A New Approach to Fuzzy Partitioning*

Frank Höppner*,**   Frank Klawonn

Department of Electrical Engineering and Computer Science
Emden University of Applied Sciences
Constantiaplatz 4
D-26723 Emden, Germany

Abstract

Fuzzy clustering algorithms like the popular fuzzy c-means algorithm (FCM) are frequently used to automatically divide up the data space into fuzzy granules (fuzzy vector quantization). In the context of fuzzy systems, in order to be intuitive and meaningful to the user, the fuzzy membership functions of the used linguistic terms have to fulfill some requirements like boundedness of support or unimodality. By rewarding crisp membership degrees, we modify FCM and obtain different membership functions that better suit these purposes. We show that the modification can be interpreted as standard FCM using distances to the Voronoi cell of the cluster rather than using distances to the cluster prototypes. In consequence, the resulting partitions of the modified algorithm are much closer to those of the crisp original methods. The membership functions can be generalized to a fuzzified minimum function. We give some bounds on the approximation quality of this fuzzification.

Keywords: fuzzy minimum function, fuzzy c-means, ISODATA, Voronoi diagram, nearest neighbour, fuzzy partition, noise sensitivity

1 Introduction

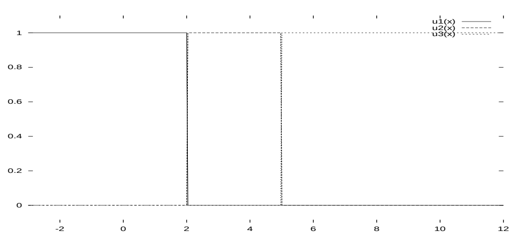
When building fuzzy systems automatically from data, we are in need of procedures that automatically divide up the input space in fuzzy granules. These granules are the building blocks for the fuzzy rules. When modelling an input/output relationship, the membership functions of these rules play the same role as basis functions in conventional function approximation tasks. To keep interpretability we usually require that the fuzzy sets are specified in *local regions*, that is, the membership functions have bounded support or decay rapidly. If this requirement is not fulfilled, many rules must be applied and aggregated simultaneously, such that the final result becomes more difficult to grasp – one is not allowed to interpret a fuzzy system *rule by rule* any longer. A second requirement is that the fuzzy sets of the primitive linguistic values should be simple and unimodal. It would be counterintuitive if the membership of the linguistic term “young”, which is high for for “17 years”, would be higher for “23 years” than for “21 years”.

To gain such fuzzy granules clustering algorithms can be used. Especially fuzzy clustering algorithms seem well suited, because they provide the user with a fuzzy membership function which could be used directly for the linguistic terms. Unfortunately, the family of the fuzzy c-means (FCM) clustering algorithms [1] and derivatives produce membership functions that do not fulfil the abovementioned requirements [5]. Figure 1(c) shows an example for FCM membership functions for a partition of the real line with cluster representatives $c_1 = 1$, $c_2 = 3$ and $c_3 = 8$. We can observe that the support of the membership functions is unbounded for all clusters, in particular for the cluster whose centre is located at $c_2 = 3$. While for $c_1 = 1$ and $c_3 = 8$ one allows even in the context of fuzzy systems for an unbounded support if $x < 1$ and $x > 8$ respectively, but at least the membership function for $c_2 = 3$ should be defined locally. Furthermore, we can observe that the membership degree for the cluster at $c_1 = 1$ increases near 5, the FCM membership functions are not unimodal. These undesired properties can be reduced by tuning a parameter of the FCM algorithm, the so-called fuzzifier, however, then we also decrease the fuzziness of the partition and finally end up with crisp indicator functions as shown in figure 1(a). The problem of unimodality can be solved by using possibilistic memberships [2], but the possibilistic c-means is not a partitional but a mode-seeking algorithm. In [5] the objective function has been completely abandoned to allow user-defined membership functions, thereby also loosing the partitional property.

