

Kernelized Fuzzy c -Means Clustering for Uncertain Data using Quadratic Penalty-Vector Regularization with Explicit Mappings

ENDO Yasunori
Dept. of Risk Eng.
University of Tsukuba
Ibaraki 305-8573, Japan

TAKAYAMA Isao
Dept. of Risk Eng.
University of Tsukuba
Ibaraki 305-8573, Japan

HAMASUNA Yukihiro
Dept. of Informatics
Kinki University
Osaka 577-8502, Japan

MIYAMOTO Sadaaki
Dept. of Risk Eng.
University of Tsukuba
Ibaraki 305-8573, Japan

Abstract—Recently, fuzzy c -means clustering with kernel functions is remarkable in the reason that these algorithms can handle datasets which consist of some clusters with nonlinear boundaries. However the algorithms have the following problems: (1) the cluster centers can not be calculated explicitly, (2) it takes long time to calculate clustering results. By the way, we have proposed the clustering algorithms using penalty-vector regularization to handle uncertain data. In this paper, we propose new clustering algorithms using quadratic penalty-vector regularization by introducing explicit mappings of kernel functions to solve the following problems. Moreover, we construct fuzzy classification functions for our proposed clustering methods.

Keywords— clustering, kernel function, uncertain data, penalty-vector regularization, classification function, explicit mapping

I. INTRODUCTION

Clustering is one of the data classification techniques and we can classify data without any external criterion using clustering methods. One of most typical and useful clustering methods is fuzzy c -means clustering (FCM) [1] and we consider FCM in this paper.

Now, studies of clustering methods for detecting nonlinear boundary have been widely discussed and kernel trick [2], [3] is focused as a very useful tool to handle datasets which consist of some clusters with nonlinear boundaries. The kernel trick uses some function called “kernel function”.

Here, we consider a mapping $\phi : \mathbb{R}^p \rightarrow \mathbb{R}^s$ ($p \ll s$). The kernel function is an operator $K : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ which satisfies $K(x, y) = \langle \phi(x), \phi(y) \rangle$. We can calculate the value of the inner product in the high-dimensional feature space with low-cost calculation by the kernel functions.

However, special problems frequently arise when we introduce kernel functions into FCM. First, the cluster centers can not be calculated directly. The reason is that the mapping ϕ can not be represented explicitly in general. The second problem is caused by the first one, that is, it takes long time to calculate clustering results. The procedure to calculate the belongingness or membership grades of each data to clusters becomes complicated. To solve the above problems,

Miyamoto et al. introduced explicit mappings into kernel trick and succeeded to calculate the cluster centers directly [4].

By the way, the data we handle have uncertainties in many cases. Until now, we have two methodologies to analyze such data, deterministic and parametric ones.

In the deterministic methodology, the uncertainty introduced data is represented as an interval. Distances or dissimilarities are the most essential measures for data analysis and therefore, the measures between intervals must be defined when we handle uncertainty of data as interval. There are many kinds of such measures, e.g., Hausdorff distance. However, we have two problems when we try to classify the dataset of such data. First, we don't have any clear principle to select which measures to use. Second, we can't discuss clustering in the framework of optimization with such measures.

On the other hand, the uncertainty is represented by some probabilistic density function (PDF) in parametric methodology. In this methodology, the clustering is regarded as a method to determine some parameters of the PDF. However, the methodology have also at least one problem, that is, selection of types of PDFs.

Therefore, we have proposed and developed the concept of tolerance in Ref. [5], [6] to solve the above problems. In our concept, tolerance vectors [5] and penalty ones [6] play main role. Uncertain data is allowed to allocate any position by those vectors as far as the constraints for those vectors are satisfied and the position is derived as an optimal solution of a given objective function. Hence, we can say that this concept is in the framework of methodology of soft computing. Penalty vector is a similar concept to tolerance one and the methods using penalty vectors become more flexible than tolerance vectors because no constraint for the vectors is needed. Moreover, the concept has been developed by using kernel trick in Ref. [7]. The method can classify datasets which consist of clusters of uncertain data with nonlinear boundary.

