ELSEVIER

Information Sciences 111 (1998) 1–17

INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL

Fuzzy feature evaluation index and connectionist realization – II: Theoretical analysis

Jayanta Basak [1], Rajat K. De *, Sankar K. Pal [2]

Machine Intelligence Unit, Indian Statistical Institute, Calcutta 700035, India

Received 26 November 1996; received in revised form 11 February 1997; accepted 15 August 1997

Abstract

The present article deals with a theoretical analysis of our earlier investigation [1] where we developed a neuro-fuzzy model for feature evaluation. This includes derivation of a fixed upper bound and a varying lower bound of the feature evaluation index. The monotonic increasing behavior of the feature evaluation index with respect to the lower bound is established. A relation of the evaluation index (lower bound) with interclass distance and weighting coefficient is also derived.

1. Introduction

The present study is a continuation of our previous investigation [1] in which a new fuzzy set theoretic feature evaluation index, in terms of individual class membership, was, first of all, defined for ranking the importance of features (or subsets of features). A neuro-fuzzy approach was then provided by designing a new connectionist model in order to perform the task of optimizing the aforesaid fuzzy evaluation index incorporating weighted distance for computing class membership values. A set of weighting coefficients representing the importance of the individual features was obtained by this optmization

* Corresponding author. E-mail: res9318@isical.ernet.in

[1] Tel.: +91 33 57 78085; e-mail: jayanta@isical.ernet.in

[2] E-mail: sankar@isical.ernet.in

PII: S0020-0255(97)10031-7

process. These weighting coefficients led to a transformation of the feature space for flexible modeling of class structures.

The present article provides a theoretical analysis of the aforesaid feature evaluation index. Here, a fixed upper bound and a varying lower bound of the evaluation index are derived. It is shown theoretically that the index monotonically increases with the lower bound. A relation of the evaluation index (lower bound) with interclass distance and weighting coefficient is established. Section 2 provides, in brief, our previous work on the fuzzy feature evaluation index and its connectionist realization [1] for the convenience of the readers. This is followed by mathematical analysis in Section 3.

2. Fuzzy feature evaluation index and connectionist realization

Let us consider an n-dimensional feature space Ω containing $x_1, x_2, x_3, \ldots, x_i, \ldots, x_n$ features (components). Let there be M classes $C_1, C_2, C_3, \ldots, C_k, \ldots, C_M$. The feature evaluation index for a subset $(\Omega_{\mathbf{x}})$ containing few of these n features, is defined as [1],

$$E = \sum_k \sum_{\mathbf{x} \in C_k} \frac{s_k(\mathbf{x})}{\sum_{k' \neq k} s_{kk'}(\mathbf{x})} \times \alpha_k, \tag{1}$$

where $\mathbf{x}$ is constituted by the features of $\Omega_{\mathbf{x}}$ only.

$$s_k(\mathbf{x}) = \mu_{C_k}(\mathbf{x}) \times (1 - \mu_{C_k}(\mathbf{x})) \tag{2}$$

and

$$s_{kk'}(\mathbf{x}) = \frac{1}{2}\left[\mu_{C_k}(\mathbf{x}) \times (1 - \mu_{C_{k'}}(\mathbf{x}))\right] + \frac{1}{2}\left[\mu_{C_{k'}}(\mathbf{x}) \times (1 - \mu_{C_k}(\mathbf{x}))\right]. \tag{3}$$

$\mu_{C_k}(\mathbf{x})$ and $\mu_{C_{k'}}(\mathbf{x})$ are the membership values of the pattern $\mathbf{x}$ in classes C_k and $C_{k'}$ respectively. α_k is the normalizing constant for class C_k which takes care of the effect of relative sizes of the classes [1].

Note that, s_k is zero (minimum) if $\mu_{C_k} = 1$ or 0, and is 0.25 (maximum) if $\mu_{C_k} = 0.5$. On the other hand, $s_{kk'}$ is zero (minimum) when $\mu_{C_k} = \mu_{C_{k'}} = 1$ or 0, and is 0.5 (maximum) for $\mu_{C_k} = 1$, $\mu_{C_{k'}} = 0$ or vice-versa. Therefore, the term $s_k / \sum_{k' \neq k} s_{kk'}$ is minimum if $\mu_{C_k} = 1$ and $\mu_{C_{k'}} = 0$ for all $k' \neq k$, i.e., if the ambiguity in the belongingness of a pattern $\mathbf{x}$ to classes C_k and $C_{k'}$ $\forall k' \neq k$ is minimum (i.e., the pattern belongs to only one class). It is maximum when $\mu_{C_k} = 0.5$ for all k. In other words, the value of E decreases as the belongingness of the patterns increases for only one class (i.e., compactness of individual classes increases) and at the same time decreases for other classes (i.e., separation between classes increases). E increases when the patterns tend to lie at the boundaries between classes (i.e., $\mu \to 0.5$). The objective is, therefore, to select those features for which the value of E is minimum.