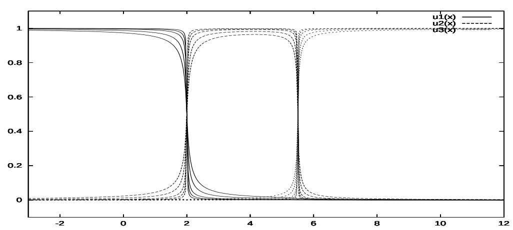
In this paper, we investigate alternative approaches to influence the fuzziness of the final partition. We consider a “reward” term for membership degrees near 0 and 1 in order to force a more crisp assignment in sec-

*This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) under grant no. Kl 648/1-1.

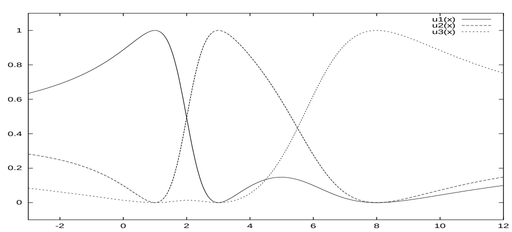
**Contact: frank.hoeppner@ieee.org



(a) Indicator functions of crisp partition.



(b) Intuitively fuzzified partitions of fig. 1(a).



(c) FCM membership functions ($m = 2.0$).

Figure 1. Different kinds of membership functions.

tion 3. If we choose a (in some sense) maximal reward, we arrive at fuzzy membership functions which are identical to those that we would obtain by using a (scaled) distance to the Voronoi cell that represents the cluster instead of the Euclidean distance to the clusters centre, as we will see in section 4. Furthermore, the membership functions – as a whole – can be interpreted as a *fuzzified minimum function* [3], for which we give an estimation of the error we make when substituting a crisp minimum function by its fuzzy version (section 5).

2 Objective-Function Based Fuzzy Clustering

In this section we briefly review the fuzzy c-means algorithm [1], for a thorough overview of objective-function based fuzzy clustering see [4], for instance. Let us denote the membership degree of datum $x_j \in X$, $j \in \{1, ..., n\}$, to cluster $p_i \in P$, $i \in \{1, ..., c\}$, by $u_{i,j} \in [0, 1]$. Denoting the Euclidean distance by d_E, we minimize the objective function

$$J_m(P, U; X) = \sum_{j=1}^{n} \sum_{i=1}^{c} u_{i,j}^m d_E^2(x_j, p_i)$$

iteratively subject to the constraints

$$\forall_{1 \le j \le n} : \sum_{i=1}^{c} u_{i,j} = 1, \quad \forall_{1 \le i \le c} : \sum_{j=1}^{n} u_{i,j} > 0 \quad (1)$$

In every iteration step, minimization with respect to $u_{i,j}$ and p_i is done separately. The necessary conditions for a minimum yield update equations for both half-steps as follows

$$u_{i,j} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_E^2(x_j, p_i)}{d_E^2(x_j, p_k)} \right)^{\frac{1}{m-1}}} \quad (2)$$

and for the prototypes

$$p_i = \frac{\sum_{j=1}^{n} u_{i,j}^m x_j}{\sum_{j=1}^{n} u_{i,j}^m} \quad (3)$$

Figure 2(a) shows an example for an FCM clustering with $c = 7$. The membership degrees are indicated by contour lines, the maximum over all membership degrees is depicted.

3 Rewarding Crisp Memberships in FCM

Some properties of the membership functions defined by (2) are undesired – at least in some application areas, as we have seen in the introduction. Let us consider the question how to reward more crisp membership degrees. We would like to avoid those small peaks of high membership degrees (cf. figure 1(c)) and are interested in broad areas of (nearly) crisp membership degrees and only narrow regions where the membership degree changes from 0 to 1 or vice versa (cf. figure 1(b)). Let us choose a couple of parameters $a_j \in \mathbb{R}_{\ge 0}$, $1 \le j \le n$, and consider the following modified objective function

$$J = \sum_{j=1}^{n} \sum_{i=1}^{c} u_{i,j}^2 d_E^2(x_j, p_i) - \sum_{i=1}^{n} a_j \sum_{j=1}^{c} \left(u_{i,j} - \frac{1}{2} \right)^2 \quad (4)$$

The first term is identical to the objective function of FCM with $m = 2$. Let us therefore examine the second term. If a data object x_j is clearly assigned to one prototype p_i, then we have $u_{i,j} = 1$ and $u_{k,j} = 0$ for all other $k \neq i$. For all these cases, the second term evaluates to $-\frac{a_j}{4}$. If the membership degrees become more fuzzy, the second term increases. Since we seek to minimize (4), this modification rewards crisp membership degrees.

