The Third AFSS(1998), 148−152

Fuzzy c-Means Clustering Algorithm with Pseudo Mahalanobis Distances

Hidetomo ICHIHASHI[a] , Masayuki OHUE[a] and Tetsuya MIYOSHI[a]

a. Department of Industrial Engineering, College of Engineering, Osaka Prefecture University
1-1 Gakuencho, Sakai, Osaka 599-8531 Japan
Tel:+81-722-54-9352 Fax:+81-722-54-9915 E-mail:ichi@ie.osakafu-u.ac.jp

Abstract

Gustafson and Kessel proposed a modified fuzzy c-Means algorithm based on the Mahalanobis distance. Though the algorithm appears more natural through the use of a fuzzy covariance matrix, it needs to calculate determinants and inverses of the c-fuzzy scatter matrices. This paper proposes a fuzzy clustering algorithm using pseudo Mahalanobis distances, which is more easy to use and flexible than the Gustafson and Kessel's fuzzy c-Means.

Keywords Fuzzy clustering, Fuzzy c-Means, linear clusters

1. Introduction

The first fuzzy clustering algorithm was developed by Ruspini [1]. Fuzzy ISODATA [2] and its extension Fuzzy c-Means [3] are the popular fuzzy clustering algorithm by the distance-based objective function methods (the within-group sum-of-squared-error (WGSS) criterion).

Other approaches are driven by optimization of a generalized fuzzy c-prototypes functional defined by a measure of similarity (or dissimilarity) between pattern (datum) and prototype. In Bezdek et al.[4, 5] the fitting prototypes are either straight lines and the measure is the orthogonal distance, or more generally, prototypes that are convex combinations of points and lines. The switching regression models [6] partitions the patterns and simultaneously provides estimates of the parameters of linear functions, which define the best-fit regression models.

The Mahalanobis distance between $\boldsymbol{x}_1, \boldsymbol{x}_2 \in R^I$ defined by the weighted inner product $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}_j$ is an important tool for pattern classification. Gustafson and Kessel [7] proposed a modified fuzzy c-Means algorithm based on the Mahalanobis distance, which appears more natural through the use of a fuzzy covariance matrix. The local variation of the norm may identify clusters of various geometric shapes including linear clusters.

Though the algorithm is much more flexible than the conventional fuzzy c-Means with the norm $\boldsymbol{x}^T \boldsymbol{x}$, it needs to calculate determinants and inverses of the c-fuzzy scatter matrices whose regularity is not guaranteed, and thus it is computationally intractable. This paper proposes a modified fuzzy c-Means clustering algorithm using pseudo Mahalanobis distances with maximizing entropy.

2. Convenient clustering algorithm with pseudo Mahalanobis distances

Let us consider a problem to partition J patterns into C clusters. Let I dimensional feature vector of jth pattern be

$$\boldsymbol{x}_j = (x_{j1}, x_{j2}, \cdots, x_{jI}) \qquad j = 1, \cdots, J \tag{1}$$

u_{cj} is the membership of pattern j to cluster c.

$$\boldsymbol{v}_c = (v_{c1}, v_{c2}, \cdots, v_{cI}) \qquad c = 1, \cdots, C \tag{2}$$

is a vector of cluster center. $\boldsymbol{A}_c = (a_{cik})$, $\boldsymbol{B}_c = (b_{cik})$ and $\boldsymbol{D}_c = (d_{cik})$ are $I \times I$ symmetric matrices which define the pseudo Mahalanobis distance.

$$\boldsymbol{A}_c = \begin{bmatrix} a_{c11} & a_{c12} & \dots & a_{c1I} \\ a_{c21} & a_{c22} & \dots & a_{c2I} \\ \vdots & \vdots & \ddots & \vdots \\ a_{cI1} & a_{cI2} & \dots & a_{cII} \end{bmatrix} \tag{3}$$

$$\boldsymbol{B}_c = \begin{bmatrix} b_{c11} & 0 & \dots & 0 \\ 0 & b_{c22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_{cII} \end{bmatrix} \tag{4}$$

$$\boldsymbol{D}_c = \begin{bmatrix} d_{c11} & d_{c12} & \dots & d_{c1I} \\ d_{c21} & d_{c22} & \dots & d_{c2I} \\ \vdots & \vdots & \ddots & \vdots \\ d_{cI1} & d_{cI2} & \dots & d_{cII} \end{bmatrix} \tag{5}$$

where $a_{cik} > 0, b_{cii} > 0$ and $d_{cik} < 0$ for all c,i,k.