The membership ($\mu_{C_k}(\mathbf{x})$) of a pattern $\mathbf{x}$ to a class C_k is defined with a multi-dimensional π-function [2] which is given by,

$$\mu_{C_k}(\mathbf{x}) = \begin{cases} 1 - 2d_k^2(\mathbf{x}) & 0 \leqslant d_k(\mathbf{x}) < \frac{1}{2}, \\ 2[1 - d_k(\mathbf{x})]^2 & \frac{1}{2} \leqslant d_k(\mathbf{x}) < 1, \\ 0 & \text{otherwise}. \end{cases} \quad (4)$$

$d_k(\mathbf{x})$ is the distance of the pattern $\mathbf{x}$ from $\mathbf{m}_k$ (the center of class C_k). It can be defined as,

$$d_k(\mathbf{x}) = \left[\sum_i \left(\frac{x_i - m_{ki}}{\lambda_k}\right)^2\right]^{\frac{1}{2}}, \quad (5)$$

where

$$\lambda_k = 2 \max_{\mathbf{x} \in C_k}[\|\mathbf{x} - \mathbf{m}_k\|], \quad (6)$$

and

$$m_{ki} = \frac{\sum_{\mathbf{x} \in C_k} x_i}{|C_k|}. \quad (7)$$

Eqs. (4)–(7) are such that the membership $\mu_{C_k}(\mathbf{x})$ of a pattern $\mathbf{x}$ is 1 if it is located at the mean of C_k, and 0.5 if it is at the boundary (i.e., ambiguous region) for a symmetric class structure.

In practice, the class structure may not be symmetric. In that case, the membership values of some patterns at the boundary of the class will be greater that 0.5. Also, some patterns of other classes may have membership values greater than 0.5 for the class under consideration. For handling this undesirable situation, the membership function corresponding to a class has been transformed [1], incorporating a weighting factor corresponding to a feature.

For this purpose, we defined weighted distance from Eq. (5) as [1],

$$d_k(\mathbf{x}) = \left[\sum_i w_i^2 \left(\frac{x_i - m_{ki}}{\lambda_k}\right)^2\right]^{\frac{1}{2}}, \quad w_i \in [0, 1]. \quad (8)$$

The membership values (μ) of the sample points of a class become dependent on w_i. The values of w_i (< 1) make the function of Eq. (4) flattened along the axis of x_i. The lower the value of w_i, the higher is the extent of flattening.

In pattern recognition literature, the weight w_i (in Eq. (8)) can be viewed to reflect the relative importance of the feature x_i in measuring the similarity (in terms of distance) of a pattern to a class. It is such that the higher the value of w_i, the more is the importance of x_i in characterizing/discriminating a class/between classes. $w_i = 1$ (0) indicates most (least) importance of x_i.

Therefore, E (Eq. (1)) is now essentially a function of $\mathbf{w}$ ($= [w_1, w_2, \ldots, w_n]$), if we consider all the n features together. The problem of feature selection/

ranking thus reduces to finding a set of w_is for which E become minimum. The task of minimization has been performed with gradient descent technique in a connectionist framework (because of its massive parallelism, fault tolerance and adaptivity) for minimizing E. A new connectionist model [1] has been developed for this purpose. When the network attains a local minimum during training, the weights of the feedforward links ($\propto w_i^2$) indicate the order of importance of the individual features.

3. Theoretical analysis

Here, we analyze mathematically the characteristics of the feature evaluation index (E) and the significance of weighting coefficients (w_i). In this regard we investigate the following:

- A fixed upper bound and a varying lower bound of $\mathscr{E}(E(\mathbf{x}))$ ($\mathscr{E}$ being the 'expectation' operator and $E(\mathbf{x})$ being the contribution of a pattern $\mathbf{x}$ to the evaluation index E) are derived. The variation of $\mathscr{E}(E(\mathbf{x}))$ with respect to the lower bound is studied.
- A relation between the lower bound, w_i and interclass distance is derived.

3.1. Upper bound and lower bound of $\mathscr{E}(E(x))$

We can write (Eq. (1)) as,

$$E = \sum_{\mathbf{x}}\sum_{k} \frac{\mu_k \times (1-\mu_k)\alpha_k}{\frac{1}{2}\sum_{k'\neq k}[\mu_k \times (1-\mu_{k'}) + \mu_{k'} \times (1-\mu_k)]}, \tag{9}$$

where $\mu_k = \mu_{C_k}(\mathbf{x})$ and $\mu_{k'} = \mu_{C_{k'}}(\mathbf{x})$. Let, $E = \sum_{\mathbf{x}} E(\mathbf{x}) = \sum_{\mathbf{x}}\sum_k E_k(\mathbf{x})$, where

$$E(\mathbf{x}) = \sum_{k} \frac{\mu_k \times (1-\mu_k)\alpha_k}{\frac{1}{2}\sum_{k'\neq k}[\mu_k \times (1-\mu_{k'}) + \mu_{k'} \times (1-\mu_k)]} \tag{10}$$

and

$$E_k(\mathbf{x}) = \frac{\mu_k \times (1-\mu_k)\alpha_k}{\frac{1}{2}\sum_{k'\neq k}[\mu_k \times (1-\mu_{k'}) + \mu_{k'} \times (1-\mu_k)]}. \tag{11}$$

That is, $E(\mathbf{x})$ is the contribution of a pattern $\mathbf{x}$ to the evaluation index (E), and $E_k(\mathbf{x})$ is that corresponding to the class C_k. For a pattern $\mathbf{x}$ in class C_k,

$$\frac{1}{2}\sum_{k'\neq k}[\mu_k(1-\mu_{k'}) + \mu_{k'}(1-\mu_k)] = \frac{1}{2}\sum_{k'\neq k}[\mu_k(1-\mu_k) + (\mu_k-\mu_{k'})^2 + \mu_{k'}(1-\mu_{k'})].$$

Since $[(\mu_k-\mu_{k'})^2 + \mu_{k'}(1-\mu_{k'})] \geqslant 0$,

$$\frac{1}{2}\sum_{k'\neq k}[\mu_k(1-\mu_{k'})+\mu_{k'}(1-\mu_k)] \geqslant \frac{M-1}{2}\mu_k(1-\mu_k),$$

where M is the number of classes. Since, $0 < \alpha_k < 1$, we can write,

$$E(\mathbf{x}) \leqslant \frac{2M}{(M-1)}. \tag{12}$$

Therefore,

$$\mathscr{E}(E(\mathbf{x})) \leqslant \frac{2M}{M-1}, \tag{13}$$

where $\mathscr{E}$ denotes the 'mathematical expectation' operator.