Since there are no additional occurrences of p_i in the second term, the prototype update step remains the same as with FCM, as given by (3).

Lemma 1 *The necessary condition for a minimum of (4) yields the following membership update equation:*

$$u_{i,j} = \frac{1}{\sum_{k=1}^{c} \frac{d_E^2(x_j,p_i)-a_j}{d_E^2(x_j,p_k)-a_j}} \tag{5}$$

Proof: Let us consider (4) for a single datum x_j. We apply Lagrange multipliers to satisfy the constraint $\sum_{i=1}^{c} u_{i,j} = 1$ for x_j (cf. (1)). We have $F =$

$$\sum_{i=1}^{c} u_{i,j}^2 \|x_j - p_i\|^2 - \sum_{i=1}^{c} a_j \left(u_{i,j} - \frac{1}{2}\right)^2 + \lambda \left(\sum_{i=1}^{c} u_{i,j} - 1\right)$$

Setting the gradient to zero yields

$$\frac{\partial F}{\partial \lambda} = \sum_{i=1}^{c} u_{i,j} - 1 = 0$$

$$\frac{\partial F}{\partial u_{k,j}} = 2u_{k,j}\|x_j - p_k\| - 2a_j(u_{k,j} - \frac{1}{2}) + \lambda$$

Note that we have fixed $m = 2$ in (4) to obtain an analytical solution. From $\frac{\partial F}{\partial u_{k,j}}$ we obtain

$$u_{k,j} = \frac{-a_j - \lambda}{2\|x_j - p_k\| - 2a_j}$$

Using $\frac{\partial F}{\partial \lambda}$, we have

$$\sum_{i=1}^{c} \frac{-a_j - \lambda}{2\|x_j - p_i\| - 2a_j} = 1$$

$$\Leftrightarrow \quad \lambda = -\frac{1}{\sum_{i=1}^{c} 2\|x_j - p_i\| - 2a_j} - a_j$$

Substituting λ in the previous equation yields (5). ■

Obviously, we immediately run into some problems when choosing $a_j > \|x_j - p_i\|$ for some $1 \le i \le c$. Then, the distance value $d_{i,j}^2 - a_j$ becomes negative and the same is true for the membership degrees ($d_{i,j} = d_E(x_j, p_i)$). Therefore, we have to require explicitly the constraint $0 \le u_{i,j} \le 1$. From the Kuhn-Tucker conditions we obtain a simple solution as long as only a single prototype has a distance smaller than a_j to x_j, in this case we obtain the minimum by setting $u_{i,j} = 1$. However, things are getting more complicated if multiple negative terms $d_{i,j} - a_j$ occur.

If we want to avoid the problem of negative memberships, we could also heuristically adapt the reward a_j such that $d_{i,j}^2 - a_j$ always remains positive. The maximal reward we can give is then

