Selecting Parameter Values for Mahalanobis Distance Fuzzy Classifiers

Peter Deer and Peter Eklund
Knowledge, Visualization and Ordering Laboratory, Griffith University
PMB 50 Gold Coast 9726, Australia
Email: p.deer@gu.edu.au, peklund@dstc.edu.au

Abstract— The fuzzy *c*-means clustering algorithm, and a related supervised classifier, require the *a priori* selection of a weighting parameter called the fuzzy exponent. This paper investigates suitable values of this fuzzy exponent using the criterion that fuzzy set memberships reflect class proportions in the mixed pixels of a remotely sensed image.

INTRODUCTION

The unsupervised fuzzy *c*-means (FCM) [2] clustering algorithm is widely used in pattern recognition applications and attempts to minimize an error function,

$$J_q = \sum_{k=1}^n \sum_{i=1}^c \mu_{ik}^q d_{ik}^2 \quad (1)$$

where there are n observations (e.g. pixels in a remotely sensed image), c classes, μ_{ik} is the membership of observation x_k in class i , q (the “fuzzy exponent”) is a parameter used to control the fuzziness of the class allocation, and $d_{ik}^2 = (x_k - m_i)^t \mathbf{A} (x_k - m_i)$ is the distance measure (or inner product norm) between x_k and the mean of class i (denoted m_i) induced by the positive definite matrix $\mathbf{A}$, with m_i determined by,

$$m_i = \frac{\sum_{k=1}^n \mu_{ik}^q x_k}{\sum_{k=1}^n \mu_{ik}^q} \quad (2)$$

The constraints on the partitioning are given by, $M_{fc} = \{U : \mu_{ik} \in [0, 1]; \sum_{k=1}^n \mu_{ik} > 0, i = 1, \dots, c; \sum_{i=1}^c \mu_{ik} = 1, k = 1, \dots, n, U \text{ is a real } c \times n \text{ matrix}\}$. Row i reveals the memberships of each observation in the class i . Column j shows how the membership of a given observation (x_j) is distributed between the classes. The error function (1) can be minimized if the fuzzy memberships are of the following form [2],

$$\mu_{ik} = \frac{(\frac{1}{d_{ik}^2})^{\frac{1}{q-1}}}{\sum_{j=1}^c (\frac{1}{d_{jk}^2})^{\frac{1}{q-1}}} \quad (3)$$

The algorithm iterates, updating the cluster centers and the fuzzy memberships of each observation until some stopping criterion is reached.

SUPERVISED FCM

Key *et al.* [11] noted that a supervised classifier could be implemented using (3) to determine fuzzy memberships. This was used in Foody [9]. In this algorithm, training sets

are selected as representative of desired classes. Training-class *means* and *co-variance* matrices (non-fuzzy) are computed and used to calculate the (squared) Mahalanobis distance between each pixel and the mean of each class,

$$d_{ik}^2 = (x_k - m_i)^t \Sigma_i^{-1} (x_k - m_i) \quad (4)$$

where d_{ik}^2 represents the (squared) Mahalanobis distance between pixel x_k and the mean of class i , m_i ; superscript t denotes (vector) transpose; and Σ_i^{-1} is the inverse of the covariance matrix of the i th class.

Fuzzy class memberships are calculated according to

$$\mu_{ik} = \frac{(\frac{1}{d_{ik}^2})^m}{\sum_{j=1}^c (\frac{1}{d_{jk}^2})^m}, \text{ where } i = 1, \dots, c \text{ and } k = 1, \dots, n$$

This substitution of m for q in (1) (2) and (3) is made for notational ease, noting that $m = \frac{1}{q-1}$.

PARAMETER SETTING FOR m

With both the unsupervised FCM clustering algorithm and the supervised Mahalanobis Distance fuzzy classifier described above, *a priori* choices about the number of classes c and the fuzzy exponent m are required. When using the supervised Mahalanobis Distance fuzzy classifier for a specific set of data, the number of classes c is largely determined by the application: the value of the fuzzy exponent m remains, however, problematic.

