Several Formulations for Graded Possibilistic Approach to Fuzzy Clustering

Katsuhiro Honda[1], Hidetomo Ichihashi[1], Akira Notsu[1],
Francesco Masulli[2], and Stefano Rovetta[3]

[1] Graduate School of Engineering, Osaka Prefecture University,
1-1 Gakuen-cho, Sakai, Osaka, 599-8531, Japan
{honda,ichi,notsu}@cs.osakafu-u.ac.jp
http://www.cs.osakafu-u.ac.jp/hi/index.html
[2] Department of Computer Science, Unversity of Pisa,
Via F. Buonarroti 2 - 56127 Pisa, Italy
masulli@disi.unige.it
[3] Department of Computer and Information Sciences,
Unversity of Genova, Via Dodecaneso 35 - 16146 Genova, Italy
ste@disi.unige.it

Abstract. Fuzzy clustering is a useful tool for capturing intrinsic structure of data sets. This paper proposes several formulations for soft transition of fuzzy memberships from probabilistic partition to possibilistic one. In the proposed techniques, the free memberships are given by introducing additional penalty term used in Possibilistic c-Means. The new features of the proposed techniques are demonstrated in several numerical experiments.

1 Introduction

Fuzzy c-Means (FCM) [1] is a well known fuzzy clustering algorithm whose goal is to partition objects (data points, individuals) into several clusters by estimating fuzzy memberships of objects to each cluster. In the FCM clustering, an additional weighting parameter called "fuzzifier" is introduced into the objective function of (hard) k-Means clustering [2] in order to fuzzify the degree of memberships, i.e., the constraints on the membership parameters are generalized so that they can take arbitrary values from the interval of $[0, 1]$ instead of $\{0, 1\}$. Because the sum of memberships of objects with respect to clusters are constrained to be 1, the fuzzy memberships are often said to be "probabilistic".

Recently, several other techniques for fuzzifying membership assignment have been proposed based on regularization approaches. Miyamoto and Mukaidono [3] considered the singularity in the hard clustering which implies the case where proper partition is not obtained by the Lagrangian multiplier method, and introduced an entropy term as the regularization term with a positive parameter into the objective function of k-Means clustering. Because the fuzzification technique derives the similar algorithm to that of entropy-constrained fuzzy clustering by Deterministic Annealing (DA) [4], the clustering model is often compared

with probabilistic mixture models [5]. Then, Ichihashi *et al.* [6] proposed a clustering algorithm, which is similar to the EM algorithm for Gaussian Mixture Models (GMMs), by using the regularization technique with Kullback-Leibler divergences (K-L information).

In spite of its usefulness, the "probabilistic" partition has a problem of noise sensitivity because the probabilistic constraint forces noise samples to belong to one or several clusters with some degree. Then, Krishnapuram and Keller proposed Possibilistic *c*-Means (PCM) [7] by giving up the constraint of sum to 1. This mode seeking algorithm is useful for outlier rejection in fuzzy membership assignment, and the memberships can be regarded as the probability that an experimental outcome coincides with one of mutually independent events. However, it is possible that sets of events are neither mutually independent nor completely mutually exclusive. Then, Masulli and Rovetta [8, 9] proposed the graded possibilistic approach to the FCM clustering with regularization by entropy. In the approach, soft transition from probabilistic to possibilistic partition is performed by using the graded possibilistic constraint.

This paper proposes two other formulations for soft transition of fuzzy memberships from probabilistic partition to possibilistic one. One is a modified version of the original FCM algorithm, in which the updating rule for memberships is a hybrid of FCM and PCM. The other is an enhanced version of the FCM algorithm with regularization by K-L information. In the proposed techniques, the free memberships drawn from the interval of $[0, 1]$ are given by introducing an additional penalty term used in PCM. The new features of the proposed techniques are demonstrated in several numerical experiments.