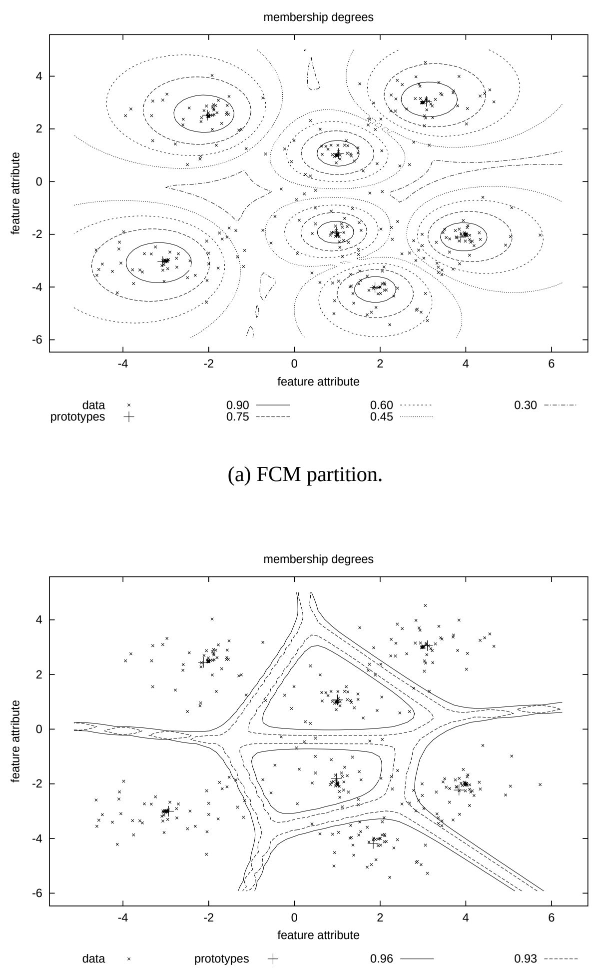
$$\min d_{*,j}^2 = \min\{\, d_{i,j}^2 \,|\, i \in \{1,..,c\}\,\} - \eta$$

and thus

$$u_{i,j} = \frac{1}{\sum_{k=1}^{c} \frac{d_{i,j}^2 - \min d_{*,j}^2}{d_{k,j}^2 - \min d_{*,j}^2}} \tag{6}$$

Without an $\eta > 0$ we find always an i such that $d_e^2(x_j, p_i) - \min_{*,j}^2 = 0$ and therefore $u_{i,j} = 1$. In other words, for $\eta = 0$ we obtain a crisp partition, the algorithm reduces to ISODATA. The choice of η influences the fuzziness of the partition, similar to the fuzzifier m with FCM. Figure 1(b) shows different partitions for η ranging from 0.01 to 0.2.

Surprisingly, besides the different shape of the membership functions, the resulting algorithm performs very similar to conventional FCM, in terms of resulting cluster centres. The modified version seems slightly less sensitive to noise and outliers, as we will see in the next section. Figure 2 compares the results of FCM and our modification for an example dataset. The maximum over all membership degrees is indicated by contour lines.



(a) FCM partition.

(b) Voronoi-like partition.

Figure 2. Example.

4 Memberships Induced by Voronoi Distance

With FCM the Euclidean distance between cluster centroids plays a central role in the definition of the membership functions. The idea is to "represent" each cluster by a single data instance – the prototype – and to use the distance between prototype and data objects as the distance between cluster and data object. Then, the relative distances (cf. (2)) define the degree of membership to a cluster, e.g., if the distance between x_j and p_1 is half the distance to p_2, the membership degree $u_{1,j}$ is twice as large as $u_{2,j}$. If we consider crisp membership degrees things are different, the membership degree does not depend on the ratio of distances, but the distances serve as threshold values. If the distance to p_1 is smaller than to p_2 – no matter how much smaller – we always have $u_{j,1} = 1$.

Let us consider (6) again and assume that p_i is closest to x_j. No matter if x_j is far away from p_i (but all other p_k are even further away) or x_j is very close to p_i, the numerator of the distance ratio is always constant η. Inside a region in which all data points are closest to p_i, the distance to cluster i is considered to be constant η. The membership degrees $u_{k,j}$ are therefore determined by the denominator, that is, mainly by $d^2_E(x_j, p_k)$. Therefore, the membership degrees obtained by (6) are no longer defined by a ratio of distances, but the maximum reward $(\min d^2_{*,j})$ has the flavour of a threshold value.