In this paper, we try to construct new FCMs for uncertain data using quadratic penalty-vector regularization with explicit mappings of kernel functions. The proposed methods have the following advantages:

- 1) The method can classify datasets which consist of clusters of uncertain data with nonlinear boundaries because of kernel trick.
- 2) The methods can calculate clustering results at shorter

ENDO Yasunori: endo@risk.tsukuba.ac.jp
TAKAYAMA Isao: takayama@soft.risk.tsukuba.ac.jp
HAMASUNA Yukihiro: yhamas@info.kindai.ac.jp
MIYAMOTO Sadaaki: miyamoto@risk.tsukuba.ac.jp

time than the conventional methods because of explicit mappings.

Moreover, we try to construct fuzzy classification functions for our proposed clustering. Fuzzy classification functions [10] calculate the values of membership grade of arbitrary points for any cluster by handling clustering results as prototypes of supervised classification. Therefore, the functions are very useful tools to give some principle to classify new observations and we can construct the functions because kernel functions can be described as explicit mapping.

II. FCM FOR UNCERTAIN DATA WITH REGULARIZATION TERM OF PENALTY VECTORS

A. Preliminaries

First of all, we define some symbols. Each data is denoted $x_k = (x_{k1}, \dots, x_{kp})^T \in \mathbb{R}^p$ and the dataset $X = \{x_1, \dots, x_n\}$ is given. Each cluster $C_i (i = 1, \dots, c)$ has a cluster center $v_i = (v_{i1}, \dots, v_{ip})^T \in \mathbb{R}^p$. V means a set of cluster centers $\{v_1, \dots, v_c\}$. A membership grade for x_k to C_i , which means belongingness of x_k to C_i , is denoted by u_{ki} . U means a partition matrix $(u_{ki})_{1 \leq k \leq n, 1 \leq i \leq c}$. Moreover, $\delta_k = (\delta_{k1}, \dots, \delta_{kp})^T \in \mathbb{R}^p$ and $w_k \in [0, +\infty)$ mean a penalty vector and a weight of the penalty vector δ_k , respectively. Δ and W mean a set of penalty vectors $\{\delta_1, \dots, \delta_n\}$ and a set of weights $\{w_1, \dots, w_n\}$, respectively.

B. Objective Functions

As mentioned above, a method using quadratic penalty-vector regularization has been proposed in Ref. [6] as methodology of soft computing. Here, we show the objective function of FCM using quadratic penalty-vector regularization (FCMP).

$$J_{sp}(U, \Delta, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ki})^m d_{ki} + \sum_{k=1}^n w_k \|\delta_{kj}\|^2, \quad (1)$$

$$J_{ep}(U, \Delta, V) = \sum_{k=1}^n \sum_{i=1}^c u_{ki} d_{ki} + \lambda^{-1} \sum_{k=1}^n \sum_{i=1}^c u_{ki} \log u_{ki} + \sum_{k=1}^n w_k \|\delta_{kj}\|^2. \quad (2)$$

The dissimilarity d_{ki} is the square of the norm as follows.

$$\begin{aligned} d_{ki} &= d(x_k + \delta_k, v_i) \\ &= \|x_k + \delta_k - v_i\|^2 \\ &= \sum_{j=1}^p (x_{kj} + \delta_{kj} - v_{ij})^2. \end{aligned} \quad (3)$$

The constraint is as follows.

$$\sum_{i=1}^c u_{ki} = 1, \quad u_{ki} \in [0, 1], \quad \forall k. \quad (4)$$

J_{sp} and J_{ep} are objective functions for the standard FCMP (sFCMP) and the entropy based FCMP (eFCMP), respectively. The solutions can be derived by using Lagrange multiplier.

C. Optimal Solutions and Algorithms

The optimal solutions of sFCMP are as follows:

$$u_{ki} = \frac{(1/d_{ki})^{\frac{1}{m-1}}}{\sum_{l=1}^c (1/d_{kl})^{\frac{1}{m-1}}}, \quad (5)$$

$$v_i = \frac{\sum_{k=1}^n (u_{ki})^m (x_k + \delta_k)}{\sum_{k=1}^n (u_{ki})^m}, \quad (6)$$

$$\delta_k = \frac{-\sum_{i=1}^c (u_{ki})^m (x_k - v_i)}{w_k + \sum_{i=1}^c (u_{ki})^m}. \quad (7)$$