Again, for a pattern $\mathbf{x}$ in class $C_k, \mu_k, \mu_{k'} \in [0,1]$, we can write,

$$\frac{1}{2}[\mu_k(1-\mu_{k'})+\mu_{k'}(1-\mu_k)] \leqslant \frac{1}{2},$$

$$\sum_{k'\neq k}\frac{1}{2}[\mu_k(1-\mu_{k'})+\mu_{k'}(1-\mu_k)] \leqslant \frac{1}{2}(M-1),$$

$$\frac{1}{\sum_{k'\neq k}\frac{1}{2}[\mu_k(1-\mu_{k'})+\mu_{k'}(1-\mu_k)]} \geqslant \frac{2}{(M-1)},$$

$$\sum_k \frac{\mu_k(1-\mu_k)\alpha_k}{\sum_{k'\neq k}\frac{1}{2}[\mu_k(1-\mu_{k'})+\mu_{k'}(1-\mu_k)]} \geqslant \frac{2}{(M-1)}\sum_k \mu_k(1-\mu_k)\alpha_k.$$

Thus,

$$E(\mathbf{x}) \geqslant \frac{2}{(M-1)}\sum_k \mu_k(1-\mu_k)\alpha_k.$$

That is,

$$\mathscr{E}(E(\mathbf{x})) \geqslant \frac{2}{(M-1)}\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right). \tag{14}$$

Therefore,

$$\frac{2}{(M-1)}\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right) \leqslant \mathscr{E}(E(\mathbf{x})) \leqslant \frac{2M}{(M-1)}. \tag{15}$$

Note that, the upper bound (UB) of $\mathscr{E}(E(\mathbf{x}))$ is fixed, whereas the lower bound (LB) is varying with $\frac{2}{(M-1)}\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right)$.

Let us now analyze the behaviour of $E(\mathbf{x})$ with respect to $\sum_k \mu_k(1-\mu_k)$. For this purpose, we substitute $\mu_k(1-\mu_k)$ by h_k in Eq. (11). In that case,

$$\frac{\mathrm{d}E_k(\mathbf{x})}{\mathrm{d}h_k} = \frac{\alpha_k\Big[\sum_{k'\neq k}[\mu_{k'}(1-\mu_k)+\mu_k(1-\mu_{k'})](1-2\mu_k)-\mu_k(1-\mu_k)\sum_{k'\neq k}(1-2\mu_{k'})\Big]}{\frac{1}{2}\Big[\sum_{k'\neq k}[\mu_{k'}(1-\mu_k)+\mu_k(1-\mu_{k'})]\Big]^2(1-2\mu_k)}$$

$$= \nu_k\alpha_k \Big/ \frac{1}{2}\left[\sum_{k'\neq k}[\mu_{k'}(1-\mu_k)+\mu_k(1-\mu_{k'})]\right]^2, \tag{16}$$

where

$$\nu_k = \frac{\sum_{k'\neq k}[\mu_{k'}(1-\mu_k)+\mu_k(1-\mu_{k'})](1-2\mu_k)-\mu_k(1-\mu_k)\sum_{k'\neq k}(1-2\mu_{k'})}{(1-2\mu_k)}. \tag{17}$$

It is clear from Eq. (16) that $\mathrm{d}E_k(\mathbf{x})/\mathrm{d}h_k$ is positive/negative if ν_k is positive/negative. In other words, $E_k(\mathbf{x})$ increases/decreases monotonically with $\mu_k(1-\mu_k)$ if ν_k is positive/negative. Simplifying the expression on the right hand side of Eq. (17) we get,

$$\nu_k = \sum_{k'\neq k}\mu_{k'} - \frac{\mu_k^2\sum_{k'\neq k}(1-2\mu_{k'})}{(1-2\mu_k)}. \tag{18}$$

In order to show that $E_k(\mathbf{x})$ monotonically increases with $\mu_k(1-\mu_k)$ for both *non-overlapping* and *overlapping* class structures, we consider the following cases.

Case 1 (Non-overlapping; (Fig. 1)): Here, for a pattern $\mathbf{x}$, if $\|\mathbf{x}-\mathbf{m}_k\| \leqslant \lambda_k/2$, $\mu_k \geqslant 0.5$ and $\mu_{k'} < 0.5$, $\forall k' \neq k$. Therefore, $\nu_k > 0$ (Eq. (18)), and as a result $\mathrm{d}E_k(\mathbf{x})/\mathrm{d}h_k > 0$. This indicates $E_k(\mathbf{x})$ is monotonically increasing with $\mu_k(1-\mu_k)$.

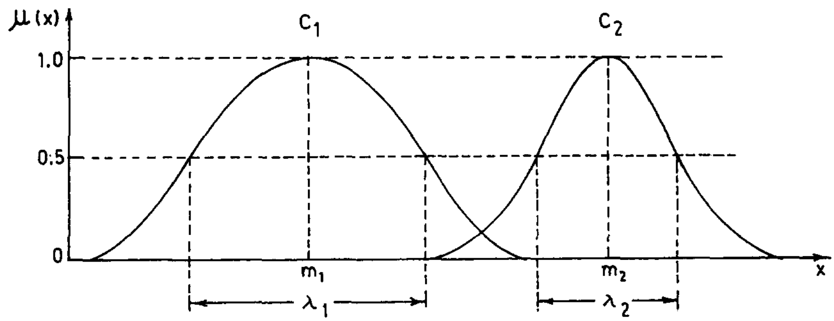

Fig. 1. Non-overlapping pattern classes modeled with π-function.

