 | 0.015 | 0.551 | 0.988 | 0.012 | 1.000 |
| 10   | 0.015 | 0.536 | 0.551 | 0.012 | 0.988 | 1.000 |
| 11   | 0.014 | 0.536 | 0.550 | 0.003 | 0.997 | 1.000 |
| 12   | 0.018 | 0.536 | 0.554 | 0.005 | 0.996 | 1.000 |
| 13   | 0.025 | 0.536 | 0.561 | 0.014 | 0.986 | 1.000 |
| 14   | 0.034 | 0.536 | 0.570 | 0.028 | 0.972 | 1.000 |
| 15   | 0.536 | 0.030 | 0.566 | 0.984 | 0.016 | 1.000 |
| 16   | 0.030 | 0.536 | 0.566 | 0.016 | 0.984 | 1.000 |
This paper also introduced a broad family of admissible constraint functions, whose form was inspired by the generalized mean formula. For a given \( m \), the proposed constraint functions resulted in infinitely many Generalized FCM algorithms, depending on \( R \in (-\infty,0) \cup (0,\infty) \). The analytically derived properties of the proposed algorithms were verified by experiments involving an artificial data set containing outliers. These experiments indicated that the membership values assigned to the feature vectors are increasingly representative of their “degree of belonging” to various clusters as \( R \) decreases below 1. For \( R \neq 1 \), the certainty of the assignment of each feature vector \( x_i \) into \( c \) clusters can be reliably measured by the corresponding parameter \( \theta_i \) assigned to it by the algorithms. The experiments also indicated that these certainty measures can be used in conjunction with the membership values to clearly identify outliers in the data set. The low computational requirements of the proposed Minimum FCM algorithms make them effective tools in applications which involve the clustering of a large number of feature vectors into a large number of clusters, such as codebook design for image compression [3].

### References