The optimal solutions of sFCMP are as follows:

$$u_{ki} = \frac{\exp(-\lambda d_{ki})}{\sum_{l=1}^c \exp(-\lambda d_{kl})}, \quad (8)$$

$$v_i = \frac{\sum_{k=1}^n u_{ki} (x_k + \delta_k)}{\sum_{k=1}^n u_{ki}}, \quad (9)$$

$$\delta_k = \frac{-\sum_{i=1}^c u_{ki} (x_k - v_i)}{w_k + 1}. \quad (10)$$

We can construct clustering algorithms using the above optimal solutions.

Algorithm 1 (sFCMP and eFCMP):

Step 1 Set the initial values U and Δ .

Step 2 Calculate V by the optimal solution on fixing U and Δ .

Step 3 Calculate U by the optimal solution on fixing Δ and V .

Step 4 Calculate Δ by the optimal solution on fixing V and U .

Step 5 Finish the algorithm if the solutions are convergent, else go back to Step 2.

III. KERNELIZED FCM FOR UNCERTAIN DATA WITH QUADRATIC REGULARIZATION TERM OF PENALTY VECTORS USING EXPLICIT MAPPINGS

In this section, we try to construct FCM for uncertain data using quadratic penalty-vector regularization with explicit mappings of kernel functions (EK-FCMP). First, we describe kernelized FCM for uncertain data using quadratic penalty-vector regularization (K-FCMP). Next, we introduce explicit mappings of kernel functions into K-FCMP to construct EK-FCMP. Last, we show an algorithm of EK-FCMP.

A. Kernel Functions

First, we define some symbols to introduce kernel functions. A mapping from the pattern space to the feature space is expressed as $\phi : \mathbb{R}^p \rightarrow \mathbb{R}^s$ ($p \ll s$). Each data in the feature space is denoted $\phi(x_k) = x_k^\phi = (x_{k1}^\phi, \dots, x_{ks}^\phi)^T \in \mathbb{R}^s$ and the dataset $X^\phi = \{x_1^\phi, \dots, x_n^\phi\}$ is given. Each cluster C_i ($i = 1, \dots, c$) has a cluster center $v_i^\phi = (v_{i1}^\phi, \dots, v_{is}^\phi)^T \in \mathbb{R}^s$. V^ϕ means a set of cluster centers $\{v_1^\phi, \dots, v_c^\phi\}$. A membership grade for x_k^ϕ to C_i , which means belongingness of x_k^ϕ to C_i , is denoted by u_{ki} . U means a partition matrix $(u_{ki})_{1 \leq k \leq n, 1 \leq i \leq c}$. Moreover, $\delta_k^\phi = (\delta_{k1}^\phi, \dots, \delta_{ks}^\phi)^T \in \mathbb{R}^s$ and $w_k \in [0, +\infty)$ mean a penalty vector and a weight of the penalty vector δ_k^ϕ , respectively. Δ and W mean a set of penalty vectors $\{\delta_1^\phi, \dots, \delta_n^\phi\}$ and a set of weights $\{w_1, \dots, w_n\}$, respectively.

Now, a kernel function $K : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ satisfies the following relation:

$$K(x, y) = \langle \phi(x), \phi(y) \rangle \quad (11)$$

From Mercer's theorem [9], K is a continuous symmetric non-negative definite kernel iff a mapping ϕ which satisfies the above relation exists. We should notice that ϕ is not explicit.

B. Objective Functions

Kernelized fuzzy c -means clustering (K-FCM) is FCM in which a kernel function is introduced into dissimilarity. An extended K-FCM by introducing penalty vectors have been proposed in Ref. [7]. We show the objective function of the standard model J_{ksp} and the one of the entropy based model J_{ksp} of K-FCM using quadratic penalty-vector regularization (K-FCMP) as follows:

$$J_{ksp}(U, \Delta^\phi, V^\phi) = \sum_{k=1}^n \sum_{i=1}^c (u_{ki})^m d_{ki} + \sum_{k=1}^n w_k \|\delta_{kj}^\phi\|^2, \quad (12)$$

$$J_{kep}(U, \Delta^\phi, V^\phi) = \sum_{k=1}^n \sum_{i=1}^c u_{ki} d_{ki} + \lambda^{-1} \sum_{k=1}^n \sum_{i=1}^c u_{ki} \log u_{ki} + \sum_{k=1}^n w_k \|\delta_{kj}^\phi\|^2. \quad (13)$$