Case 2 (Overlapping; (Fig. 2)): In this case, for a pattern $\mathbf{x}$, if $\|\mathbf{x} - \mathbf{m}_k\| \leqslant \lambda_k/2$, $\mu_k \geqslant 0.5$ and $\mu_{k'} \gtrless 0.5$, $\forall k' \neq k$. Since the classes are overlapped, we consider two different possibilities: $\mathbf{x}$ lying outside the overlapping zone (i.e., $\|\mathbf{x} - \mathbf{m}_k\| \leqslant \lambda_k/2$ and $\|\mathbf{x} - \mathbf{m}_{k'}\| > \lambda_{k'}/2$) and $\mathbf{x}$ lying within the overlapping zone (i.e., $\|\mathbf{x} - \mathbf{m}_k\| < \lambda_k/2$ and $\|\mathbf{x} - \mathbf{m}_{k'}\| < \lambda_{k'}/2$).

If the pattern $\mathbf{x}$ lies outside the overlapping zone, then $\mu_{k'} < 0.5$ and thereby $v_k > 0$ (Eq. (18)). This indicates $E_k(\mathbf{x})$ monotonically increases with $\mu_k(1 - \mu_k)$.

If $\mathbf{x}$ lies within the overlapping zone, both $\mu_k, \mu_{k'} > 0.5$. Then we have three possibilities: (a) $\mu_k > \mu_{k'}$, (b) $\mu_k \approx \mu_{k'}$ and (c) $\mu_k < \mu_{k'}$.

(a) $\mu_k > \mu_{k'}$: Let $\mu_{k'} = \mu_k - \epsilon_{kk'}$, where $\epsilon_{kk'} > 0$. Therefore, from Eq. (18) we get,

$$v_k = \sum_{k' \neq k} (\mu_k - \epsilon_{kk'}) - \frac{\mu_k^2 \sum_{k' \neq k} (1 - 2\mu_k + 2\epsilon_{kk'})}{(1 - 2\mu_k)}, \tag{19}$$

i.e.,

$$v_k = (M - 1)\mu_k - \sum_{k' \neq k} \epsilon_{kk'} - \frac{2\mu_k^2 \sum_{k' \neq k} \epsilon_{kk'} - \mu_k^2 (2\mu_k - 1)(M - 1)}{1 - 2\mu_k}. \tag{20}$$

Thus, $E_k(\mathbf{x})$ increases monotonically with $\mu_k(1 - \mu_k)$ if

$$(M - 1)\mu_k - \sum_{k' \neq k} \epsilon_{kk'} - \frac{2\mu_k^2 \sum_{k' \neq k} \epsilon_{kk'} - \mu_k^2 (2\mu_k - 1)(M - 1)}{1 - 2\mu_k} > 0. \tag{21}$$

i.e., if

$$\frac{1}{M - 1} \sum_{k' \neq k} \epsilon_{kk'} > - \frac{\mu_k (1 - \mu_k)(2\mu_k - 1)}{(1 - \mu_k)^2 + \mu_k^2}. \tag{22}$$

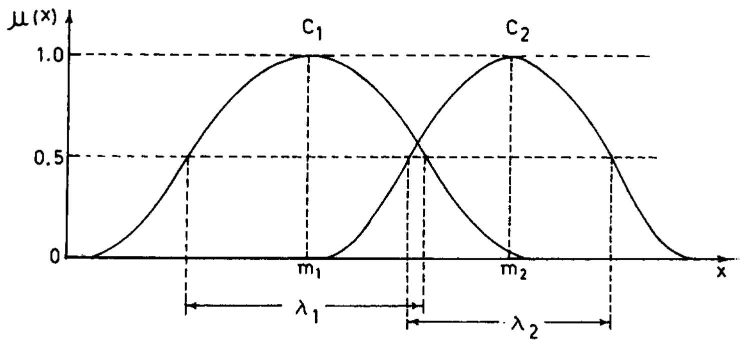

Fig. 2. Overlapping pattern classes modeled with π-function.

Since $\epsilon_{kk'} > 0$, the above inequality always holds, and therefore, in such cases, $E_k(\mathbf{x})$ always increases monotonically with $\mu_k(1 - \mu_k)$.

(b) $\mu_k \approx \mu_{k'}$: In this case, $\epsilon_{kk'} \approx 0$, and therefore, the inequality (22) always holds. Thus, in this case also, we get a monotonic increasing nature of $E_k(\mathbf{x})$ with respect to $\mu_k(1 - \mu_k)$.

(c) $\mu_k < \mu_{k'}$: In this case, $\epsilon_{kk'} < 0$. Let us replace $\epsilon_{kk'}$ by $-\varepsilon_{kk'}$, i.e., $\mu_{k'} = \mu_k + \varepsilon_{kk'}$. Then, the condition for $E_k(\mathbf{x})$ being monotonically increasing function with respect to $\mu_k(1 - \mu_k)$ becomes,

$$\frac{1}{M-1}\sum_{k' \neq k}\varepsilon_{kk'} < \frac{\mu_k(1-\mu_k)(2\mu_k - 1)}{(1-\mu_k)^2 + \mu_k^2}. \tag{23}$$

This condition provides an upper bound on the average value of $\varepsilon_{kk'}$ (hence on the average value of $\mu_{k'}$) that can be allowed in order to get a monotonic increasing behavior of $E_k(\mathbf{x})$ with respect to $\mu_k(1 - \mu_k)$.