2   Fuzzy *c*-Means and Possibilistic *c*-Means

2.1   Fuzzy *c*-Means and Several Fuzzification Techniques

Fuzzy *c*-Means (FCM) [1] is an unsupervised classification technique that is a fuzzified version of *k*-Means clustering [2]. In the *k*-Means (hard *c*-Means) clustering, objects to be classified are assigned to one of C clusters where each cluster has its prototypical mean vector. The membership assignment is based on minimization of within-group-sum-of-errors, i.e., nearest prototype classification. Then, the two step iterative algorithm is composed of calculation of mean vectors and assignment of objects.

The FCM algorithm proposed by Bezdek *et al.* [1] uses the objective function of generalized within-group-sum-of-errors

$$L_{fcm}^s = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}^\theta d_{ci}^2. \tag{1}$$

d_{ci}^2 is the clustering criterion of the distance between the ith object $\boldsymbol{x}_i$ and the cth prototypical mean vector (cluster center) $\boldsymbol{b}_c$, and $u_{ci} \in [0, 1]$ represents the membership of the ith object to the cth cluster. θ is an additional weighting

exponent. If $\theta = 1$, the clustering model is reduced to the (hard) k-Means model. The larger the θ, the fuzzier the memberships. So, the weighting exponent is usually set to be $\theta > 1$ and is called the "fuzzifier". The memberships are often calculated under the constraint of

$$\sum_{c=1}^{C} u_{ci} = 1, \qquad i = 1, \cdots, n. \tag{2}$$

Because the memberships are obtained by a formula similar to the updating rule for posterior probabilities in the EM algorithm with probabilistic mixture models, the constraint is called the "probabilistic constraint" [10].

In the original FCM algorithm, called the "standard FCM algorithm", the updating rules for parameters are given as

$$\boldsymbol{b}_c = \frac{\sum_{i=1}^{n} u_{ci}^{\theta} \boldsymbol{x}_i}{\sum_{i=1}^{n} u_{ci}^{\theta}}, \tag{3}$$

$$u_{ci} = \left[\sum_{l=1}^{C} \left(\frac{d_{ci}^2}{d_{li}^2} \right)^{\frac{1}{\theta-1}} \right]^{-1}. \tag{4}$$

Another approach to fuzzification of the hard c-Means clustering is the regularization of the objective function. Miyamoto and Mukaidono [3] introduced a regularization term with a positive parameter λ into the objective function. Using the entropy term, the objective function of the FCM clustering based on the regularization technique is defined as

$$L_{fcm}^e = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} d_{ci}^2 + \lambda \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci} \log u_{ci}, \tag{5}$$

where the entropy term works like the weighting exponent in the standard FCM algorithm, and transforms the linear programming problem into the nonlinear optimization problem with respect to memberships u_{ci}. The parameter λ plays a role for tuning the degree of fuzziness of membership values. The larger the λ, the fuzzier the memberships. This fuzzification technique is called the "regularization by entropy." The updating rules for memberships and cluster centers are derived as follows:

$$u_{ci} = \frac{\exp(-\frac{1}{\lambda} d_{ci}^2)}{\sum_{l=1}^{c} \exp(-\frac{1}{\lambda} d_{li}^2)}, \tag{6}$$

$$\boldsymbol{b}_c = \frac{\sum_{i=1}^{n} u_{ci} \boldsymbol{x}_i}{\sum_{i=1}^{n} u_{ci}}. \tag{7}$$

The regularization approach can also be performed by using other regularization terms. Ichihashi et $al.$ [6] generalized the regularized objective function

replacing the entropy term with K-L information term and proposed an FCM-type counterpart of the GMMs with full unknown parameters. The clustering technique is called the FCM clustering with regularization by K-L information (KLFCM) and the objective function is defined as follows:

$$L_{klfcm} = \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}d_{ci}^2 + \lambda \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \log \frac{u_{ci}}{\pi_c} + \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \log |\Sigma_c|, \quad (8)$$