The dissimilarity d_{ki} between the k -th uncertain data and the i -th cluster is the square of the norm on the feature space as follows:

$$\begin{aligned} d_{ki} &= d(x_k^\phi + \delta_k^\phi, v_i^\phi) \\ &= \|x_k^\phi + \delta_k^\phi - v_i^\phi\|^2 \\ &= \sum_{j=1}^s (x_{kj}^\phi + \delta_{kj}^\phi - v_{ij}^\phi)^2. \end{aligned} \quad (14)$$

The constraint is as follows:

$$\sum_{i=1}^c u_{ki} = 1, \quad u_{ki} \in [0, 1], \quad \forall k. \quad (15)$$

C. Optimal Solutions

We can derive the optimal solutions by using Lagrange multiplier. The optimal solutions of standard K-FCMP are as follows:

$$u_{ki} = \frac{(1/d_{ki})^{\frac{1}{m-1}}}{\sum_{l=1}^c (1/d_{kl})^{\frac{1}{m-1}}}, \quad (16)$$

$$v_i^\phi = \frac{\sum_{k=1}^n (u_{ki})^m (x_k^\phi + \delta_k^\phi)}{\sum_{k=1}^n (u_{ki})^m}, \quad (17)$$

$$\delta_k^\phi = \frac{-\sum_{i=1}^c (u_{ki})^m (x_k^\phi - v_i^\phi)}{w_k + \sum_{i=1}^c (u_{ki})^m}. \quad (18)$$

The optimal solutions of entropy based K-FCMP are as follows:

$$u_{ki} = \frac{\exp(-\lambda d_{ki})}{\sum_{l=1}^c \exp(-\lambda d_{kl})}, \quad (19)$$

$$v_i^\phi = \frac{\sum_{k=1}^n u_{ki} (x_k^\phi + \delta_k^\phi)}{\sum_{k=1}^n u_{ki}}, \quad (20)$$

$$\delta_k^\phi = \frac{-\sum_{i=1}^c u_{ki} (x_k^\phi - v_i^\phi)}{w_k + 1}. \quad (21)$$

We can formally obtain the optimal solutions as above. However, we cannot calculate v_i^ϕ and δ_k^ϕ directly by this method. The reason is that the map ϕ is not explicit and we have to use x_k , not x_k^ϕ . Therefore, Algorithm 1 cannot be adapted and the procedure of K-FCMP becomes more complicated. Hence, We introduce explicit mappings into K-FCMP to calculate v_i^ϕ directly. It is expected that introducing the explicit mappings makes the procedure more simple and makes calculation time short.

D. Optimal Solutions with Explicit Mappings of Kernel Functions

In this paragraph, we introduce explicit mappings of kernel functions [4]. As above, the kernel function K satisfies $K(x, y) = \langle \phi(x), \phi(y) \rangle = \langle x^\phi, y^\phi \rangle$. Now, we consider a kernel function with the kernel matrix $K = (K_{kl}) = (K(x_k, x_l))$. As K is a positive symmetric matrix, we can introduce an inner product $\langle x, y \rangle_K = xKy$ into $\mathbb{R}^s$. We assume that $\phi : X \rightarrow \mathbb{R}^n$ and $\phi(x_k) = x_k^\phi = e_k$. Here e_k is a unit vector. Because ϕ can be represented explicitly, we call ϕ using the kernel matrix K explicit mapping. Therefore, the inner product is as follows:

$$\langle e_g, e_h \rangle_K = K_{gh}. \quad (22)$$

Thus, the dissimilarity d_{ki} can be calculated as follows:

$$\begin{aligned} d_{ki} &= d(x_k^\phi + \delta_k^\phi, v_i^\phi) \\ &= d(e_k + \delta_k^\phi, v_i^\phi) \\ &= \langle e_k + \delta_k^\phi - v_i^\phi, e_k + \delta_k^\phi - v_i^\phi \rangle_K \\ &= (e_k + \delta_k^\phi - v_i^\phi)^T K (e_k + \delta_k^\phi - v_i^\phi). \end{aligned} \quad (23)$$