The pseudo Mahalanobis distance is defined by a matrix $\boldsymbol{E}_c = (e_{cik})$

$$e_{cik} = (1 - \delta_{ik})a_{cik} + \delta_{ik}b_{cik} + (1 - \delta_{ik})d_{cik} \tag{6}$$

where δ_{ik} is a Kronecker's symbol

$$\delta_{ik} = \begin{cases} 1 & (i = k) \\ 0 & (i \neq k) \end{cases}$$

Let the objective function (Lagrange function) be

$$\begin{aligned} \min\ L = & \sum_{j=1}^{J}\sum_{c=1}^{C}\sum_{i=1}^{I}\sum_{k=1}^{I} u_{cj}\{(1 - \delta_{ik})a_{cik} \\ & +\delta_{ik}b_{cik} + (1 - \delta_{ik})d_{cik}\}(x_{ji} - v_{ci})(x_{jk} - v_{ck}) \\ & -\sum_{c=1}^{C}\lambda_c^a\left(\sum_{i=1}^{I}\sum_{k=1}^{I} a_{cik} - 1\right) \\ & -\sum_{c=1}^{C}\lambda_c^b\left(\sum_{i=1}^{I} b_{cii} - 1\right) \\ & -\sum_{c=1}^{C}\lambda_c^d\left(\sum_{i=1}^{I}\sum_{k=1}^{I} d_{cik} + 1\right) \\ & +\lambda_0\sum_{c=1}^{C}\left\{\sum_{i=1}^{I}\sum_{k=1}^{I}(a_{cik}\log a_{cik}\right. \\ & \left. +(-d_{cik})\log(-d_{cik})) + \sum_{i=1}^{I} b_{cii}\log b_{cii}\right\} \\ & -\sum_{j=1}^{J}\tau_j\left(\sum_{c=1}^{C} u_{cj} - 1\right) \\ & +\tau_0\sum_{j=1}^{J}\sum_{c=1}^{C} u_{cj}\log u_{cj} \end{aligned} \tag{7}$$

The first term of L represents the sum of pseudo Mahalanobis distances defined by the matrix $\boldsymbol{E}_c$. a_{cik}'s , b_{cik}'s and d_{cik}'s are decision variables. The second and third terms represent the constraints that the components of matrix $\boldsymbol{A}_c$ and $\boldsymbol{B}_c$ sum to one respectively, and λ_c^a's and λ_c^b's are the Lagrange multipliers. The fourth term represents the constraint that the components of $\boldsymbol{D}_c$ sum to minus one. λ_c^d's are the Lagrange multipliers. The fifth term is to make $a_{cik} > 0$, $b_{cik} > 0$ and $d_{cik} < 0$ and by the second to fourth terms, $a_{cik}, b_{cii}, -d_{cik} \in (0,1)$. The sixth term represents a constraint that membership u_{cj} sum to one and τ_j's are the Lagrange multipliers. The seventh term is an entropy term introduced by Miyamoto and Mukaidono [8] to obtain a fuzzy partition. The larger $\tau_0(> 0)$ is the more fuzzy clusters are obtained.

From the necessary condition of the optimality of L

$$\frac{\partial L}{\partial a_{cik}} = 0 \tag{8}$$

Since $a_{cik} = a_{cki}$ we have

$$\sum_{j=1}^{J}(1 - \delta_{ik})u_{cj}(x_{ji} - v_{ci})(x_{jk} - v_{ck}) \\ -\lambda_c^a + \lambda_0(\log a_{cik} + 1) = 0 \tag{9}$$

$$\frac{\partial L}{\partial \lambda_c^a} = \sum_{i=1}^{I}\sum_{k=1}^{I} a_{cik} - 1 = 0 \tag{10}$$