where d_{ci}^2 is the (squared) Mahalanobis distance $d_{ci}^2 = (\boldsymbol{x}_i-\boldsymbol{b}_c)^\top \Sigma_c^{-1}(\boldsymbol{x}_i-\boldsymbol{b}_c)$,and all the elements of Σ_c are also decision variables. Eq.(8) is minimized under the condition that both the sum of u_{ci} and the sum of π_c with respect to c equal 1, respectively. If $u_{ci} \simeq \pi_c$ for all i and c, the K-L information term becomes 0 and the membership assignment is very fuzzy; but when λ is 0 the solution u_{ci}'s are obtained at the extremal point (0 or 1). Fuzziness of the partition can be controlled by λ. From the necessary conditions, the updating rules for u_{ci}, π_c, Σ_c are given as follows:

$$u_{ci} = \frac{\pi_c \exp\left(-\frac{1}{\lambda}d_{ci}\right)|\Sigma_c|^{-\frac{1}{\lambda}}}{\sum_{l=1}^{C} \pi_l \exp\left(-\frac{1}{\lambda}d_{li}\right)|\Sigma_l|^{-\frac{1}{\lambda}}}, \quad (9)$$

$$\pi_c = \frac{1}{n}\sum_{i=1}^{n} u_{ci}, \quad (10)$$

$$\Sigma_c = \frac{\sum_{i=1}^{n} u_{ci}(\boldsymbol{x}_i - \boldsymbol{b}_c)(\boldsymbol{x}_i - \boldsymbol{b}_c)^\top}{\sum_{i=1}^{n} u_{ci}}, \quad (11)$$

and the cluster center $\boldsymbol{b}_c$ is given by Eq.(7). Because π_c represents the proportion of objects belonging to the cth cluster, it is regarded as the capacity of the cluster. The algorithm is equivalent to the EM algorithm with GMMs if and only if the fuzzification coefficient $\lambda = 2$. When $\lambda \neq 2$, there is no corresponding mixture density. In the KLFCM clustering, K-L information term is used for both optimization of cluster capacities and fuzzification of memberships while Hathaway [5] interpreted the clustering criterion as the sum of K-L information for updating memberships [11].

2.2 Possibilistic c-Means

Krishnapuram and Keller [7] proposed the possibilistic clustering by giving up the probabilistic constraint. The objective function is formulated as

$$L_{pcm}^s = \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci}^\theta d_{ci}^2 + \sum_{c=1}^{C} \eta_c \sum_{i=1}^{n} (1 - u_{ci})^\theta. \quad (12)$$

where $\eta_c, c = 1, \cdots, C$ are suitable positive numbers. The first term demands that the distances from the objects to the prototypes be as low as possible,

whereas the second term forces the u_{ci} to be as large as possible avoiding the trivial solution. Then, the updating rule for memberships is given as

$$u_{ci} = \frac{1}{1 + \left(\frac{d_{ci}^2}{\eta_c}\right)^{\frac{1}{\theta-1}}}. \tag{13}$$

The value of η_c controls the bandwidth of the possibility (membership) distribution for each cluster and determines the distance at which the membership value of an object in a cluster becomes 0.5. In [7], η_c was calculated by using the fuzzy intra-cluster distance as

$$\eta_c = K \frac{\sum_{i=1}^n u_{ci}^\theta d_{ci}^2}{\sum_{i=1}^n u_{ci}^\theta}, \tag{14}$$

and K is typically chosen to be 1.

Another formulation for Possibilistic c-Means can be derived by modifying the FCM clustering with regularization by entropy [12, 13],

$$L_{pcm}^e = \sum_{c=1}^C \sum_{i=1}^n u_{ci} d_{ci}^2 + \lambda \sum_{c=1}^C \sum_{i=1}^n (u_{ci} \log u_{ci} - u_{ci}) \tag{15}$$

The updating rule for memberships is given as $u_{ci} = \exp(-\frac{1}{\lambda} d_{ci}^2)$.