Bezdek *et al.* [2] suggested that for the unsupervised FCM clustering algorithm q should be in the range 1 to 30, with the range 1.5 to 3 giving good results — these values for q correspond to values of m in the range positive ∞ to 0.03, with good results in the range 2.0 to 0.5. The authors noted no strong theoretical or empirical evidence for these choices. Cannon *et al.* [4] also noted no theoretical basis for choosing a good value suggesting $1.1 \leq q \leq 5$ as the most useful range. McBratney and Moore [12] found a value of q approximately 2 ($m = 1$) optimal (but dependent on the number of classes). Choe and Jordan [6] found the unsupervised FCM algorithm insensitive to the value of q while in the range 1.1 to 30. They further suggest that the value $q = 12$ ($m \approx 0.09$) as optimal. Cannon and Jacobs [5] found for image applications that $1.1 \leq q \leq 2.5$ proved adequate for practical purposes. Foody [9] used the value $q = 1.25$ for the supervised Mahalanobis Distance fuzzy classifier stating that this value was qualitatively assessed to give good results.

FUZZY IMAGE CLASSIFICATION

Many image classification problems are characterized by gradual class transitions: this makes fuzzy classification quite appealing in this domain. We also commonly find what is termed the mixed pixel problem. This occurs when there are a number of classes contributing to the observed spectral response of a pixel in an image. Earlier work on interpreting fuzzy set memberships of pixels in remotely sensed images as mixed pixel class proportions has produced good results. Fisher and Pathirana [8], Pathirana [15], Wang [16], Foody [9], Nishida *et al.* [14], Foody and Cox [10], Maselli *et al.* [13] and Atkinson *et al.* [1] all suggest a strong relationship between fuzzy memberships and proportions of ground cover. Although Campbell and Hashim [3] question the interpretation of fuzzy class memberships as proportional representation, noting no theoretical justification for it, such interpretation has considerable support in the literature: this offers an approach to determining a suitable value for the fuzzy exponent m in the supervised Mahalanobis Distance fuzzy classifier (and it is expected that the results could reasonably be extended to guide selection of q in the FCM).

INVESTIGATION OF THE VALUE OF m

Deer and Eklund [7] investigated empirically the values of m for the case of two classes of equal proportions. It is also possible to investigate the two class mixed pixel case analytically (allowing for arbitrary proportions). Consider a simple situation in which there are two classes and two image bands. Let the mean of class 1 be $\bar{\mu}_1 = (x_1, y_1)^t$ and the mean of class 2 be $\bar{\mu}_2 = (x_2, y_2)^t$.

Consider a mixed pixel in a real proportion $\lambda_{1p} : \lambda_{2p}$. For convenience we omit the subscript p . Without loss of generality we require, $\lambda_1 + \lambda_2 = 1$. Under the linear mixing assumption we would expect such a mixed pixel to exhibit a spectral vector $\bar{x}_p = (x_p, y_p)^t$ given by,

$$\bar{x}_p = ((\lambda_1 x_1 + \lambda_2 x_2), (\lambda_1 y_1 + \lambda_2 y_2))^t$$

Denote the class 1 (squared) Mahalanobis Distance of $\bar{x}_p$ by $(D_{1p})^2$. Now,

$$\begin{aligned} (D_{1p})^2 &= (\bar{x}_p - \bar{\mu}_1)^t \Sigma_1^{-1} (\bar{x}_p - \bar{\mu}_1) \\ &= ((\lambda_1 x_1 + \lambda_2 x_2) - x_1, (\lambda_1 y_1 + \lambda_2 y_2) - y_1)^t \\ &\quad \Sigma_1^{-1} ((\lambda_1 x_1 + \lambda_2 x_2) - x_1, (\lambda_1 y_1 + \lambda_2 y_2) - y_1) \\ &= (\lambda_2 x_2 - (1 - \lambda_1)x_1, \lambda_2 y_2 - (1 - \lambda_1)y_1)^t \\ &\quad \Sigma_1^{-1} (\lambda_2 x_2 - (1 - \lambda_1)x_1, \lambda_2 y_2 - (1 - \lambda_1)y_1) \end{aligned}$$