We can calculate the optimal solutions by using the above inner product and dissimilarity. First, we show the solutions of standard EK-FCMP (sEK-FCMP) with the explicit mappings as follows:

$$u_{ki} = \frac{(1/d_{ki})^{\frac{1}{m-1}}}{\sum_{l=1}^c (1/d_{kl})^{\frac{1}{m-1}}}, \quad (24)$$

$$v_i^\phi = \frac{\sum_{k=1}^n (u_{ki})^m (e_k + \delta_k^\phi)}{\sum_{k=1}^n (u_{ki})^m}, \quad (25)$$

$$\delta_k^\phi = \frac{-\sum_{i=1}^c (u_{ki})^m (e_k - v_i^\phi)}{w_k + \sum_{i=1}^c (u_{ki})^m}. \quad (26)$$

Next, we show the solutions of entropy based EK-FCMP (eEK-FCMP) with the explicit mappings as follows:

$$u_{ki} = \frac{\exp(-\lambda d_{ki})}{\sum_{l=1}^c \exp(-\lambda d_{kl})}, \quad (27)$$

$$v_i^\phi = \frac{\sum_{k=1}^n u_{ki} (e_k + \delta_k^\phi)}{\sum_{k=1}^n u_{ki}}, \quad (28)$$

$$\delta_k^\phi = \frac{-\sum_{i=1}^c u_{ki} (e_k - v_i^\phi)}{w_k + 1}. \quad (29)$$

E. Algorithm

Here, we construct an algorithm for EK-FCMP using the above optimal solutions with the explicit mappings as follows.

Algorithm 2 (sEK-FCMP and eEK-FCMP):

Step 1 Set the initial values U and Δ^ϕ . Generate the kernel matrix K .

Step 2 Calculate V^ϕ by the optimal solutions with the explicit mappings on fixing U and Δ^ϕ .

Step 3 Calculate U by the optimal solutions with the explicit mappings on fixing Δ^ϕ and V^ϕ .

Step 4 Calculate Δ^ϕ by the optimal solutions with the explicit mapping on fixing V^ϕ and U .

Step 5 Finish if the solutions are convergent, else go back to Step 2.

F. Fuzzy Classification Functions

Fuzzy classification functions [10] calculate the value of membership grade of an arbitrary point for any cluster by handling the clustering result as a prototype of supervised classification. Therefore, The functions are very useful tools to give some indication to classify new observations and it is significant to discuss how to construct the functions in case of the work in this paper. However, it is difficult to construct the functions because the optimal solution of the membership grades includes the term of penalty vectors and the value cannot be calculated directly. Thus, we have to consider another way to calculate the values of membership grades and we try to obtain the values by iterative optimization for an arbitrary point on the pattern space. We show a standard model algorithm to calculate the value of membership grade $u_i(x) = u_i$ of an arbitrary point $x \in \mathbb{R}^p$ for the i -th cluster.

Algorithm 3:

Step 1 Suppose that a new observation x is given. Derive a cluster center set V^ϕ by Algorithm 2. Introduce $\hat{V}^\phi = \{\hat{v}_i^\phi \mid \hat{v}_i^\phi = (v_{i1}^\phi, \dots, v_{is}^\phi, 0)^T\}$ and generate a kernel matrix $\hat{K} \in \mathbb{R}^{(n+1) \times (n+1)}$ for the new dataset $\hat{X} = \{X, x\}$. Set an initial penalty vector $\hat{\delta}^\phi \in \mathbb{R}^{n+1}$ for the observation x .

Step 2 Calculate u_i by the following equation on fixing $\hat{\delta}^\phi$

$$u_i = \frac{(1/d_i)^{\frac{1}{m-1}}}{\sum_{l=1}^c (1/d_l)^{\frac{1}{m-1}}}, \quad (30)$$

$$d_i = (e_{n+1} + \hat{\delta}^\phi - \hat{v}_i^\phi)^T \hat{K} (e_{n+1} + \hat{\delta}^\phi - \hat{v}_i^\phi). \quad (31)$$

Step 3 Calculate $\hat{\delta}^\phi$ by the following equation on fixing u_i .

$$\hat{\delta}^\phi = \frac{-\sum_{i=1}^c (u_i)^m (e_{n+1} - \hat{v}_i^\phi)}{w(x) + \sum_{i=1}^c (u_i)^m}. \quad (32)$$

Step 4 Finish if the solutions are convergent, else go back to Step 2.

