FUZZY C - MEANS AS A REGULARIZATION AND MAXIMUM ENTROPY APPROACH

Sadaaki Miyamoto[a], Masao Mukaidono[b]

[a] *University of Tsukuba, Institute of Information Sciences and Electronics, Tsukuba-shi, Ibaraki 305, Japan*
[b] *Department of Computer Science, Meiji University, Kawasaki-shi 214, Japan*

Abstract. A maximum entropy method has been proposed by one of the authors as a variation of the fuzzy c - means. In this paper this method and the standard fuzzy c-means are regarded as *regularizations* of the crisp k - means. As a result, further variations of the maximum entropy method are developed that are parallel to the corresponding methods of the ordinary fuzzy c - means. Moreover, the fuzzy k - nearest neighbor classification rule which has been derived from the standard fuzzy c - means is transformed into the corresponding method within the maximum entropy framework.

Keywords. Fuzzy c - means, maximum entropy, regularization, fuzzy classifier, k - nearest neighbor rule.

1 Introduction

Fuzzy c - means clustering (abbreviated as FCM) [5, 6, 2], which is the fuzzy version of the k - means [17], has been studied by many researchers. Recently, one of the authors has proposed a maximum entropy method [15, 16] that can be used instead of the standard FCM.

In this paper we show that the standard fuzzy c - means and the maximum entropy method are different types of *regularizations* for the crisp k - means.

Regularization is an old technique to solve ill-posed problems of functional equations [23] and has been applied to many real problems. For example, the ridge estimator in regression analysis [10] is closely related to regularization. Regularization in general implies modification of a given problem that is singular in some sense into a regular problem. The singular problem is difficult to solve but the latter problem is easier to handle. The latter problem is called the regularization of the original problem when the solution of the regularized problem approximates the original solution.

We recognize fuzzy c - means to be a regularization for the crisp k - means. It is strange to say that the well-known method of k - means is singular, since the solution of the crisp k - means is by no means ill-posed.

Although the crisp k - means is a well-posed problem, the crisp solution is characterized by the extremal points when the optimal clustering is generalized into the fuzzy membership case [2]. Thus, we regard extremal points as singular solutions, whereas intermediate points are considered to be regular solutions.

This idea is fruitful in the sense that we have possibility of other regularizations than the standard one by Dunn [5, 6] and Bezdek [2].

The concept of the regularization in FCM has briefly been suggested in Miyamoto [20]. In this paper we show how this concept works for the maximum entropy approach. As a result we obtain a number of new algorithms that are parallel to the standard FCM.

Moreover, a method of supervised fuzzy classification that has been derived from the k - nearest neighbor rule [4] can be transformed into the corresponding method of classification by the maximum entropy approach.

Thus, what we present is not a single method but a framework of maximum entropy in which we can develop a family of new methods.

2 Fuzzy c - means as a regularization

2.1 Review of fuzzy c - means

Let us briefly review fuzzy c - means for discussing regularizations. In the following we are referring to description by Bezdek [2].

Let us first assume that the individuals to be clustered be $X = \{x_1, ..., x_n\}$ and suppose that $x_1, ..., x_n$ are points of p dimensional Euclidean space, although we consider other spaces later. Remark also that the number of clusters is given by c. Hereafter we refer to the methods as c - means, not as k - means.

The introduction of the standard fuzzy c - means starts from a version of crisp c - means algorithm formulated as an optimization problem. Namely, the function to be minimized is

$$J_1(U, v) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} d(x_k, v_i), \tag{1}$$

where $U = (u_{ik})$ and u_{ik} is a binary variable showing whether x_k belongs to the cluster i ($u_{ik} = 1$) or not

$(u_{ik} = 0)$; $v = (v_1, ..., v_c)$ and each v_i is the center for the cluster i. $d(x_k, v_i)$ is a dissimilarity measure between x_k and v_i, and for the most part $d(x_k, v_i) = \|x_k - v_i\|_2^2$ where $\|\cdot\|_2$ is the Euclidean norm. Since each individual belongs to one and only one cluster, the admissible set for u_{ik} is

$$M_c = \{(u_{ik}) | \sum_{i=1}^{c} u_{ik} = 1, u_{ik} \in \{0,1\}, k = 1, ..., n\}.$$

In most cases of different c - means, a two stage alternative optimization procedure is used for minimization of the objective function.