but $(1 - \lambda_1) = \lambda_2$ therefore,

$$\begin{aligned} (D_{1p})^2 &= (\lambda_2 x_2 - \lambda_2 x_1, \lambda_2 y_2 - \lambda_2 y_1)^t \Sigma_1^{-1} (\lambda_2 x_2 - \lambda_2 x_1, \lambda_2 y_2 - \lambda_2 y_1) \\ &= \lambda_2^2 ((x_2 - x_1), (y_2 - y_1))^t \Sigma_1^{-1} ((x_2 - x_1), (y_2 - y_1)) \\ &= \lambda_2^2 (D_{12})^2 = ((1 - \lambda_1)^2 (D_{12})^2) \end{aligned}$$

Similarly $(D_{2p})^2 = \lambda_1^2 (D_{21})^2 = ((1 - \lambda_2)^2 (D_{21})^2)$ where $(D_{12})^2$ is the class 1 (squared) Mahalanobis Distance of $\bar{\mu}_2$

(the mean of class 2), and $(D_{21})^2$ is the class 2 (squared) Mahalanobis Distance of $\bar{\mu}_1$ (the mean of class 1).

This development permits us to derive analytically a value for the exponent m in the Mahalanobis Distance fuzzy classifier. In the simple case where there are only two classes, we denote the fuzzy set membership of the pixel p in the class j ,

$$\mu_{jp} = \frac{\left(\frac{1}{D_{jp}^2}\right)^m}{\left(\frac{1}{D_{1p}^2}\right)^m + \left(\frac{1}{D_{2p}^2}\right)^m} \quad j = 1, 2.$$

Interpreting fuzzy set memberships as proportions of mixed pixels leads to $\mu_{1p} : \mu_{2p} = \lambda_1 : \lambda_2$, i.e. the ratio of memberships should equal the ratio of proportions. Therefore we require

$$\frac{\mu_{1p}}{\mu_{2p}} = \frac{\left(\frac{1}{D_{1p}^2}\right)^m}{\left(\frac{1}{D_{2p}^2}\right)^m} = \frac{\lambda_1}{\lambda_2}$$

$$\frac{\left(\frac{1}{\lambda_2^2 (D_{12})^2}\right)^m}{\left(\frac{1}{\lambda_1^2 (D_{21})^2}\right)^m} = \frac{\lambda_1}{\lambda_2}$$

$$\left(\frac{\lambda_1}{\lambda_2}\right)^{2m} \left(\frac{D_{21}}{D_{12}}\right)^{2m} = \frac{\lambda_1}{\lambda_2}$$

$$\left(\frac{D_{12}}{D_{21}}\right)^{2m} - \left(\frac{\lambda_1}{\lambda_2}\right)^{2m-1} = 0$$

For convenience, denote the ratios $\left(\frac{D_{12}}{D_{21}}\right)$ and $\left(\frac{\lambda_1}{\lambda_2}\right)$ as K_1 and K_2 respectively.

$$\begin{aligned} (K_1)^{2m} &= (K_2)^{2m-1} \\ 2m \ln K_1 &= (2m - 1) \ln K_2 \\ \ln K_2 &= 2m(\ln K_2 - \ln K_1) \\ m &= \frac{\ln K_2}{2(\ln K_2 - \ln K_1)} \end{aligned} \quad (5)$$

Note that: (i) this derivation is not valid for $K_1 = K_2$ (divide by zero); (ii) if $K_2 = 1$, then $\ln K_2 = 0$, ($m = 0$, and this case is uninteresting); and (iii) if $K_1 = 1$, then $\ln K_1 = 0$, and $m = 0.5$, irrespective of the value of K_2 .