We omit the algorithm for eEK-FCMP.

IV. NUMERICAL EXAMPLES

We show numerical examples in Fig. 1 ~ Fig. 24. The pattern space is $[0, 1] \times [0, 1]$. We use the following Gaussian kernel:

$$K(x, y) = \exp(-\beta \|x - y\|^2).$$

The parameters are as follows: $m = 1.2$, $\beta = 70$, $\lambda = 1.2$ and $w_k = 3.0$ for all k . We use the same initial values at all case.

We apply the proposed methods to three types of datasets. First, we compare the conventional and the proposed methods through the nesting dataset. Fig. 1 ~ Fig. 4, Fig. 5 ~ Fig. 8, Fig. 9 ~ Fig. 12 and Fig. 13 ~ Fig. 16 are the results by K-sFCM, K-eFCM, sEK-FCMP and eEK-FCMP, respectively. We also show the cluster centers by using kernel principal component analysis (K-PCA) [11].

Next, we apply sEK-FCMP to other two types of datasets. The clustering results are shown in Fig. 17 (double-circle dataset) and Fig. 21 (double-mass dataset). Fig. 18 and Fig. 22 show the cluster centers in the feature space. The dimension of the feature space by Gaussian Kernel is infinite so that we use the kernel principal component analysis (K-PCA) [11] to show the two dimensional Euclidean space. It became possible to show the cluster centers explicitly by our proposed methods. Fig. 19 and Fig. 20, and Fig. 23 and Fig. 24 show the fuzzy classification functions of each cluster for three types of datasets. We can classify the new given data by the classification functions.

In Table I we show the range of w_k when desirable clustering results are obtained. In the range, the larger w_k gets, the faster the algorithm converges and the clearer the shape of the fuzzy classification functions becomes. EK-FCMP, in particular the entropy based model, with the adequate w_k gives more natural results than K-FCM.

TABLE I

THE RANGE OF w_k WHEN OBTAINING DESIRABLE CLUSTERING RESULTS

sEK-FCMP ($\beta = 70, m = 1.2$)	$1.62 \leq w_k \leq 4.96$
sEK-FCMP ($\beta = 40, m = 1.2$)	None
eEK-FCMP ($\beta = 70, \lambda = 10$)	$2.52 \leq w_k \leq 5.68$
eEK-FCMP ($\beta = 40, \lambda = 5.0$)	$13.2 \leq w_k$

V. CONCLUSION

In this paper, we developed FCM for uncertain data using quadratic penalty-vector regularization by introducing explicit mappings of kernel functions. The proposed algorithm has the following features:

- 1) The algorithm can handle datasets which consist of some clusters with nonlinear boundary.
- 2) Because of explicit mappings of kernel functions, we reduce the calculation time.
- 3) The algorithm becomes more simple than the conventional ones.

Moreover, we construct the fuzzy classification functions.

In the forthcoming paper, we will discuss how to find appropriate parameters and apply to real data.

ACKNOWLEDGMENT

This study is partly supported by the Grant-in-Aid for Scientific Research (C) (Project No.21500212) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

- [1] J. C. Bezdek: *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York, 1981.
- [2] V. N. Vapnik: *Statistical Learning Theory*, Wiley, New York, 1998.
- [3] V. N. Vapnik: *The nature of Statistical Learning Theory*, 2nd edn., Springer, New York, 2000.
- [4] S. Miyamoto, Y. Kawasaki, K. Sawazaki: *An Explicit Mapping for Fuzzy c-Means using Kernel Function and Application to Text Analysis*, IFSA/EUSFLAT 2009, 2009.
- [5] Y. Endo, R. Murata, H. Haruyama, S. Miyamoto: *Fuzzy c-Means for Data with Tolerance*, Proc. 2005 International Symposium on Nonlinear Theory and Its Applications, pp.345-348, 2005.
- [6] Y. Hasegawa, Y. Endo, Y. Hamasuna: *On Fuzzy c-Means for Data with uncertainty using Spring Modulus*, SCIS&ISIS 2008, 2008.
- [7] Y. Kanzawa, Y. Endo, S. Miyamoto: *Entropy Regularized Fuzzy c-Means for Data with Tolerance introducing Penalty Term in Feature Space*, SCIS&ISIS 2008, 2008.