First of all, the chance of $\mu_k < \mu_{k'}$ is low for a pattern in class C_k. Even if this happens (say, for overlapping case), the chance of

$$\frac{1}{M-1}\sum_{k' \neq k}\varepsilon_{kk'} > \frac{\mu_k(1-\mu_k)(2\mu_k - 1)}{(1-\mu_k)^2 + \mu_k^2}$$

happening is very low (as illustrated in the following two examples). Therefore, $E_k(\mathbf{x})$ is most likely monotonically increasing with $\mu_k(1 - \mu_k)$.

Example 1. Let, $\mu_1 = 0.6$ for a pattern $\mathbf{x}$ lying within the region $\|\mathbf{x} - \mathbf{m}_1\| < \lambda_1/2$ in class C_1. Then, the condition (23) becomes,

$$\frac{1}{M-1}\sum_{k' \neq k}\varepsilon_{kk'} < 0.1.$$

In order to violate this condition, the average membership value of $\mathbf{x}$ (say, μ_2) to classes other than C_1 should be at least 0.7. It can also be seen that whatever be the value of μ_1 (> 0.5), the value of μ_2 should be greater than μ_1. This is unusual. Thus, we can say that in this case the above inequality (23) will be satisfied and thereby, we can expect a monotonic increasing behavior of $E_1(\mathbf{x})$ with respect to $\mu_1(1 - \mu_1)$.

Example 2. Let, $\mu_1 = 0.5$. In that case, the condition (23) becomes

$$\frac{1}{M-1}\sum_{k' \neq k}\varepsilon_{kk'} < 0.$$

That is, the average membership value of $\mathbf{x}$ to classes other than C_1 should be greater than or equal to 0.5. This situation occurs when the classes are highly overlapped. In other words, if there is high amount of overlap, the behavior of $E(\mathbf{x})$ becomes unpredictable for ambiguous patterns.

Thus, we can say that $E_k(\mathbf{x})$ always monotonically increases with $\mu_k(1-\mu_k)$ except for some ambiguous patterns in highly overlapping regions. If we take average of $E(\mathbf{x})(=\sum_k E_k(x))$, we can expect this average ($\mathscr{E}(E(\mathbf{x}))$) to be always monotonically increasing with $\mathscr{E}(\sum_k \mu_k(1-\mu_k))$, the average of $\sum_k \mu_k(1-\mu_k)$. Therefore, it may be concluded that $\mathscr{E}(E(\mathbf{x}))$ is a monotonically increasing function of $\frac{2}{M-1}\,\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$, as $\frac{2}{M-1}$ and α_k are positive constants.

Note from Eq. (1) that it is difficult to compute $\mathscr{E}(E(\mathbf{x}))$, and so as to determine its relation with w_i and interclass distance. Again, $\mathscr{E}(E(\mathbf{x}))$ is found (in Section 3.1) to be a monotonically increasing function of the lower bound (LB). Therefore, if one can find a relation between LB, w_i and interclass distance, it will reflect the one between $\mathscr{E}(E(\mathbf{x}))$, w_i and interclass distance. The next subsection addresses this issue.

3.2. Relation between LB, interclass distance and w_i

Let us now derive a relation of the lower bound of $\mathscr{E}(E(\mathbf{x}))$ with interclass distance and weighting coefficients for some well defined class structures.

- Let us assume that the classes $C_1, C_2, \ldots, C_k, \ldots, C_M$ have independent, identical Gaussian distributions with respective means $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_k, \ldots, \mathbf{m}_M$ and with the same variance σ^2. Let $\mathscr{P}(\mathbf{x}|C_k)$ be the class-conditional probability density function for class C_k. Then

$$\mathscr{P}(\mathbf{x}|C_k) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\sum_i \frac{(x_i - m_{ki})^2}{2\sigma^2}\right) \tag{24}$$

- Let the membership of a pattern $\mathbf{x}$ in a class C_k be given by,

$$\mu_k = \mu_k(\mathbf{x}) = \exp\left(-\sum_i \frac{(x_i - m_{ki})^2 w_i^2}{2\lambda^2}\right), \tag{25}$$

where λ is the bandwidth of the class C_k, and is the same for all the classes. The lower bound of $\mathscr{E}(E(\mathbf{x}))$ is given by,

$$\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right) = \int_{\mathbf{x}} \left(\sum_k \mu_k(1-\mu_k)\alpha_k\right) \mathscr{P}(\mathbf{x})\,\mathrm{d}\mathbf{x}, \tag{26}$$

where

$$\mathscr{P}(\mathbf{x}) = \sum_k P_k \mathscr{P}(\mathbf{x}|C_k); \tag{27}$$

P_k being a priori probability of class C_k. Evaluating the right hand side of Eq. (26) (see Appendix A), we have

$$\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right) = \rho^n \sum_k \alpha_k P_k \left[\prod_i \left(\frac{1}{\rho^2+w_i^2}\right)^{\frac{1}{2}} - \prod_i \left(\frac{1}{\rho^2+2w_i^2}\right)^{\frac{1}{2}}\right]$$
$$+ \rho^n \sum_k \sum_{k' \neq k} \alpha_k P_{k'} \left[\prod_i \frac{\exp\left(-c_{kk'i}^2 w_i^2/(2\sigma^2(\rho^2+w_i^2))\right)}{(\rho^2+w_i^2)^{1/2}}\right.$$
$$\left. - \prod_i \frac{\exp\left(-c_{kk'i}^2 w_i^2/(\sigma^2(\rho^2+2w_i^2))\right)}{(\rho^2+2w_i^2)^{1/2}}\right], \qquad (28)$$

where $\rho = \lambda/\sigma$ and $c_{kk'i} = m_{ki} - m_{k'i}$ is a measure of interclass distance between the classes C_k and $C_{k'}$ along the feature axis x_i.