Procedure CM.

CM1. Set initial values for $\bar{U}$ and $\bar{v}$.

CM2. Minimize $J(U, \bar{v})$ with respect to $U \in M$ and let the optimal solution be $\bar{U}$.

CM3. Minimize $J(\bar{U}, v)$ with respect to v and let the optimal solution be $\bar{v}$.

CM4. Check the stopping criterion. If the criterion is not satisfied, go to **CM2**.

If we put $J = J_1$ and the admissible set $M = M_c$, the above procedure is reduced to a standard algorithm of crisp c -means which is sometimes referred to as Forgy's method [1, 7].

The first step toward the fuzzy c - means is to generalize the binary valued U to fuzzy U. Thus, u_{ik} may take any value in the unit interval, and therefore the admissible set becomes

$$M_f = \{(u_{ik}) | \sum_{i=1}^{c} u_{ik} = 1, u_{ik} \in [0,1], k = 1, ..., n\}.$$

This generalization is insufficient. Suppose that we use the procedure **CM** with $J = J_1$ and $M = M_f$. We obtain crisp solutions, since the objective function is linear for u_{ik} and hence the step **CM2** is a linear programming. It is well-known that the optimal solution for a linear programming is on an extremal point: in this case u_{ik} is zero or unity.

Consequently, Dunn [5, 6] and Bezdek [2] studied a modified criterion

$$J_m(U, v) = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ik})^m d(x_k, v_i)$$

using a parameter $m > 1$. By using this parameter, we have fuzzy memberships; it is easily seen that $0 < u_{ik} < 1$, unless the point x_k is on a center v_i.

It is well-known that the solution for $\bar{U}$ with a given $\bar{v}$ in step **CM2** is

$$\bar{u}_{ik} = [\sum_{j=1}^{c} (\frac{d(x_k, \bar{v}_i)}{d(x_k, \bar{v}_j)})^{\frac{1}{m-1}}]^{-1}, \tag{2}$$

while the solution for $\bar{v}$ with a given $\bar{v}$ in step **CM3** is

$$\bar{v}_i = \frac{\sum_{k=1}^{n} (\bar{u}_{ik})^m x_k}{\sum_{k=1}^{n} (\bar{u}_{ik})^m} \tag{3}$$

in the case of the Euclidean space $d(x_k, v_i) = \|x_k - v_i\|_2^2$. Moreover we sometimes write d_{ik} instead of $d(x_k, v_i)$. In particular, the last symbol may imply variations of the dissimilarity $d(x_k, v_i)$.

As mentioned earlier, we regard binary solutions as extremal and hence singular, when the admissble set M_f of fuzzy memberships is assumed. Thus, the regularizing parameter is m, and the objective function is transformed by using this parameter which results in regularized solutions that are fuzzy. Notice that when we refer to regularizations, the regularized functional is similar to the original one and the regularized solution is *near* the original solution. The last statement is difficult to prove exactly, but we have many empirical evidence that fuzzy solutions approximate corresponding crisp solutions.

What is the advantage of taking this standpoint of the regularization? The answer is that this argument implies that other types of regularizations are possible. Indeed, a recent method by one of the authors is considered to be a typical regularization of the crisp c - means.

In order to see this, notice that a typical regularization for the optimization of J is done by adding regularizing functional K with a positive parameter α:

$$J_\alpha = J + \alpha K$$

and the minimization of J_α is considered instead of minimizing J. The regularizing parameter α should be adjusted so that good quality solutions are obtained. The maximum entropy method has the last form, whereas the standard fuzzy c - means does not. In this sense the standard fuzzy c - means is a *nonlinear* regularization.