More generally the ideal value for the fuzzy exponent m in the Mahalanobis Distance Fuzzy Classifier (under the requirement that the fuzzy memberships equal proportions in the mixed pixel case), is shown to be a function of both the ratio of the Mahalanobis Distances, and the ratio of proportions. This shows we cannot select an optimal (in the sense of returning fuzzy class memberships in proportions of mixtures) value for m unless we already know the class proportions.

We can explore (5) further. If $K_2 < 1$, then $\ln K_2 < 0$. Recall we require $m \geq 0$, then $(\ln K_2 - \ln K_1)$ must be < 0 . Therefore $\ln K_1 > \ln K_2$. Therefore $K_1 > K_2$. Therefore $K_1 < K_2 < 1$ implies that no analytic solution is possible. If $K_2 > 1$, then $\ln K_2 > 0$. For $m \geq 0$, then $(\ln K_2 - \ln K_1)$ must be > 0 . Therefore $\ln K_2 > \ln K_1$. Therefore $K_2 >$

K_1 . Therefore $K_1 > K_2 > 1$ also implies that no solution is possible. All other cases lead to analytic solutions.

Cases for equal Mahalanobis Distances (i.e. $K_1 = 1$), and $K_1 = K_2 = 1$, are trivial. The case $K_1 = K_2$ is generally precluded. There are six remaining, more general cases:

$K_1 > 1$, then

$$K_2 > K_1 > 1 \quad (6)$$

$$K_1 > K_2 \geq 1 \quad (7)$$

$$K_1 > 1 \geq K_2 \quad (8)$$

$K_1 < 1$, then

$$K_2 < K_1 < 1 \quad (9)$$

$$K_1 < K_2 \leq 1 \quad (10)$$

$$K_1 < 1 \leq K_2 \quad (11)$$

To examine these relationships numerically, we require some typical values for K_1 (for K_2 , the ratio of class proportions it suffices to select values 0.1 to 0.9 in increments of 0.1). We can derive typical values for K_1 from the training data in our case study.

VALUES FROM CASE STUDY

Our study area is located on the western end of Kangaroo Island, South Australia. Landsat 5 TM (partial) image datasets were used (1024 × 1024 pixels, bands 2, 3, 4 and 5), and a digital ground truth dataset. The 'ground truth' was compiled for the same date as the image data by interpretation of photographic imagery supported by limited field survey. There were 12 classes in ground truth dataset: forest (normal), forest (uncertain/unknown), forest (modified), natural vegetation (other), sea, swamp, lake, pine, pasture/cropland, 3 road classes (vehicular track, sealed road, and unsealed road). The TM image was rectified/registered to the ground truth image using a polynomial warping approach with nearest neighbor re-sampling.

A number of regions were selected for each training class by inspection of the image and consultation with the ground truth. The regions did not include boundary areas: no attempt was made to deliberately include mixed pixels in the training data. The data from the selected regions were grouped to form the training class data for each class. The classes of primary interest were *Forest*, *Sea*, *Pasture* and *Pine*.

Denote class i 's Mahalanobis Distance (Eqn (4)) to the mean of class j as D_{ij}^2 , the *Forest* Mahalanobis Distance of the mean of class *Sea* as D_{fs}^2 , ..., the *Pine* Mahalanobis Distance of the mean of class *Forest* as $D_{\pi f}^2$, ..., and the *Pine* Mahalanobis Distance of the mean of class *Pasture* as $D_{\pi p}^2$. Values for ratios of Mahalanobis Distances (i.e. K_1) from training data are shown in Table I.