3   Soft Transition to Possibilistic Partition

3.1   DA-Based Soft Transition

Masulli and Rovetta [8, 9] proposed the graded possibilistic approach to clustering, in which soft transition from probabilistic to possibilistic constraint is performed by using the graded possibilistic constraint. Assume that a class of constraints is expressed by a unified formulation: $\Psi = \sum_{c=1}^C u_{ci}^{[\xi]} - 1$, where $[\xi]$ is an interval variable representing an arbitrary real number included in the range $[\underline{\xi}, \overline{\xi}]$, i.e., there must exist a scalar exponent $\xi^* \in [\underline{\xi}, \overline{\xi}]$ such that the equality $\Psi = 0$ holds. The constraint can be implemented by using a running parameter α. The extrema of the interval are written as a function of α, where $\underline{\xi} = \alpha, \overline{\xi} = \frac{1}{\alpha}$ and $\alpha \in [0, 1]$. Then, the constraint with an interval is represented as a set of two inequalities: $\sum_{c=1}^C u_{ci}^\alpha \geq 1$ and $\sum_{i=1}^C u_{ci}^{\frac{1}{\alpha}} \leq 1$.

For implementation of the graded possibilistic clustering, the following algorithm can be used. When we use the entropy regularization (or DA approach), the memberships are updated as

$$u_{ci} = \frac{\phi_{ci}}{\kappa_i}, \tag{16}$$

where ϕ_{ci} is a free membership of $\boldsymbol{x}_i$ to the cth cluster drawn from the interval of $[0, 1]$ and is given as

$$\phi_{ci} = \exp\left\{-\frac{d_{ci}^2}{\lambda}\right\}. \tag{17}$$

κ_i takes one of the following three values.

$$\kappa_i = \left(\sum_{l=1}^{C} \phi_{li}^{\frac{1}{\alpha}} \right)^{\alpha} \qquad \text{if} \sum_{l=1}^{C} \phi_{li}^{\frac{1}{\alpha}} > 1, \tag{18}$$

$$\kappa_i = \left(\sum_{l=1}^{C} \phi_{li}^{\alpha} \right)^{\frac{1}{\alpha}} \qquad \text{if} \sum_{l=1}^{C} \phi_{li}^{\alpha} < 1, \tag{19}$$

$$\kappa_i = 1 \qquad \qquad \text{else.} \tag{20}$$

Eq.(18) transforms memberships that are above the upper boundary onto the boundary ($\sum_{l=1}^{C} u_{li}^{\frac{1}{\alpha}} = 1$) while Eq.(19) transforms memberships under the lower boundary onto the boundary ($\sum_{l=1}^{C} u_{li}^{\alpha} = 1$). When $\alpha = 1$, Eqs.(18) and (19) derive $\kappa_i = \sum_{l=1}^{C} \phi_{li}$, and memberships u_{ci}'s are reduced to the probabilistic ones ($\sum_{c=1}^{C} u_{ci} = 1$). On the other hand, $\alpha = 0$ provides the possibilistic membership assignment because all of κ_i's are given by Eq.(20). Then, the value of α should be gradually decreased from 1 to 0.

3.2   Soft Transition with Standard Fuzzifiation Technique

The key in generalizing the graded possibilistic approach to other fuzzification techniques is how to define the free memberships of Eq.(17). In this subsection, a free membership drawn from the interval of $[0, 1]$ is formulated in the frame of the standard fuzzification technique. In order to generalize the updating rule for memberships, the additional penalty term used in PCM is introduced into the objective function of FCM as follows:

$$L_{fcm}^{gs} = \sum_{c=1}^{C} \sum_{i=1}^{n} u_{ci}^{\theta} d_{ci}^2 + \sum_{c=1}^{C} \eta_c \sum_{i=1}^{n} (1 - u_{ci})^{\theta}. \tag{21}$$

Then, the free membership ϕ_{ci} is given as

$$\phi_{ci} = \frac{1}{1 + \left(\frac{d_{ci}^2}{\eta_c} \right)^{\frac{1}{\theta-1}}}, \tag{22}$$

where η_c is a predefined constant. We can see that $d_{ci}^2 = 0$ derives $\phi_{ci} = 1$, and ϕ_{ci} moves toward 0 as $d_{ci}^2 \to \infty$.