- [8] S. Miyamoto and M. Mukaidono: *Fuzzy c-Means as a Regularization and Maximum Entropy Approach*, Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97), Vol.2, pp.86-92, 1997.
- [9] J. Mercer: *Functions of Positive and Negative Type and Their Connection with the Theory of Integral Equations*, Philosophical Transactions of the Royal Society A, 209, pp.415-446, 1909.
- [10] S. Miyamoto, K. Umayahara, M. Mukaidono: *Fuzzy Classification Functions in the Methods of Fuzzy c-Means and Regularization by Entropy*, Journal of Japan Society for Fuzzy Theory and Systems Vol.10, No.3, pp.548-557, 1998.
- [11] B. Schölkopf, A. Smola, and K.-R. Müller: *Nonlinear Component Analysis as a Kernel Eigenvalue Problem*, Neural Computation, Vol.10, pp.1299-1319, 1998.

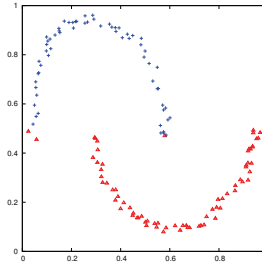


Fig. 1. K-sFCM

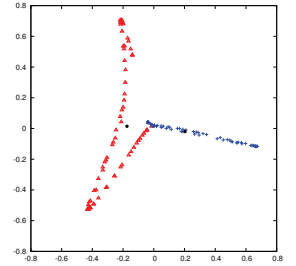


Fig. 2. K-PCA for K-sFCM

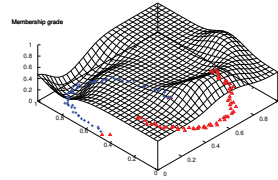


Fig. 3. The fuzzy classification function for K-sFCM (the first cluster)

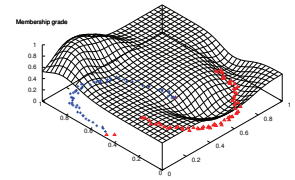


Fig. 4. The fuzzy classification function for K-sFCM (the second cluster)

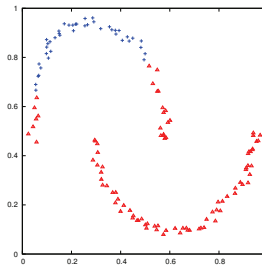


Fig. 5. K-eFCM

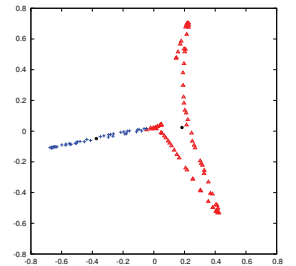


Fig. 6. K-PCA for K-eFCM

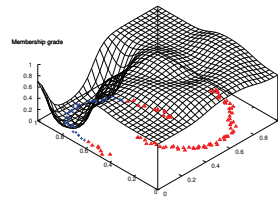


Fig. 7. The fuzzy classification function for K-eFCM (the first cluster)

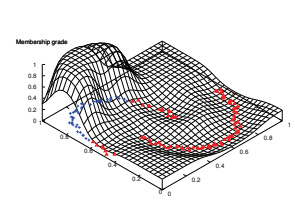


Fig. 8. The fuzzy classification function for K-eFCM (the second cluster)