Table I indicates that the *Forest* Mahalanobis Distance to the mean of class *Sea* divided by the *Sea* Mahalanobis Distance the mean of class *Forest* in 0.15, and so on. Table I reveals that ratios of Mahalanobis Distances can typically be found in the range 0.15 to 7. Using values in this range, we can investigate the various possibilities given by (6) to (11). Recall that the requirement is to solve (5), rewritten,

	D_{ij}^2/D_{ji}^2	D_{ji}^2/D_{ij}^2	D_{ij}/D_{ji}	D_{ji}/D_{ij}
D_{fs}/D_{sf}	0.02	44.83	0.15	6.70
D_{fp}/D_{pf}	0.32	3.12	0.57	1.77
$D_{f\pi}/D_{\pi f}$	0.06	15.53	0.25	3.94
D_{sp}/D_{ps}	27.49	0.04	5.24	0.19
$D_{s\pi}/D_{\pi s}$	13.79	0.07	3.71	0.27
$D_{p\pi}/D_{\pi p}$	0.13	7.81	0.36	2.79

TABLE I
EMPIRICAL VALUES FOR THE RATIOS OF CLASS MAHALANOBIS
DISTANCES TO CLASS MEANS.

$$y = (K_1)^{2x} - (K_2)^{2x-1}$$

for $y = 0$. Values for K_1 and K_2 selected appropriately enable us to find the correct value for the fuzzy exponent value (represented by x) and also permit us to examine the sensitivity of our solution. Case (6), ($K_2 > K_1 > 1$) – (with typical values inserted from Table I) – shown graphically in Fig. 1, has a (positive) root at 0.8.

Fig. 1. Solving for Fuzzy Exponent m when $K_2 > K_1 > 1$

Case (7), ($K_1 > K_2 \geq 1$) is plotted in Fig. 2. This has no positive root (it is one of the cases identified earlier as having no solution).

Case (8), ($K_1 > 1 \geq K_2$) is plotted in Fig. 3, and has a root at about 0.4. Case (9), ($K_2 < K_1 < 1$) is plotted in Fig. 4, and has a root at 0.6. Case (10), ($K_1 < K_2 \leq 1$) is plotted in Fig. 5. This has no (positive finite) root (but is satisfactory for practical purposes at about $x = 3$). Eqn (10) was shown to have no analytic solution however in practice a value of $m = 3$ will suffice. Case (11), ($K_1 < 1 \leq K_2$) is plotted in Fig. 6, showing a root at $x = 0.4$.

Therefore we can say that where the equation for the analytic value for the fuzzy exponent (for the two class mixed pixel case) has positive (and finite) roots found in the range 0.25 to 1.5 (over a range of 'reasonable' values for ratios of distances and ratios of proportions); figures which accord reasonably well with earlier empirical investigations. It is noted that in several of the cases plotted, the slope of the graph around the root ($y = 0$) suggests a high degree of sensitivity to the value of the fuzzy exponent.

CONCLUSION

A theoretical investigation of the value of m is presented subject to the condition that the fuzzy memberships reflect the proportions of contributing classes for the two class mixed pixel case. An expression is derived for the value of

Fig. 2. Solving for Fuzzy Exponent m when $K_1 > K_2 \geq 1$

Fig. 3. Solving for Fuzzy Exponent m when $K_1 > 1 \geq K_2$

m . This is shown in general to be a function of both the ratio of the class distance metrics (i.e. the observed data) and the ratio of proportions of contributing classes (the very thing which we would often seek to estimate with the fuzzy classification approach). Some special cases are described; most notably, that $m = 0.5$ returns the ideal fuzzy memberships for all class proportions in the case of identical distributions, and that, for some combinations of ratios of Mahalanobis Distances and class proportions, there is no correct choice for m (i.e. no value which will return fuzzy memberships equaling class proportions). Because of the relationship between this classifier and the FCM clustering algorithm, it is expected that these results will be extensible to FCM.