Let us consider two classes C_1 and C_2, with two features x_1 and x_2. Let, C_1 and C_2 have unit normal distribution, i.e., $\sigma = 1.0$. Let, $\lambda = 1.0$ and $P_k = \alpha_k = 0.5$ $(\forall k)$. c_{121} and c_{122} are the interclass distances between class C_1 and class C_2 along the feature axes x_1 and x_2, respectively. We now demonstrate graphically the variation of $\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right)$ with respect to c_{121} and c_{122}, and w_1 and w_2.

Fig. 3 shows the variation of $\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right)$ with respect to c_{121} and c_{122} with $w_1 = w_2 = 1$. $\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right)$ is maximum when $c_{121} = c_{122} = 0$,

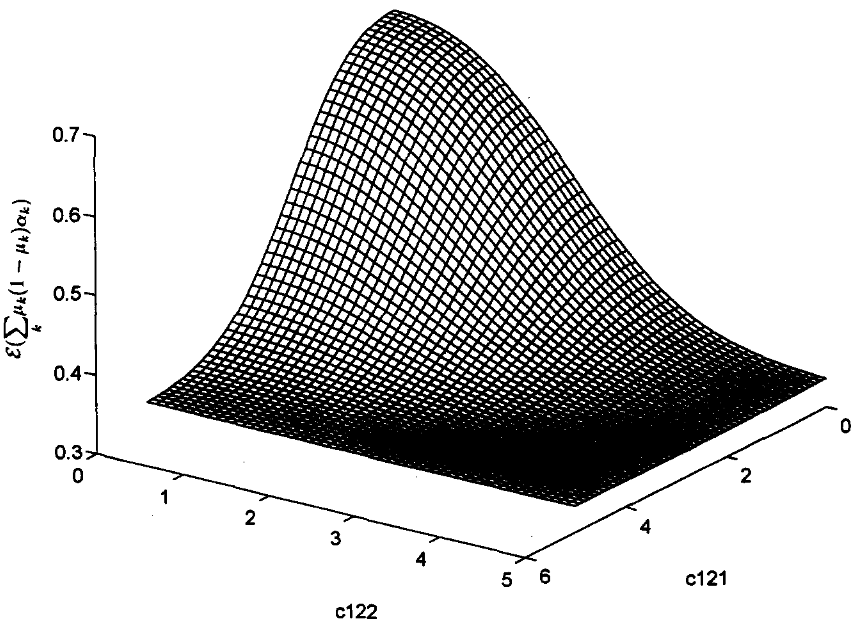

Fig. 3. Graphical representation of $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ with respect to c_{121} and c_{122} with $w_1 = w_2 = 1.0$.

i.e., when the two classes completely overlap. $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ decreases with the increase in c_{121} and c_{122}. This variation is symmetric with respect to both c_{121} and c_{122}. The rate of decrease in $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ also decreases as c_{121} (and c_{122}) increases. Finally, after a certain value of c_{121} (and c_{122}) the rate of decrease in $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ becomes infinitesimally small. This is also evident from the way of computing μ-value where μ_2 of a pattern $\mathbf{x}$ with fixed μ_1 decreases with increase in interclass distance. If the interclass distance exceeds a certain value, μ_2 becomes very small. Thus, the contribution of the pattern to the evaluation index does not get affected further by the extent of the class separation.

Fig. 4(a)–(f) show the variations of $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ with respect to the weighting coefficients w_1 and w_2 for different interclass distances. In Fig. 4(a)–(e) we have considered $c_{122}=0$ throughout whereas c_{121} is considered to be 0.0, 0.5, 1.0, 1.5 and 2.0, respectively. The purpose of this is to show the variation of the extent of the neighborhood region around a local minimum in the $w_1 - w_2$ plane with interclass distance along an axis. Note that, $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ is zero when $w_1 = w_2 = 0$ which is a trivial case. $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ increases as w_1 (and w_2) increase upto a certain value; beyond which $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ decreases upto some local minimum value. Note from Fig. 4(a)–(e) that, the extent of the neighborhood region around a local minimum in the weight space (basin of attraction in the weight space) of the network increases as the interclass distance (c_{121}) increases. This neighborhood region constitutes a zone of operation of the network. Therefore, if interclass distance increases, zone of operation of the network increases, i.e., the freedom of choosing the initial weights increases. In other words, given a set of weights, if the basin of attraction in the weight space is large, the network has higher probability of getting converged into the corresponding local minimum.

Here we have chosen $c_{122}=0$, i.e., the feature x_2 does not have any discriminating power between the classes. This is also illustrated in Fig. 4(a)–(e), where the value of $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ decreases steadily with a decrease in w_2. In other words, for a minimum value of $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ in the zone of operation of the network, w_2 will be very close to zero, which signifies that the feature x_2 is not important for classification. Similarly, a steady decrease in $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ with respect to w_1 in the zone of operation of the network signifies the fact that the feature x_1 is important for classification.