2.2 Maximum entropy approach

The maximum entropy method [15, 16] has been introduced as the following optimization problem:

$$\begin{aligned} &maximize \quad -\sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} \log u_{ik} \\ &subject\ to \quad \sum_{i=1}^{c} u_{ik} = 1, \quad J_1(U, v) = K \end{aligned}$$

where $J_1(U, v)$ is given by (1) and K be a parameter. Let the Lagrange multipliers be ν_k and λ, and the Lagrange function be $\mathcal{L}$. Then we have

$$\begin{aligned} \mathcal{L} \quad = \quad & -\sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} \log u_{ik} \\ & +\nu_k (\sum_{i=1}^{c} u_{ik} - 1) + \lambda (J_1 - K). \end{aligned}$$

The stationary point of $\mathcal{L}$ is

$$u_{ik} = \frac{e^{-\lambda d_{ik}}}{\sum_{j=1}^{c} e^{-\lambda d_{jk}}}. \tag{4}$$

where $d_{ik} = d(x_k, v_i)$ for simplicity. The multiplier ν_k has already been determined while λ is still in (4). This parameter should be determined so that the constraint $J_1(U, v) = K$ is satisfied. The number K is, however, difficult to determine beforehand.

The last statement implies that the multiplier λ should not be treated as a uniquely determined number, but an adjustable parameter. Thus, actually we consider a regularized problem which is equivalent to the above:

$$\begin{aligned} &\textit{minimize} \quad J_1(U, v) + \lambda^{-1} \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} \log u_{ik} \\ &\textit{subject to} \quad \sum_{i=1}^{c} u_{ik} = 1, \end{aligned}$$

where λ is a positive regularizing parameter. Notice that the number K can be neglected. It is easily seen that the solution of the last problem is given by the same formula of (4). Remark also that in the Euclidean space $d(x_k, v_i) = \|x_k - v_i\|_2^2$, the center $\bar{v} = (\bar{v}_1, ..., \bar{v}_c)$ in **CM** is given by

$$\bar{v}_i = \frac{\sum_{k=1}^{n} \bar{u}_{ik} x_k}{\sum_{k=1}^{n} \bar{u}_{ik}}. \tag{5}$$

When $\bar{u}_{ik}$ is replaced by $(\bar{u}_{ik})^m$, the above formula becomes that of calculating centers in the standard fuzzy c - means.

2.3 Logarithmic transformation of dissimilarity

A simple calculation shows a relationship between (2) and (4). Let $\bar{d}_{ik} = d(x_k, \bar{v}_i)$ and $\alpha = \frac{1}{m-1} > 0$ in (2). Since $a^z = e^{z \log a}$, we have

$$\begin{aligned} \bar{u}_{ik} &= [\sum_{j=1}^{c} (\frac{\bar{d}_{ik}}{\bar{d}_{jk}})^{\alpha}]^{-1} && (6) \\ &= [\sum_{j=1}^{c} e^{\alpha(\log \bar{d}_{ik} - \log \bar{d}_{jk})}]^{-1} && (7) \\ &= \frac{e^{-\alpha \log \bar{d}_{ik}}}{\sum_{j=1}^{c} e^{-\alpha \log \bar{d}_{jk}}}. && (8) \end{aligned}$$

The last expression is similar to (4). Indeed, the equation (4) coincides with (8) by the substitutions $\lambda = \alpha$ and $d_{jk} = \log \bar{d}_{jk}$.

Since the calculations of the centers are different between the both methods, we cannot say that the two methods are equivalent. Nevertheless, the above logarithmic transformation shows that the underlying idea is essentially the same.

3 Methods derived from maximum entropy approach

3.1 Methods of clustering

The maximum entropy approach is not a single method, but is providing a framework in which different algorithms are incorporated. Namely, the consideration in the previous section immediately leads us to a list of new methods of clustering. Some of them are as follows.

(I) As shown by Bezdek and other researchers, there are a number of variations such as the fuzzy c - varieties [2] and fuzzy c - regression [8, 21]. These models can be reconsidered by the maximum entropy approach.

(II) Recently, L_1 and L_p spaces are considered by a number of researchers [3, 9, 18, 19]. Their methods can also be studied within the present framework.

Let us examine these statements in more detail, by showing two examples: one in (I) and the other in (II). In the alternative optimization **CM**, u_{ik} is calculated with a given $\bar{v}$ in **CM2**, and v is calculated with a given $\bar{U}$ in **CM3**. This structure of the algorithm facilitates development of new methods in the maximum entropy framework.

Example 1 (FCV).