REFERENCES

- [1] Atkinson P.M., M.E.J. Cutler and H. Lewis, Mapping Sub-Pixel Proportional Land Cover With AVHRR Imagery. *International Journal of Remote Sensing*, 18, pp. 917-935, 1997.
- [2] Bezdek J.C., R. Ehrlich and W. Full. FCM: The Fuzzy c -Means Clustering Algorithm. *Computers and Geoscience*, 10, pp. 191 - 203, 1984.
- [3] Campbell J.G. and A.A. Hashim. Fuzzy Sets, Pattern Recognition, Linear Estimation, and Neural Networks - a Unification of the Theory with Relevance to Remote Sensing. *Proceedings of the 18th Annual Conference of the Remote Sensing Society* Cracknell A.P. and R.A. Vaughan (eds.) University of Dundee, pp. 508-517, 1992.
- [4] Cannon R.L., J.V. Dave and J.C. Bezdek. Efficient Implementation of the Fuzzy c -Means Clustering Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, pp. 248-255, 1986.
- [5] Cannon R.L. and C. Jacobs. *Multispectral Pixel Classification with Fuzzy Objective Functions*. Center for Automation Research Technical Report CAR-TR-51, University of Maryland, College Park, 1984.
- [6] Choe H. and J.B. Jordan. On the Optimal Choice of Parameters in a Fuzzy c -Means Algorithm. *Proc. of the IEEE International Conf. on Fuzzy Systems* pp. 349-354, 1992.
- [7] Deer P.J. and P.W. Eklund. On Parameter Values For a Mahalanobis Distance Fuzzy Classifier. 8th Int. Conf. on Information Processing and the Management of Uncertainty (IPMU00) pp. 1957-1960, Consejo Superior de Investigaciones Científicas, Madrid, Spain, 2000.
- [8] Fisher P.F. and S. Pathirana. The Evaluation of Fuzzy Membership of Land Cover Classes in the Suburban Zone. *Remote Sensing of Environment*, 34, pp. 121-132, 1990.
- [9] Foody G.M. A Fuzzy Sets Approach to the Representation of Vegetation Continua from Remotely Sensed Data: An Example from Lowland Heath. *Photogrammetric Engineering and Remote Sensing*, 58, pp. 221 - 225, 1992.
- [10] Foody G.M. and D.P. Cox. Sub-Pixel Land Cover Composition Estimation Using a Linear Mixture Model and Fuzzy Set Membership. *International Journal of Remote Sensing*, 15, pp. 619-631, 1994.
- [11] Key J.R., J.A. Maslanik and R.G. Barry. Cloud Classification From Satellite Data Using a Fuzzy Sets Algorithm: A Polar Example. *International Journal of Remote Sensing*, 10, pp. 1823-1842, 1989.
- [12] McBratney A.B. and A.W. Moore. Application of Fuzzy Sets to Climatic Classification. *Agricultural and Forest Meteorology*, 35, pp. 165-185, 1985.
- [13] Maselli F., A. Rodolfi and C. Conese. Fuzzy Classification of Spatially Degraded Thematic Mapper Data for Estimation of Sub-Pixel Components. *International Journal of Remote Sensing*, 17, pp. 537-551, 1996.
- [14] Nishida M., K. Hayashi, N. Yoshimura and K. Ohtsuka. Application of Fuzzy Reasoning to Estimation of Class Mixture Proportion of Mixed Pixel on Remote Sensing Images. *Proceedings of the 1993 International Geoscience and Remote Sensing Symposium (IGARSS 93)*, pp. 1965-1968, 1993.
- [15] Pathirana S. *Fuzzy Membership Approach to the Mixed Pixel Problem of Remotely Sensed Data: An Application to a Sub-Urban Fringe Zone of Northeast Ohio*. PhD dissertation, Kent State University, 1990.
- [16] Wang F. Fuzzy Supervised Classification of Remotely Sensed Images. *IEEE Transactions on Geoscience and Remote Sensing*, 28, pp. 194-201, 1990.

Fig. 4. Solving for Fuzzy Exponent m when $K_2 < K_1 < 1$

Fig. 5. Solving for Fuzzy Exponent m when $K_1 < K_2 \leq 1$

Fig. 6. Solving for Fuzzy Exponent m when $K_1 < 1 \leq K_2$