Here, it is obvious that the free membership of Eq.(22) derives the possibilistic partition of PCM when $\alpha = 0$. Then, the parameter η_c plays a similar role with that of PCM and can be given in the same way with PCM. On the other hand, when $\alpha = 1$, the updating rule has some connection with that of FCM. Substituting Eq.(22), Eq. (16) with $\alpha = 1$ is written as

$$u_{ci} = \frac{\phi_{ci}}{\sum_{l=1}^{C} \phi_{li}} = \frac{\frac{(d_{ci}^2)^{-\frac{1}{\theta-1}}}{(d_{ci}^2)^{-\frac{1}{\theta-1}} + (\eta_c)^{-\frac{1}{\theta-1}}}}{\sum_{l=1}^{C} \frac{(d_{li}^2)^{-\frac{1}{\theta-1}}}{(d_{li}^2)^{-\frac{1}{\theta-1}} + (\eta_c)^{-\frac{1}{\theta-1}}}}. \tag{23}$$

By the way, the updating rule of Eq.(4) can be written as

$$u_{ci} = \frac{(d_{ci}^2)^{-\frac{1}{\theta-1}}}{\sum_{l=1}^C (d_{li}^2)^{-\frac{1}{\theta-1}}} = \frac{\frac{(d_{ci}^2)^{-\frac{1}{\theta-1}}}{\sum_{l=1}^C (d_{li}^2)^{-\frac{1}{\theta-1}}}}{\sum_{l=1}^C \frac{(d_{li}^2)^{-\frac{1}{\theta-1}}}{\sum_{k=1}^C (d_{ki}^2)^{-\frac{1}{\theta-1}}}}. \tag{24}$$

Then, the proposed model is equivalent to the standard FCM if $(\eta_c)^{-\frac{1}{\theta-1}} = \sum_{l=1}^C (d_{li}^2)^{-\frac{1}{\theta-1}} - (d_{ci}^2)^{-\frac{1}{\theta-1}}$ for all i.

Assume that η_c is fixed for all i in each cluster. When η_c is large, all u_{ci}'s tend to take the same value of $1/C$, i.e., the membership assignment becomes very fuzzy. On the other hand, when η_c is enough close to 0, the value of Eq.(23) becomes similar to that of Eq.(24) because

$$\frac{(d_{ci}^2)^{-\frac{1}{\theta-1}}}{(d_{ci}^2)^{-\frac{1}{\theta-1}} + (\eta_c)^{-\frac{1}{\theta-1}}} \simeq \frac{(d_{ci}^2)^{-\frac{1}{\theta-1}}}{(\eta_c)^{-\frac{1}{\theta-1}}}, \tag{25}$$

and

$$\sum_{l=1}^C \frac{(d_{li}^2)^{-\frac{1}{\theta-1}}}{(d_{li}^2)^{-\frac{1}{\theta-1}} + (\eta_c)^{-\frac{1}{\theta-1}}} \simeq \frac{1}{(\eta_c)^{-\frac{1}{\theta-1}}} \sum_{l=1}^C (d_{li}^2)^{-\frac{1}{\theta-1}}. \tag{26}$$

Therefore, the value of η_c should be gradually increased from a positive small value to Eq.(14).

3.3   Soft Transition with Regularization by K-L Information

In the same manner, the graded possibilistic approach to the FCM clustering with regularization by K-L information can be formulated as follows:

$$L_{klfcm}^g = \sum_{c=1}^C \sum_{i=1}^n u_{ci} d_{ci}^2 + \lambda \sum_{c=1}^C \sum_{i=1}^n u_{ci} \log \frac{u_{ci}}{\pi_c} + \sum_{c=1}^C \sum_{i=1}^n u_{ci} \log |\Sigma_c|$$
$$+ \sum_{c=1}^C \lambda \log \eta_c \sum_{i=1}^n (1 - u_{ci}) + \lambda \sum_{c=1}^C \sum_{i=1}^n (1 - u_{ci}) \log(1 - u_{ci}), \tag{27}$$

where d_{ci}^2 is the (squared) Mahalanobis distance. Then, the free membership ϕ_{ci} is given as

$$\phi_{ci} = \frac{\pi_c \exp\left(-\frac{1}{\lambda} d_{ci}\right) |\Sigma_c|^{-\frac{1}{\lambda}}}{\pi_c \exp\left(-\frac{1}{\lambda} d_{ci}\right) |\Sigma_c|^{-\frac{1}{\lambda}} + \frac{1}{\eta_c}}, \tag{28}$$

where η_c is a predefined constant.