By substituting Eq.(9) into Eq.(10), λ_c^a is obtained. Then again substitute λ_c^a into Eq.(9) which yields

$$a_{cik} = \frac{\exp(F_{cik})}{\sum_{m=1}^{I}\sum_{n=1}^{I}\exp(F_{cmn})} \tag{11}$$

where

$$F_{cmn} = -\frac{1}{\lambda_0}\sum_{j=1}^{J}(1 - \delta_{mn})u_{cj} \\ (x_{jm} - v_{cm})(x_{jn} - v_{cn}) \tag{12}$$

By the similar manner we have

$$b_{cii} = \frac{\exp(G_{cii})}{\sum_{m=1}^{I}\exp(G_{cmm})} \tag{13}$$

where

$$G_{cmm} = -\frac{1}{\lambda_0}\sum_{j=1}^{J}\delta_{mn}u_{cj}(x_{jm} - v_{cm})^2 \tag{14}$$

and

$$d_{cik} = \frac{-\exp(H_{cik})}{\sum_{m=1}^{I}\sum_{n=1}^{I}\exp(H_{cmn})} \tag{15}$$

where

$$H_{cmn} = \frac{1}{\lambda_0}\sum_{j=1}^{J}(1 - \delta_{mn})u_{cj}$$

$$(x_{jm} - v_{cm})(x_{jn} - v_{cn}) \tag{16}$$

From

$$\frac{\partial L}{\partial v_{ci}} = 0 \tag{17}$$

we have

$$\sum_{k=1}^{I}\{(1-\delta_{ik})a_{cik} + \delta_{ik}b_{cik} + (1-\delta_{mn})d_{cik}\}$$

$$\left(\sum_{j=1}^{J} u_{cj}x_{jk} - v_{ck}\sum_{j=1}^{J} u_{cj}\right) = 0 \tag{18}$$

Thus,

$$v_{ck} = \frac{\sum_{j=1}^{J} u_{cj}x_{jk}}{\sum_{j=1}^{J} u_{cj}} \quad , \quad k = 1,\cdots,I \tag{19}$$

And, from

$$\frac{\partial L}{\partial u_{cj}} = 0 \tag{20}$$

we have

$$u_{cj} = \frac{\exp(R_{cj})}{\sum_{q=1}^{C}\exp(R_{qj})} \tag{21}$$

where

$$R_{cj} = -\frac{1}{\tau_0}\sum_{i=1}^{I}\sum_{k=1}^{I}\{(1-\delta_{ik})a_{cik} + \delta_{ik}b_{cik}$$

$$+(1-\delta_{ik})d_{cik}\}(x_{ji} - v_{ci})(x_{jk} - v_{ck}) \tag{22}$$

The clustering algorithm is the iteration through necessary conditions Eqs.(11), (13), (15), (19),and (21).

step 1 : Fix C, $2 \le C \le J$, fix λ_0 and τ_0. Initialize $u_{cj}, j = 1,\cdots,J, c = 1,\cdots,C$ such that $\sum_{c=1}^{C} u_{cj} = 1$.

step 2 : Calculate the fuzzy cluster centers $\boldsymbol{v}_c$ with Eq.(19).

step 3 : Calculate $\boldsymbol{A}_c$ with Eq.(11) , $\boldsymbol{B}_c$ with Eq.(13) and $\boldsymbol{D}_c$ with Eq.(15).

step 4 : Update u_{cj} with Eq.(21).

step 5 : If

$$|u_{cj} - u_{cj}^{OLD}| < \varepsilon$$

is satisfied then stop. Otherwise, return to step 2.

3. Numerical example

The artificial data [7] shown in Fig.1 consists of J=20 points(patterns) in R^2. These data form two visually apparent linear clusters in the shape of a cross; the coordinates of each point are listed in column 1 of Table 1. The fuzzy 2-partition attained in 10 iterations with $\varepsilon = 0.00001$ is exhibited as column 2 of Table 1 and in Fig.1 where $\times$ and $\bigcirc$ denote the data of cluster 1 and 2 respectively. Evidently the proposed algorithm successfully label all 20 data points correctly (although the membership of points 6 and 18 at the center of the cross are fortuitous). The actual hard(sample) covariance matrices ($\boldsymbol{M}_1$, $\boldsymbol{M}_2$), fuzzy covariance matrices ($\boldsymbol{N}_1$, $\boldsymbol{N}_2$) by Gustafson and Kessel and our ($\boldsymbol{E}_1^{-1}, \boldsymbol{E}_2^{-1}$) with $\lambda_0 = 0.10$, $\tau_0 = 0.05$ are