Fig. 4(f) shows the variation of $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ when $c_{121}=c_{122}=3.0$. Here, the variation of $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ is symmetric with respect to both w_1 and w_2, which indicates that both the features x_1 and x_2 are equally important. The zone of operation of the network becomes more or less flat which signifies that the network can settle into any point in this zone. This is due to the fact that any one feature or a weighted combination of the two is sufficient for classification.

4. Conclusions

In this article, we have provided a theoretical analysis for the performance of our earlier investigation on feature evaluation with fuzzy set theory and neural networks [1]. It is shown that the evaluation index has a fixed upper bound and a varying lower bound. The monotonic increasing behavior of the evaluation index with respect to the lower bound is established for different

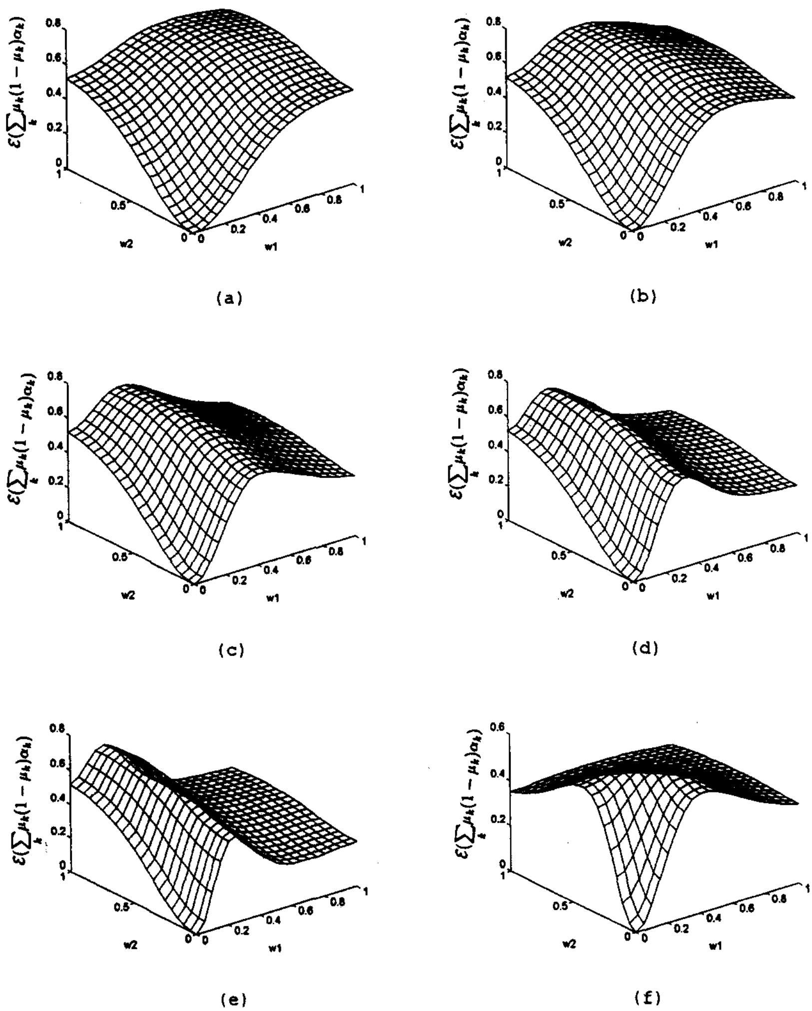

Fig. 4. Graphical representation of $\mathscr{E}(\sum_k \mu_k(1-\mu_k)\alpha_k)$ with respect to w_1 and w_2 for different values of c_{121} and c_{122}.

cases. A relation of the evaluation index (LB), interclass distance and weighting coefficients is derived. It is shown graphically that the zone of operation of the network increases with increase in interclass distance. Given a set of weights, if the zone of operation of the network is large, the network has higher probability of getting converged into the corresponding local optimum.

Acknowledgements

Mr. Jayanta Basak is grateful to Ms. Basabi Chakraborty for the helpful discussions with her. Mr. Rajat K. De is grateful to the Department of Atomic Energy, Government of India, for providing him a Dr. K. S. Krishnan Senior Research Fellowship.

Appendix A. Evaluation of the expression $\mathscr{E}\left(\sum_k \mu_k(-\mu_k)\alpha_k\right)$.

Using Eq. (27), $\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right)$ is given by,

$$\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right) = \sum_k \int_{\mathbf{x}} \mu_k(1-\mu_k)\alpha_k \mathscr{P}(\mathbf{x})\mathrm{d}\mathbf{x}$$

$$= \sum_k \int_{x_1=-=\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} \mu_k(1-\mu_k)\alpha_k \left(P_k \mathscr{P}(\mathbf{x}|C_k) + \sum_{k' \neq k} P_{k'} \mathscr{P}(\mathbf{x}|C_{k'}) \right) \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

Let,

$$J_k = \int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} \mu_k(1-\mu_k)\alpha_k P_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\sum_i \frac{(x_i - m_{ki})^2}{2\sigma^2} \right)$$

$$= \int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} (\mu_k - \mu_k^2)\alpha_k P_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\sum_i \frac{(x_i - m_{ki})^2}{2\sigma^2} \right),$$

$$J_{kk'} = \int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} \mu_k(1-\mu_k)\alpha_k P_{k'} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\sum_i \frac{(x_i - m_{k'i})^2}{2\sigma^2} \right)$$

$$= \int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} (\mu_k - \mu_k^2)\alpha_k P_{k'} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\sum_i \frac{(x_i - m_{k'i})^2}{2\sigma^2} \right),$$

so that

$$\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right) = \sum_k \left(J_k + \sum_{k'\neq k} J_{kk'}\right). \tag{A.1}$$