Here, the free membership of Eq.(28) derives a possibilistic partition when $\alpha = 0$, and the parameter η_c plays a similar role with that of PCM. Considering

the similarity between KLFCM and GMMs, it is a natural choice that $\frac{1}{\eta_c} = K\pi_c \exp\left(-\frac{1}{\lambda}\right)|\Sigma_c|^{-\frac{1}{\lambda}}$ because the within-group variance is 1 in the Mahalanobis distance. Then K can be chosen to be 1. On the other hand, when $\alpha = 1$, the updating rule becomes more similar to that of KLFCM as η_c moves toward 0. Therefore, the value of η_c should be gradually increased from a positive small value to $\left[K\pi_c \exp\left(-\frac{1}{\lambda}\right)|\Sigma_c|^{-\frac{1}{\lambda}}\right]^{-1}$ in the graded possibilistic clustering.

By the way, other parameters can be updated in the same manner with KLFCM so long as we use the same constraint except for the probabilistic constraint of Eq.(2).

4   Numerical Experiments

This section shows the results of numerical experiments that were performed by using an artificial data set consisting of 100 samples with 2-D observations. The data set was partitioned into 2 clusters using the proposed formulations.

First, the standard FCM and its generalized model were applied to the data set. The derived fuzzy classification functions are shown in Figs. 1 and 2, in which the objects to be classified are represented by "○" and the gray scale shows the maximum membership value, i.e., the membership degree belonging to the nearest cluster center. In the graded possibilistic approach, η_c was given as $\eta_c = 0.0001\alpha + (1-\alpha)\frac{\sum_{i=1}^{n} u_{ci}^{\theta} d_{ci}^2}{\sum_{i=1}^{n} u_{ci}^{\theta}}$. The figures show that the possibilistic partition is a good property of the mode seeking algorithm and the graded possibilistic approach performs the soft transition well. Furthermore, the intermediate model ($\alpha = 0.5$) reflects the features of both of probabilistic and possibilistic partition.

Next, the graded possibilistic approach is performed using the regularization by KL information. Here, it has been shown that the KLFCM algorithm is sensitive to initial partition and often falls into local minima where we have a very large cluster (global cluster) and several very small clusters with a few objects. In this experiment, the initial partition was given by the standard FCM algorithm and the upper limit of variance (covariance) was set as 0.02 in order to avoid a global cluster. The derived fuzzy classification functions are shown in Figs. 3 and 4. η_c was given as $\eta_c = 0.0001\alpha + (1-\alpha)\left[\pi_c \exp\left(-\frac{1}{\lambda}\right)|\Sigma_c|^{-\frac{1}{\lambda}}\right]^{-1}$. The figures show that the KLFCM-based model could capture not only the cluster centers but also the capacities and the shapes of clusters.

In this way, the proposed approach is useful for performing the soft transition from probabilistic to possibilistic partition.

5   Conclusion

This paper proposed several formulations for graded possibilistic approach in the FCM clustering. In the proposed techniques, the free memberships drawn from the interval of $[0, 1]$ are given by introducing an additional penalty term used in
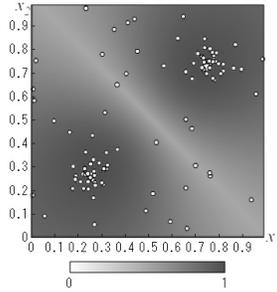
Fig. 1. Fuzzy classification function by standard FCM with $\theta = 2.0$

(a) $\alpha = 0.5$ (b) $\alpha = 0.0$
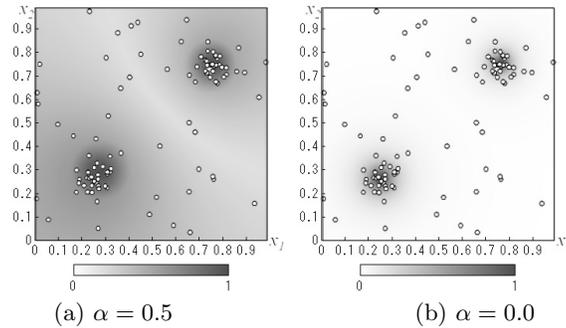
Fig. 2. Fuzzy classification function by graded possibilistic approach with standard fuzzification with $\theta = 2.0$

PCM. The probabilistic partition of the conventional clustering algorithm can be derived by using a small penalty weight while the weight plays a role for tuning the bandwidth of the possibility (membership) distribution for each cluster in the possibilistic partition. Application to real world data sets is remained in future work.