Let us consider a crisp partition, which is induced by cluster centroids. The resulting partition is usually referred to as the Voronoi diagram. The Euclidean distance of a data object x_j to the hyperplane that separates the clusters p_i and p_s is given by $|(x_j - h_s)^\top n_s|$ where h_s is a point on the hyperplane, e.g., $h_s = (p_s + p_i)/2$, and n_s is the normal vector $n_s = \beta_s \cdot (p_i - p_s)$ with $\beta_s = \frac{1}{\|p_i - p_s\|}$ for $s \neq i$. How can we define the distance of a data object x_j to the Voronoi cell of cluster i rather than to a separating hyperplane? If we do not take absolute values, we obtain *directed* distances $(x_j - h_s)^\top n_s$, which become positive if x_j lies on the same side as the cluster centre and negative if x_j lies on the opposite side. Taking the absolute value of the minimum over all the directed distances yields the distance to the *border of the cell* (see also [3] for the case of rectangles in shell clustering). If x_j lies within the Voronoi cell of cluster i, then the distance to the *cell* is zero. We can formalize this special case easily by setting $b_s = 1$ and defining:

$$d_V(x_j, p_i) = \left| \min_{1 \le s \le c} (x - h_s)^\top n_s \right|$$

In figure 3, x_j is closest to the separating line between p_1 and p_2, therefore this distance serves as the distance to the Voronoi cell of p_1. The graph of d_V for the 4 clusters of figure 3 is shown in figure 4.

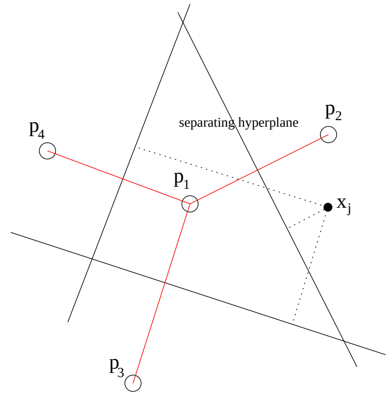


Figure 3. Voronoi cell of centroid p_1.

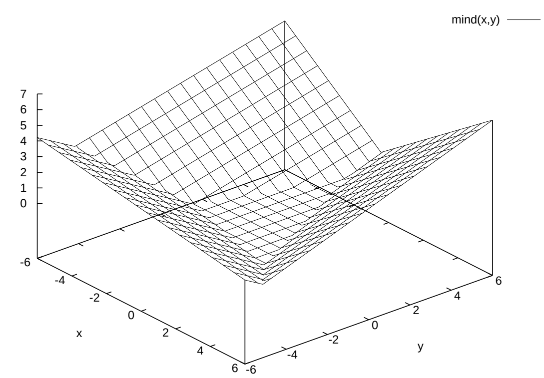


Figure 4. Distance to Voronoi cell.

If we do not scale the normal vectors n_s to unit length, but assume $\beta_s = 1$ for all s, we preserve the shape of d_V (position of hyperplanes does not change), only the gradient of the different hyperplanes varies. The following lemma establishes a connection between the scaled Voronoi distance and the approach discussed in the previous section.

Lemma 2 *Given a Voronoi diagram induced by a set of distinct points p_i, $1 \le i \le c$, and a point x. Using $\beta_s = 1$ for all $1 \le s \le c$, the (scaled) distance between x and the Voronoi cell of point p_i is given by*

$$d_V(x, p_i) = \frac{1}{2}\left(d^2_E(x, p_i) - \min_{1 \le s \le c} d^2_E(x, p_s) \right) \quad (7)$$

Proof: Some simple transformations yield the following chain of equalities

$$\begin{aligned} & d_V(x, p_i) \\ = & \left| \min_{1 \le s \le c} \left(x - \frac{p_s + p_i}{2} \right)^\top (p_i - p_s) \right| \\ = & \left| \min_{1 \le s \le c} x^\top (p_i - p_s) - \frac{1}{2}(p_i^\top p_i - p_s^\top p_s) \right| \\ = & \frac{1}{2} \left| \min_{1 \le s \le c} x^\top x - 2x^\top p_s + p_s^\top p_s - \right. \end{aligned}$$

$$
\begin{aligned}
& \left. + (x^\top x - 2x^\top p_i + p_i^\top p_i) \right| \\
= \;& \frac{1}{2} \left| \min_{1\le s\le c} \|x - p_s\|^2 - \|x - p_i\|^2 \right| \\
\overset{(\star)}{=} \;& \frac{1}{2} \left(\|x - p_i\|^2 - \min_{1\le s\le c} \|x - p_s\|^2 \right)
\end{aligned}
$$