$$\boldsymbol{M}_1 = \begin{bmatrix} 0.13 & 1.50 \\ 1.50 & 24.44 \end{bmatrix}, \boldsymbol{M}_2 = \begin{bmatrix} 33.92 & -0.03 \\ -0.03 & 0.07 \end{bmatrix}$$

$$\boldsymbol{N}_1 = \begin{bmatrix} 0.06 & 0.72 \\ 0.72 & 25.65 \end{bmatrix}, \boldsymbol{N}_2 = \begin{bmatrix} 36.33 & 0.88 \\ 0.88 & 0.05 \end{bmatrix}$$

$$\boldsymbol{E}_1^{-1} = \begin{bmatrix} 3.46 & 32.45 \\ 32.45 & 429.25 \end{bmatrix}, \boldsymbol{E}_2^{-1} = \begin{bmatrix} 1407.04 & 23.94 \\ 23.94 & 1.41 \end{bmatrix}$$

respectively. The all diagonal entries reflect the predominantly linear structure of cluster 1 and 2, cluster 1 having large variance along the x_2-axis, while cluster 2 has large variance along the x_1-axis.

The fuzzy partition of rotated data at an angle of 45° attained in 13 iterations with $\varepsilon = 0.00001$ is exhibited as column 3 of Table 1. The center of rotation is $(0,0)$. The matrices ($\boldsymbol{E}_1^{-1}, \boldsymbol{E}_2^{-1}$) with $\lambda_0 = 0.10$, $\tau_0 = 0.05$ are

$$\boldsymbol{E}_1^{-1} = \begin{bmatrix} 10.72 & 6.72 \\ 6.72 & 5.75 \end{bmatrix}, \boldsymbol{E}_2^{-1} = \begin{bmatrix} 16.37 & -14.26 \\ -14.26 & 14.29 \end{bmatrix}$$

respectively. Fig.2 shows the resultant clusters. The linear structure is detected even in the rotated linear clusters.

We apply artificial data consists of J=60 points in R^3, which forms three linear clusters. Fig.3 shows the result where crisply clustered data are depicted by $\times, \bigcirc$ and $\triangle$ on the $x_1 - x_3$ plane.

4. Concluding remarks

We have proposed the convenient fuzzy clustering algorithm using pseudo Mahalanobis distances and maximizing entropy approach. Since discriminant analysis with Mahalanobis distances has been efficiently applied to various pattern recognition problems, our future works includes simultaneous determination of a data partition and classification by the proposed algorithm.

Table 1 Obtained fuzzy clusters

Data $\boldsymbol{x}_j$	Terminal memberships in fuzzy cluster 1	
	Example 1	Example 2
(-9.75,-0.15)	0.007	0.011
(-6.44,0.34)	0.102	0.115
(-4.69,-0.30)	0.234	0.248
(-2.04,0.37)	0.445	0.441
(-1.24,0.45)	0.484	0.478
(0.33,-0.08)	0.502	0.502
(5.04,-0.21)	0.241	0.279
(5.86,-0.25)	0.171	0.212
(7.54,0.16)	0.068	0.100
(7.67,0.24)	0.063	0.094
(-0.30,-8.07)	0.944	0.926
(0.13,-7.13)	0.897	0.873
(-0.37,-5.18)	0.745	0.725
(0.03,-3.33)	0.596	0.587
(0.35,-2.63)	0.554	0.548
(0.23,-2.68)	0.557	0.551
(-0.05,-2.00)	0.527	0.524
(0.41,0.37)	0.510	0.508
(0.69,4.75)	0.800	0.773
(0.74,8.87)	0.988	0.980