Let us take an example in the fuzzy c - varieties, in which the objective function is

$$J(U, w, s) = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ik})^m d_{ik}, \tag{9}$$

where

$$d_{ik} = \|x_k - w_i\|^2 - \langle x_k - w_i, s_i \rangle^2. \tag{10}$$

Namely, the cluster i is represented by a line $w_i + \beta s_i$ where β is a scalar variable, w_i means the center of the cluster, and s_i is the unit vector showing the direction of the line. The line shows the principal axis of the cluster. Thus, the dissimilarity is measured by the distance between an individual and the line.

Consequently, the alternative algorithm requires three steps:

Procedure FCV.

FCV1. Set initial values for $\bar{U}$, $\bar{w}$, $\bar{s}$.

FCV2. Minimize $J(\bar{U}, w, \bar{s})$ with respect to w. Let the optimal solution be $\bar{w}$.

FCV3. Minimize $J(\bar{U}, \bar{w}, s)$ with respect to s. Let the optimal solution be $\bar{s}$.

FCV4. Minimize $J(U, \bar{w}, \bar{s})$ with respect to U. Let the optimal solution be $\bar{U}$.

FCV5. Check stopping criterion. If the criterion is not satisfied, go to FCV2.

It is known that in FCV2,

$$\bar{w}_i = \frac{\sum_{k=1}^{n} (\bar{u}_{ik})^m x_k}{\sum_{k=1}^{n} (\bar{u}_{ik})^m}.$$

In FCV3, let $\bar{s}_i$ be the normalized eigenvector for the maximum eigenvalue of the following matrix

$$A^i = \sum_{\ell=1}^{c} (\bar{u}_{i\ell})^m (x_\ell - \bar{w}_i)(x_\ell - \bar{w}_i)^T.$$

In FCV4,

$$\bar{u}_{ik} = [\sum_{j=1}^{c} (\frac{\bar{d}_{ik}}{\bar{d}_{jk}})^{\frac{1}{m-1}}]^{-1}.$$

as usual. (For simplicity, we write $\bar{d}_{ik} = \|x_k - \bar{w}_i\|^2 - \langle x_k - \bar{w}_i, \bar{s}_i \rangle^2$.)

Now, it is easy to obtain the maximum entropy method in which the objective function L is given by

$$L = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} d_{ik} + \lambda^{-1} \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} \log u_{ik}.$$

The solution in FCV2 is

$$\bar{w}_i = \frac{\sum_{k=1}^{n} \bar{u}_{ik} x_k}{\sum_{k=1}^{n} \bar{u}_{ik}}.$$

In FCV3, let $\bar{s}_i$ be the normalized eigenvector for the maximum eigenvalue of the matrix

$$B^i = \sum_{\ell=1}^{c} \bar{u}_{i\ell} (x_\ell - \bar{w}_i)(x_\ell - \bar{w}_i)^T.$$

Finally, in FCV4,

$$\bar{u}_{ik} = \frac{e^{-\lambda \bar{d}_{ik}}}{\sum_{j=1}^{c} e^{-\lambda \bar{d}_{jk}}}.$$

which is the same formula as (4).

Example 2 (L_1 FCM).

As another example, let us consider the L_1 space based fuzzy c - means [3, 9, 18]. Namely we assume that

$$d(x_k, v_i) = \|x_k - v_i\|_1 = \sum_{\ell=1}^{p} |x_k^\ell - v_i^\ell|, \tag{11}$$

where x_k^ℓ is the ℓth component of the vector x_k. Thus, the dissimilarity is given by the L_1 norm.

Miyamoto and Agusta [18] show a fast algorithm in the step **CM3**. (Remark that u_{ik} is calculated by the same formula (2).) As in the Euclidean case, the solution $\bar{v}_i$ is calculated componentwise. Hence we consider ℓth component alone. Assume that when $\{x_1^\ell, ..., x_n^\ell\}$ is ordered, subscripts are changed using a permutation function $q(k)$, $k = 1, ..., n$, that is, $x_{q(1)}^\ell \leq x_{q(2)}^\ell \leq ... \leq x_{q(n)}^\ell$. Using $\{x_{q(k)}^\ell\}$, the following algorithm calculates $\bar{v}_i^\ell$, in which we put $\gamma_{ik} = (u_{ik})^m$.