In equation ($\star$) we have used the trivial fact that any $d_E(x, p_i)$ is greater than or equal to $\min_{1\le s\le c} d_E(x, p_s)$. ■

Thus, the lemma tells us, by using a maximum reward the resulting membership values are identical to those that we would obtain by using standard FCM membership functions and a (scaled) Voronoi cell distance instead of Euclidean centroid distance.

By replacing the Euclidean distance with the Voronoi distance during membership calculation, we obtain different membership functions which are much closer to those of the original ISODATA (cf. figure 2(b)). In this sense we can speak of a new ISODATA fuzzification.

Note that with FCM squared Euclidean distances are used to determine the membership degrees, but if we use the maximum reward/Voronoi distance we use Euclidean distances to the Voronoi cell, which are not squared. Therefore, the modification might be less sensitive to noise and outliers.

5 Interpretation as Fuzzified Minimum Function

In the previous sections we have seen how the introduction of a reward term leads us to a fuzzy partition which is more closely related to the results of the crisp ISODATA (or a Voronoi partition) than the standard FCM partition. The ISODATA algorithm minimizes the objective function

$$\sum_{j=1}^{n} \min_{1\le i\le c} \|x_j - p_i\|^2$$

The crisp minimum function can be reformulated as

$$\min_{1\le i\le c} \|x_j - p_i\|^2 = \sum_{i=1}^{c} u_{i,j} \|x_j - p_i\|^2 \qquad (8)$$

using crisp membership degrees $u_{i,j}$ defined by $u_{i,j} = 1 \Leftrightarrow i = \text{argmin}\|x_j - p_i\|^2$ (0 otherwise). If the partition of the discussed algorithm can be interpreted as a fuzzified Voronoi diagram, is it also possible to interpret the term $\sum_{i=1}^{c} u_{i,j}^2 d_V(x_j, p_i)$ as a fuzzified minimum function? We have faced the problem of a fuzzified minimum function before in [3]. There, we considered the terms $d_i = b_i - \min_{1\le s\le k} b_s$ in a minimum term $\min(b_1, b_2, ..., b_k)$ as the "distance of argument i to the minimum" and used the standard FCM membership degrees to assign a "degree of minimality" to each argument b_i. Note that this leads to the same equations as we have discussed in the previous sections.

Regarding the approximation quality, we state the following theorem:

Theorem 1 (Fuzzified Minimum Function) *Let $f : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ be a strictly increasing function with $f(x) \ge x$, let $\eta \in \mathbb{R}_{\ge 0}$. Then for all $d = (d_1, ..., d_k) \in \mathbb{R}^k$, $D_s = (f(d_s - \min\{d_1, .., d_k\}) + \eta)^q$, $q \ge 1$, the following inequality holds:*

$$
\begin{aligned}
& \left| \sum_{s=1}^{k} u_s\, d_s - \min\{d_1, d_2, ..., d_k\} \right| \\
& < \eta^q r + \eta(k - r - 1) \le \eta(k-1)
\end{aligned}
$$

where $u_s = \frac{1}{\sum_{i=1}^{k} \frac{D_s}{D_i}}$ and r is the number of indices s for which d_s has at least a distance of $1-\eta$ from the minimum: $r = |\{s \,|\, 1 \le s \le k,\ d_s - \min\{d_1, d_2, ..., d_k\} > 1 - \eta\}|$