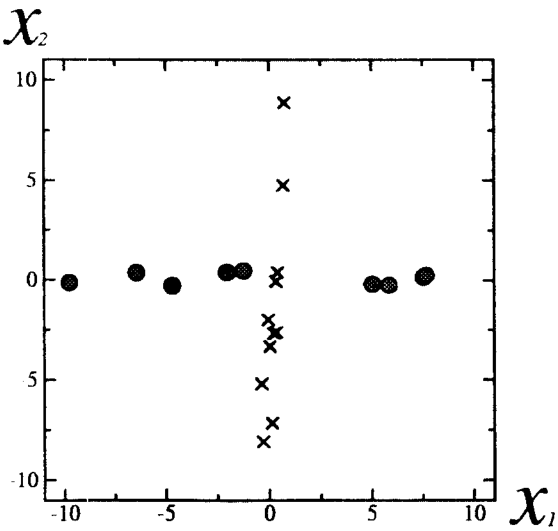

Fig. 1 Gustafson's cross

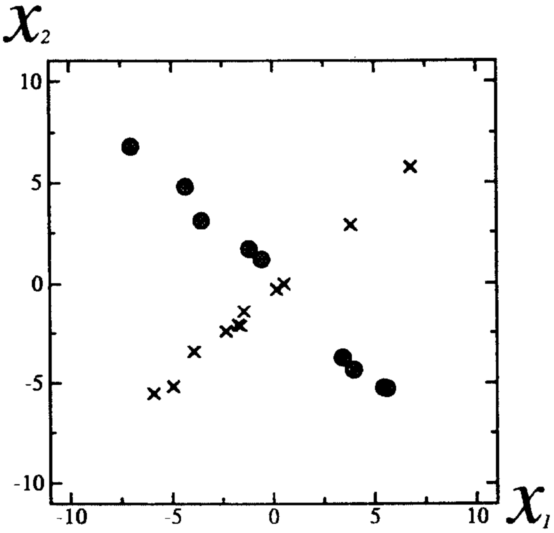

Fig. 2 Rotated patterns

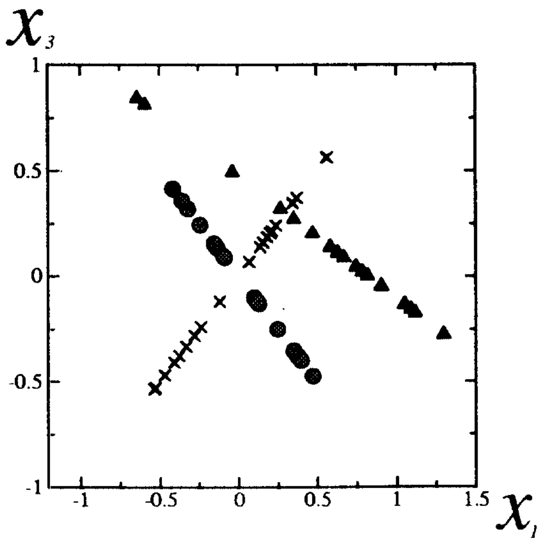

Fig. 3 Three linear clusters in R^3

References

1. E. H. Ruspini , " Numerical methods for fuzzy clustering " , Information Sciences, Vol.2, No.3, pp.319-350, 1970

2. J. C. Dunn , " A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters " , J. Cybernet, Vol.3, pp.32-57, 1974

3. J. C. Bezdek , " Pattern Recognition with Fuzzy Objective Function Algorithms " , Plenum Press, New York, 1981

4. J. C. Bezdek, C. Coray, R. Gunderson and J. Watson , " Detection and characterization of cluster substructure. I. linear structure " , fuzzy c-lines, SIAM J. Appl. Math., Vol.40, No.2, pp.339-357, 1981

5. J. C. Bezdek, C. Coray, R. Gunderson and J. Watson , " Detection and characterization of cluster substructure. II. fuzzy c-varieties and convex combinations thereof " , SIAM J. Appl. Math., Vol.40, No.2, pp.358-372, 1981

6. R. J. Hathaway and J. C. Bezdek , " Switching regression models and fuzzy clustering " , IEEE Trans. on Fuzzy Systems, Vol.1, No.3, pp.195-204, 1993

7. D. E. Gustafson and W. Kessel , " Fuzzy clustering with a fuzzy covariance matrix " , in Proc. IEEE-CDC, Vol.2, pp.761-766, 1979

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), June 25-30, Prague, Chech, Vol.II, pp.86-92, 1997