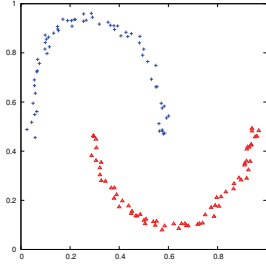


Fig. 9. sEK-FCMP

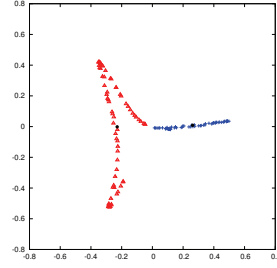


Fig. 10. K-PCA for sEK-FCMP

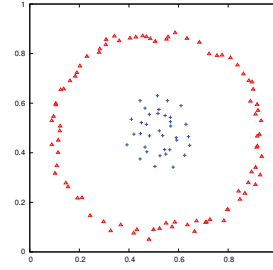


Fig. 17. sEK-FCMP

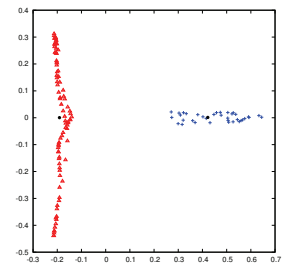


Fig. 18. K-PCA for sEK-FCMP

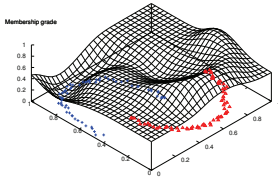


Fig. 11. The fuzzy classification function for sEK-FCMP (the first cluster)

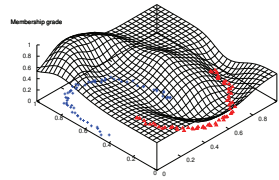


Fig. 12. The fuzzy classification function for sEK-FCMP (the second cluster)

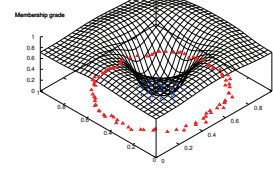


Fig. 19. The fuzzy classification function for sEK-FCMP (the first cluster)

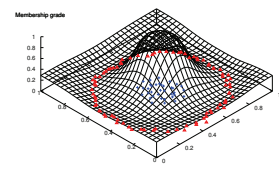


Fig. 20. The fuzzy classification function for sEK-FCMP (the second cluster)

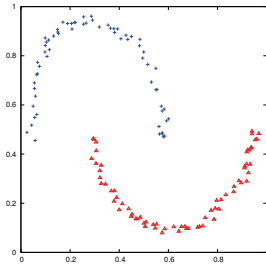


Fig. 13. eEK-FCMP

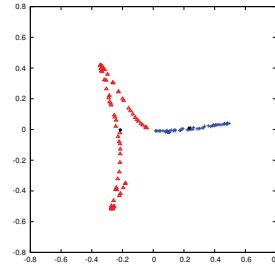


Fig. 14. K-PCA for eEK-FCMP

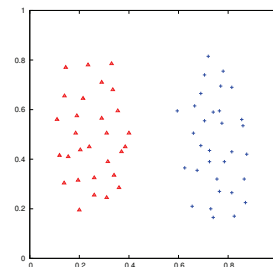


Fig. 21. sEK-FCMP

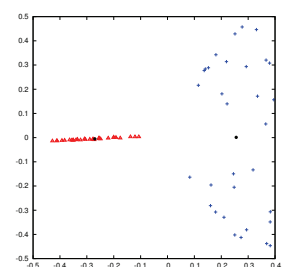


Fig. 22. K-PCA for sEK-FCMP

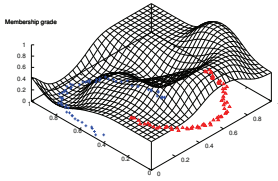


Fig. 15. The fuzzy classification function for eEK-FCMP (the first cluster)

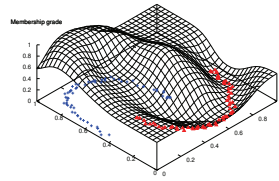


Fig. 16. The fuzzy classification function for eEK-FCMP (the second cluster)

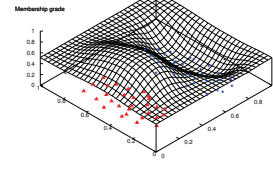


Fig. 23. The fuzzy classification function for sEK-FCMP (the first cluster)

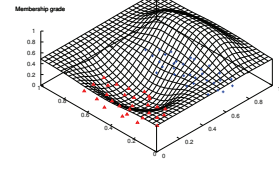


Fig. 24. The fuzzy classification function for sEK-FCMP (the second cluster)