Algorithm CC.

begin
  $S := -\sum_{k=1}^{n} \gamma_{ik}$;
  $r := 0$;
  while ($S < 0$)do begin
    $r := r + 1$;
    $S := S + 2\gamma_{iq(r)}$
  end;
  output $v_i^\ell = x_{q(r)}^\ell$ as the ℓth coordinate of
  the cluster center $\bar{v}_i$
end.

Now, we can develop a maximum entropy algorithm in which the objective function is

$$J_1(U, v) + \lambda^{-1} \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik} \log u_{ik},$$

where $J_1(U, v)$ is defined by (1) and the dissimilarity given by (11). By a discussion parallel to the previous work [18], the center is calculated by the algorithm CC *with the replacement* $\gamma_{ik} = u_{ik}$. Notice that the membership u_{ik} is calculated by (4). It is obvious that the latter method has the advantage of less calculation, since calculation of the power m in $(u_{ik})^m$ is necessary in the former method.

These considerations show that a variety of other methods can easily be transformed into the corresponding algorithms in the framework of the maximum entropy.

3.2 Fuzzy classifiers

Clustering is frequently called unsupervised classification, whereas supervised classification means determination of classification rules given a number of classes of individuals. Frequently the latter is simply referred to as classification problem. There are various methods for this class of problems, which include classical descriminant analysis [11] as well as the nonparametric technique of the k - nearest neighbor method (kNN).

Fuzzy supervised classification techniques have been studied as extensions of crisp nonparametric methods. In particular, a fuzzy kNN and a method of fuzzy nearest prototype has been studied by Keller *et al.* [13].

Their fuzzy kNN classifier, which assigns the membership of x in the class i, is given by

$$u_j(x) = \sum_{i=1}^{k} u_{ji} [\sum_{\ell=1}^{k} (\frac{d(x,y_i)}{d(x,y_\ell)})^{\frac{1}{m-1}}]^{-1} \quad (12)$$

where $y_1, ..., y_k$ are k nearest individuals from x, and $d(x,y_i) = \|x - y_i\|_2^2$.

Moreover, the fuzzy nearest prototype classifier therein is

$$u_j(x) = [\sum_{\ell=1}^{c} (\frac{d(x,z_j)}{d(x,z_\ell)})^{\frac{1}{m-1}}]^{-1} \quad (13)$$

where z_i $(i = 1, ..., c)$ are prototype vectors for the class i.

It is immediate to see that these classifiers are closely related to the rule (2) of determining memberships in the standard fuzzy c - means.

Now, we obtain, by analogy, fuzzy classifiers from the maximum entropy principle:

(I) Fuzzy kNN classifier:

$$u_j(x) = \frac{\sum_{i=1}^{k} u_{ji} e^{-\lambda d(x,y_i)}}{\sum_{\ell=1}^{k} e^{-\lambda d(x,y_\ell)}}. \quad (14)$$

($y_1, ..., y_k$ are k nearest individuals from x.)

(II) Fuzzy prototype classifier:

$$u_j(x) = \frac{e^{-\lambda d(x,z_j)}}{\sum_{\ell=1}^{c} e^{-\lambda d(x,z_\ell)}}. \quad (15)$$

(z_i, $i = 1, ..., c$, are prototype vectors for the class i.)

An advantage of the present classifiers is that the combination of exponential functions provides a clear view for behaviors of the function. For example, the prototype classifier satisfies

$$u_j(x)^{-1} = 1 + \sum_{\ell \neq j} e^{-\lambda (d(x,z_\ell) - d(x,z_j))}. \quad (16)$$

When λ is large enough, the fuzzy classifier is shown to approach the crisp nearest prototype classifier by which x is classified into the class of the nearest prototype. To see this, let z_j be the nearest prototype to x, then the terms $d(x,z_\ell) - d(x,z_j)$ are all positive. When $\lambda \to 0$, the second term in (16) becomes sufficiently small and hence $u_j(x) \to 1$. From $\sum_\ell u_\ell(x) = 1$, we have $u_\ell(x) \to 0$ $(\ell \neq j)$. Thus, the classifier tends to crisp one. It should be remarked that this property is obtained from the assumption that the parameter $\lambda > 0$ is *adjustable*.