Proof: We have the following equality

$$
\begin{aligned}
& \sum_{s=1}^{k} \frac{d_s}{D_s \sum_{i=1}^{k} \frac{1}{D_i}} \\
= \;& \sum_{s=1}^{k} \frac{d_s}{D_s \frac{\sum_{i=1}^{k} \prod_{t=1, t\ne i}^{k} D_t}{\prod_{i=1}^{k} D_i}} \\
= \;& \sum_{s=1}^{k} \frac{d_s \prod_{i=1}^{k} D_i}{D_s \sum_{i=1}^{k} \prod_{t=1, t\ne i}^{k} D_t} \\
= \;& \sum_{s=1}^{k} \frac{d_s \prod_{i=1, i\ne s}^{k} D_i}{\sum_{i=1}^{k} \prod_{t=1, t\ne i}^{k} D_t} \\
= \;& \frac{\sum_{s=1}^{k} d_s \prod_{i=1, i\ne s}^{k} D_i}{\sum_{i=1}^{k} \prod_{t=1, t\ne i}^{k} D_t} \qquad (9)
\end{aligned}
$$

Using the abbreviations $M = \min\{d_1, d_2, ..., d_k\}$ we estimate the approximation error as follows

$$
\begin{aligned}
& \left| \frac{\sum_{s=1}^{k} d_s \prod_{i=1, i\ne s}^{k} D_i}{\sum_{s=1}^{k} \prod_{i=1, i\ne s}^{k} D_i} - M \right| \\
= \;& \left| \frac{\left(\sum_{s=1}^{k} d_s \prod_{i=1, i\ne s}^{k} D_i\right) - M \left(\sum_{s=1}^{k} \prod_{i=1, i\ne s}^{k} D_i\right)}{\sum_{s=1}^{k} \prod_{i=1, i\ne s}^{k} D_i} \right| \\
= \;& \left| \frac{\sum_{s=1}^{k} (d_s - M) \prod_{i=1, i\ne s}^{k} D_i}{\sum_{s=1}^{k} \prod_{i=1, i\ne s}^{k} D_i} \right| \\
\overset{\star^1}{=} \;& \left| \frac{\sum_{s=2}^{k} (d_s - M) \prod_{i=1, i\ne s}^{k} D_i}{\sum_{s=1}^{k} \prod_{i=1, i\ne s}^{k} D_i} \right| \\
\overset{\star^2}{\le} \;& \left| \frac{\eta^q \sum_{s=2}^{k} (d_s - M) \prod_{i=2, i\ne s}^{k} D_i}{\sum_{s=1}^{k} \prod_{i=1, i\ne s}^{k} D_i} \right| \\
\overset{\star^3}{<} \;& \left| \frac{\eta^q \sum_{s=2}^{k} (d_s - M) \prod_{i=2, i\ne s}^{k} D_i}{\prod_{i=2}^{k} D_i} \right| \\
= \;& \left| \eta^q \sum_{s=2}^{k} \frac{(d_s - M)}{D_s} \right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \quad \eta^q \sum_{s=2}^{k} \left| \frac{(d_s - M)}{D_s} \right| \\
&\stackrel{\star^4}{<} \quad \eta^q \sum_{s=2}^{k} \left| \frac{(d_s - M)}{(d_s - M)(d_s - M + \eta)^{q-1}} \right| \\
&= \quad \eta^q \sum_{s=2}^{k} \left| \frac{1}{(d_s - M + \eta)^{q-1}} \right| \\
&\stackrel{\star^5}{\leq} \quad \eta^q \sum_{s=2}^{k} \left| \frac{1}{\eta^{q-1}} \right| \\
&= \quad \eta(k-1)
\end{aligned}
$$

Remarks:

- $\star^1$ Without loss of generality we have assume that d_1 is the minimum and have $(d_1 - M) = 0$.
- $\star^2$ From $d_1 = M$ we can conclude $D_1 = (f(d_1 - d_1) + \eta)^q \leq \eta^q$.
- $\star^3$ We have dropped all summands in the denominator $\sum_{s=1}^{k} \prod_{i=1, i \neq s}^{k} D_i$ that contain D_1. All summands are positive.
- $\star^4$ We drop one η in the denominator $D_s = (d_s - M + \eta)(d_s - M + \eta)^{q-1}$ which makes the term smaller.
- $\star^5$ Here we assume the worst case that all d_s are minimal and thus $d_s - M = 0$. (However, if this would actually be the case, we can see from the equality $\star^1$ that the approximation error is zero.) We also obtain an equality if $q = 1$.