References

1. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press (1981)
2. MacQueen, J.B.: Some Methods of Classification and Analysis of Multivariate Observations. Proc. 5th Berkeley Symposium on Math. Stat. and Prob. (1967) 281–297
3. Miyamoto, S., Mukaidono, M.: Fuzzy c-Means as a Regularization and Maximum Entropy Approach. Proc. 7th Int. Fuzzy Syst. Assoc. World Cong. **2** (1997) 86–92
4. Rose, K., Gurewitz, E., Fox, G.: A Deterministic Annealing Approach to Clustering. Pattern Recognition Letters **11** (1990) 589–594
5. Hathaway, R.J.: Another Interpretation of the EM Algorithm for Mixture Distributions. Statistics & Probability Letters **4** (1986) 53–56
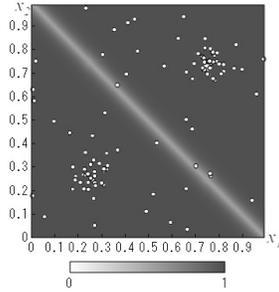
Fig. 3. Fuzzy classification function by KLFCM with $\lambda = 2.0$

(a) $\alpha = 0.1$ (b) $\alpha = 0.0$
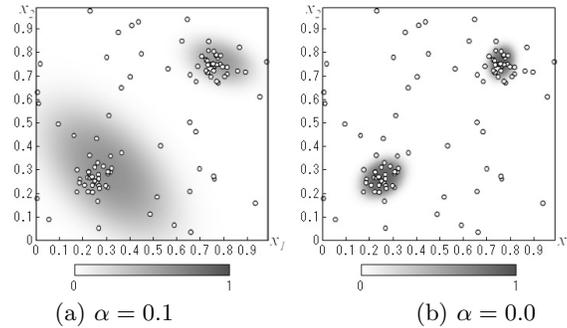
Fig. 4. Fuzzy classification function by graded possibilistic approach with regularization by KL information with $\lambda = 2.0$

6. Ichihashi, H., Miyagishi, K., Honda, K.: Fuzzy c-Means Clustering with Regularization by K-L Information. Proc. of 10th IEEE Int. Conf. Fuzzy Systems **3** (2001) 924–927
7. Krishnapuram, R., Keller, J.M.: A Possibilistic Approach to Clustering. IEEE Trans. Fuzzy Systems **1** (1993) 98–110
8. Masulli, F., Rovetta, S.: The Graded Possibilistic Clustering Model. IJCNN 2003 Conference Proceedings (2003) 791–796
9. Masulli, F., Rovetta, S.: Soft Transition from Probabilistic to Possibilistic Fuzzy Clustering. IEEE Trans. Fuzzy Systems (in press)
10. Höppner, F., Klawonn, F., Kruse, R., Runkler, T.: Fuzzy Cluster Analysis, Jhon Wiley & Sons (1999)
11. Honda, K., Ichihashi, H.: Regularized Linear Fuzzy Clustering and Probabilistic PCA Mixture Models. IEEE Trans. Fuzzy Systems, **13** (2005) 508–516
12. Davé, R.N., Krishnapuram, R.: Robust Clustering Methods: A Unified View. IEEE Trans. Fuzzy Systems, **5** (1997) 270–293
13. Shibuya, K., Miyamoto, S., Takata, O., Umayahara, K.: Regularization and Constraints in Fuzzy c-Means and Possibilistic Clustering. J. Japan Society for Fuzzy Theory and Sysmtems, **13** (2001) 707–715 (in Japanese)