Another observation is that the above classifiers are not limited in the Euclidean space, since $d(x,v)$ may be defined in terms of other spaces such as the L_1 space.

Moreover, we can derive fuzzy classifiers based on fuzzy c - varieties by using

$$u_j(x) = \frac{e^{-\lambda d(x,V_j)}}{\sum_{\ell=1}^{c} e^{-\lambda d(x,V_\ell)}}, \quad (17)$$

where V_j, $j = 1, ..., c$, is not a single vector, but a set of parameters characterizing a variety. In relation to the FCV clustering shown as Example 1, $d(x,V_j)$ is the distance from x to the line representing the class.

4 Methods related to regularization

Regularization is somewhat similar to the penalty method for optimization by considering the regularizing functional to be a penalty. Another relation between regularization and bicriteria optimization is observed. Namely, the regularized objective function $J_\alpha = J + \alpha K$ keeps balance between the original objective J and the regularizing functional αK.

The latter idea of the bicriteria optimization has been used in fuzzy clustering [24]. Combining their method and the consideration herein, we have a new algorithm using the maximum entropy. The detailed description is omitted here.

Another example of bicriteria fuzzy clustering is the possibilistic approach by Krishnapuram and Keller [14]. They remove the assumption of the fuzzy partition $\sum_{i=1}^{c} u_{ik} = 1$ and consider an objective function

$$J = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ik})^m d_{ik} + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} (1 - u_{ik})^m.$$

Is should be noticed that the second term is introduced in order to *regularize* the objective function: without this term the solution becomes trivial. At the same time the first and second terms should be balanced in the optimization, and hence the method is regarded as a bicriteria clustering algorithm.

An interesting exercise is to design a regularized objective function using the entropy that leads us to a very simple membership allocation scheme. See Appendix.

5 Conclusions

We have presented a number of methods that are parallel to existing methods within the framework of maximum entropy. Thus, the major purpose of the present paper is to show potential of this framework or approach. Therefore examination of performances of the individual methods presented here by numerical simulations and real data will be done as further works. In this sense, the present paper is still a preliminary

report, and more studies should be done using this approach of the maximum entropy.

The crucial point in the present approach is that the parameter λ is rather free to choose, whereas $m = 2$ has frequently been used in the standard FCM, since this value simplifies the calculation. From the viewpoint of regularization, a regularizing parameter should not uniquely be determined. For example, properties of a regularized method can be made clear by moving the regularizing parameter, as shown in Section 3.2.

For some readers it seems strange to use the maximum entropy for clustering, since the entropy criterion has sometimes been referred to in validation of clusters [2]. Actually we abandan the use of entropy in validating clusters by employing the entropy criterion in cluster generation. Thus, the relationship between the criterion for clustering and that for validation is delicate in general.

As another direction for future studies, mixture with other methods and the approach here is promising. There have been studies of clustering using Kohonen's associative memory [12]. Pal, Bezdek, and Tsao [22] develop a crisp clustering algorithm GLVQ based on the learning vector quantization. Fuzzy version of this algorithm using regularization should be attempted.

Since the membership allocation rule obtained by the maximum entropy has a simple form of exponential functions, there are rooms for observing many theoretical properties. Thus, the present approach provides a new viewpoint for fuzzy clustering which has many possibilities for future studies.

Acknowledgment

This research has partly been supported by TARA (Tsukuba Advanced Research Alliance), the University of Tsukuba.

Appendix

Let us consider an objective function that leads us to a simple membership allocation scheme of an exponential type. Put

$$L = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}d_{ik} + \sum_{i=1}^{c}\zeta_i^{-1}\sum_{k=1}^{n} u_{ik}(\log u_{ik} - 1),$$

in which $\sum_{i=1}^{c} u_{ik} = 1$ is not assumed. From

$$\frac{\partial L}{\partial u_{ik}} = d_{ik} + \zeta_i^{-1}\log u_{ik} = 0$$

we have

$$u_{ik} = e^{-\zeta_i d_{ik}}.$$

References

[1] M. R. Anderberg, *Cluster Analysis for Applications*, Academic Press, New York, 1973.