If some d_s, $s \in \{2, 3, .., k\}$, have reached a distance $d_s - M \geq 1 - \eta$ from the minimum, the estimation can be improved[1]. If we continue from the result after $\star^3$ we have $d_s - M < f(d_s - M) + \eta < (f(d_s - M) + \eta)^q = D_s$ and thus may substitute $(d_s - M)$ by D_s. This leads us to an error below $\eta^q(k-1)$.

To summarize both estimations, if there are r values that have a distance of at least $d_s > 1 - \eta + M$, we have an error smaller than $\eta(k - r - 1) + \eta^q r$. ■

Although we deal only with non-negative distances in the context of clustering, note that the fuzzified minimum function does also work with negative terms. If there are negative arguments, the minimum will also be negative, and subtracting the (negative) minimum from all other arguments yields a set of non-negative arguments. Also note that the fuzzified minimum is once differentiable for $q > 1$. Figure 5 shows an example where we take the pointwise minimum of three functions. The resulting fuzzified minimum is displayed for two different values of $\eta = 0.1/0.2$ (solid lines) using $q = 1.5$. According to the theorem, the error is bounded by $0.06/0.18$ if the minimum is clearly separated from the other values and $0.2/0.4$ in general.

[1]This additional condition has not been mentioned in [3].

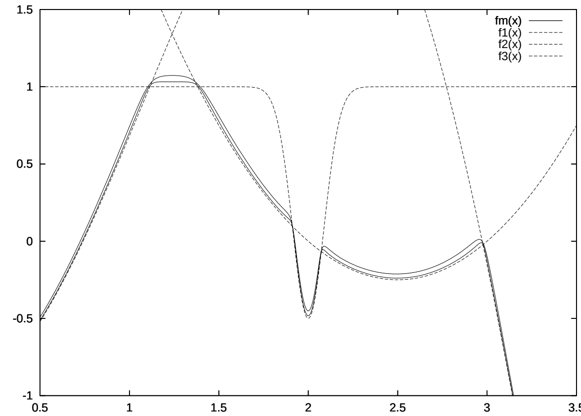


Figure 5. Minimum of three functions.

6 Conclusions

In this paper, we have presented a modification of FCM, which is more closely related to the original (non-fuzzy) ISODATA algorithm. This can be desirable for certain applications, for example if we want to attach linguistic labels to the membership functions. We have proposed a modification of the objective function that is minimized by FCM to reward nearly crisp memberships. If we (heuristically) select a (in some sense) "maximum reward", we have shown that the membership functions correspond to membership functions that would be obtained by using the distance between the Voronoi cell and a data object.

The obtained membership functions can also be interpreted as a fuzzified minimum function. In retroperspective, we can consider the modification as a substitution of the crisp minimum function of ISODATA by a fuzzified variant.

References

[1] J. C. Bezdek. A convergence theorem for the fuzzy ISODATA clustering algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2(1):1–8, Jan. 1980.

[2] R. N. Davé and R. Krishnapuram. Robust clustering methods: A unified view. *IEEE Trans. on Fuzzy Systems*, 5(2):270–293, May 1997.

[3] F. Höppner. Fuzzy shell clustering algorithms in image processing: Fuzzy c-rectangular and 2-rectangular shells. *IEEE Trans. on Fuzzy Systems*, 5(4):599–613, Nov. 1997.

[4] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler. *Fuzzy Cluster Analysis*. John Wiley & Sons, Chichester, England, 1999.

[5] T. A. Runkler and J. C. Bezdek. Alternating cluster estimation: A new tool for clustering and function approximation. *IEEE Trans. on Fuzzy Systems*, 7(4):377–393, Aug. 1999.