[2] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York, 1981.

[3] L. Bobrowski and J. C. Bezdek, c-means clustering with the ℓ_1 and ℓ_∞ norms, *IEEE Trans. on Systems, Man, and Cybern.*, Vol. 21, pp. 545–554, 1991.

[4] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

[5] J. C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *J. of Cybernetics*, Vol.3, pp. 32–57, 1974.

[6] J. C. Dunn, Well-separated clusters and optimal fuzzy partitions, *J. of Cybernetics*, Vol.4, pp. 95–104, 1974.

[7] E. W. Forgy, Cluster analysis of multivariate data: efficiency vs. interpretability of classifications, *Biometrics*, Vol.21, pp. 768–769 (Abstract) 1965.

[8] R. J. Hathaway and J. C. Bezdek, Switching regression models and fuzzy clustering, *IEEE Trans. on Fuzzy Syst.*, Vol.1, pp. 195–204, 1993.

[9] K. Jajuga, L_1-norm based fuzzy clustering, *Fuzzy Sets and Systems*, Vol. 39, pp. 43–50, 1991.

[10] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol. 2, 4th Ed.*, Griffin, London, 1979.

[11] M. G. Kendall, A. Stuart, and J. K. Ord, *The Advanced Theory of Statistics, Vol. 3, 4th Ed.*, Griffin, London, 1983.

[12] T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, Heiderberg, 1989.

[13] J. M. Keller, M. R. Gray, and J. A. Givens, Jr., A fuzzy k - nearest neighbor algorithm, *IEEE Trans., on Syst., Man, and Cybern.*, Vol. 15, pp. 580–585, 1985.

[14] R. Krishnapuram and J. M. Keller, A possibilistic approach to clustering, *IEEE Trans. on Fuzzy Syst.*, Vol.1, pp. 98–110, 1993.

[15] R.-P. Li and M. Mukaidono, Gaussian clustering and its application to rock classification, *Proc. of Eleventh Fuzzy System Symposium*, Japan Society of Fuzzy Theory and Systems, pp .697–698, 1995.

[16] R.-P. Li and M. Mukaidono, A maximum entropy approach to fuzzy clustering, *Proc. of the 4th IEEE Intern. Conf. on Fuzzy Systems (FUZZ-IEEE/IFES'95)*, Yokohama, Japan, March 20-24, 1995, pp. 2227–2232.

[17] J. B. McQueen, Some methods of classification and analysis of multivariate observations, *Proc. of 5th Berkeley Symposium on Math. Stat. and Prob.*, pp. 281–297, 1967.

[18] S. Miyamoto and Yudi Agusta, An efficient algorithm for ℓ_1 fuzzy c-means and its termination, *Control and Cybernetics* Vol. 24, pp. 421–436, 1995.

[19] S. Miyamoto and Yudi Agusta, Efficient algorithms for ℓ_p fuzzy c-means clustering and their termination properties. *Fifth Conference of the International Federation of Classification Societies (IFCS 96)*, Kobe, Japan, March 27-30, Vol.2, pp. 255–258, 1996.

[20] S. Miyamoto, Fuzzy c - means and its variations, *Journal of the Japan Society of Fuzzy Theory and Systems*, Vol.8, pp. 423–430, 1996 (in Japanese).

[21] Y. Nakamori, Fuzzy clustering and regression analysis, *Journal of the Japan Society of Fuzzy Theory and Systems*, Vol.8, pp. 431–439, 1996 (in Japanese).

[22] N. K. Pal, J. C. Bezdek, and E. C.-K. Tsao, Generalized clustering networks and Kohonen's self-organizing scheme, *IEEE Trans. on Neural Networks*, Vol.4, pp. 549–557, 1993.

[23] A. N. Tihonov, Solutions of incorrectly formulated problems and the regularization method, *Dokl. Akad. Nauk. SSSR*, Vol.151, pp. 1035–1038, 1963.

[24] H.-F. Wang and G.-Y. Wu, Bi-criteria fuzzy clustering systems, *IV IFSA World Congress*, São Paulo, Brazil, 1995, Vol.1, pp. 633–636, 1995.