Let us also assume that,

$$J_{ki1} = \int_{x_i=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i-m_{ki})^2}{2\sigma^2} - \frac{(x_i-m_{ki})^2 w_i^2}{2\lambda^2}\right] \mathrm{d}x_i,$$

$$J_{ki2} = \int_{x_i=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i-m_{ki})^2}{2\sigma^2} - \frac{(x_i-m_{ki})^2 w_i^2}{\lambda^2}\right] \mathrm{d}x_i,$$

$$J_{kk'i1} = \int_{x_i=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i-m_{k'i})^2}{2\sigma^2} - \frac{(x_i-m_{ki})^2 w_i^2}{2\lambda^2}\right] \mathrm{d}x_i,$$

$$J_{kk'i2} = \int_{x_i=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i-m_{k'i})^2}{2\sigma^2} - \frac{(x_i-m_{ki})^2 w_i^2}{\lambda^2}\right] \mathrm{d}x_i,$$

so that

$$J_k = \alpha_k P_k \left(\prod_i J_{ki1} - \prod_i J_{ki2}\right)$$

and

$$J_{kk'} = \alpha_k P_{k'} \left(\prod_i J_{kk'i1} - \prod_i J_{kk'i2}\right).$$

Therefore, from Eq. (A.1) we have

$$\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right) = \sum_k \alpha_k P_k \left(\prod_i J_{ki1} - \prod_i J_{ki2}\right) + \sum_k \sum_{k'\neq k} \alpha_k P_{k'} \left(\prod_i J_{kk'i1} - \prod_i J_{kk'i2}\right). \tag{A.2}$$

For evaluating the integrals J_{ki1}, J_{ki2}, $J_{kk'i1}$ and $J_{kk'i2}$, we use the result of the following integral,

$$J = \int_{-\infty}^{\infty} \exp\left(-(\alpha x^2 + \beta x + \gamma)\right) \mathrm{d}x.$$

Now,

$$J = \int_{-\infty}^{\infty} \exp\left(-\alpha\left(x^2 + 2x\frac{\beta}{2\alpha} + \frac{\beta^2}{4\alpha^2}\right) + \left(\frac{\beta^2}{4\alpha} - \gamma\right)\right) \mathrm{d}x$$
$$= \exp\left(\frac{\beta^2}{4\alpha} - \gamma\right) \int_{-\infty}^{\infty} \exp\left(-\alpha\left(x + \frac{\beta}{2\alpha}\right)^2\right) \mathrm{d}x$$
$$= \exp\left(\frac{\beta^2}{4\alpha} - \gamma\right) \int_{-\infty}^{\infty} \exp(-\alpha y^2) \mathrm{d}y,$$

where

$$y = x + \frac{\beta}{2\alpha}.$$

Therefore,

$$J = 2\exp\left(\frac{\beta^2 - 4\alpha\gamma}{4\alpha}\right) \int_0^{\infty} \exp(-\alpha y^2)\, \mathrm{d}y$$
$$= 2\exp\left(\frac{\beta^2 - 4\alpha\gamma}{4\alpha}\right) \int_0^{\infty} \frac{1}{2\sqrt{\alpha}} \exp(-z) z^{-\frac{1}{2}}\, \mathrm{d}z,$$

where

$$z = \alpha y^2.$$

Hence,

$$J = \frac{\exp\left(\frac{\beta^2 - 4\alpha\gamma}{4\alpha}\right)\sqrt{\pi}}{\sqrt{\alpha}}. \tag{A.3}$$

We use the following transformation for evaluating J_{ki1}, J_{ki2}, $J_{kk'i1}$ and $J_{kk'i2}$.

$$y_i = \left(\frac{x_i - m_{ki}}{\sqrt{2}\lambda}\right) w_i,$$
$$\mathrm{d}x_i = \frac{\sqrt{2}\lambda}{w_i} \mathrm{d}y_i.$$

Then we can write

$$\left[\frac{(x_i - m_{ki})^2 w_i^2}{2\lambda^2} + \frac{(x_i - m_{ki})^2}{2\sigma^2}\right] = y_i^2 + \frac{\rho^2}{w_i^2} y_i^2 = \left(1 + \frac{\rho^2}{w_i^2}\right) y_i^2,$$

$$\left[\frac{(x_i - m_{ki})^2 w_i^2}{\lambda^2} + \frac{(x_i - m_{ki})^2}{2\sigma^2}\right] = 2y_i^2 + \frac{\rho^2}{w_i^2} y_i^2 = \left(2 + \frac{\rho^2}{w_i^2}\right) y_i^2,$$

$$\begin{aligned}\left[\frac{(x_i - m_{ki})^2 w_i^2}{2\lambda^2} + \frac{(x_i - m_{k'i})^2}{2\sigma^2}\right] &= y_i^2 + \frac{[(x_i - m_{ki}) + (m_{ki} - m_{k'i})]^2}{2\sigma^2} \\ &= \left(1 + \frac{\rho^2}{w_i^2}\right) y_i^2 + \frac{(x_i - m_{ki}) c_{kk'i}}{\sigma^2} + \frac{c_{kk'i}^2}{2\sigma^2} \\ &= \left(1 + \frac{\rho^2}{w_i^2}\right) y_i^2 + \frac{\sqrt{2}\lambda y_i c_{kk'i}}{\sigma^2 w_i} + \frac{c_{kk'i}^2}{2\sigma^2},\end{aligned}$$

and

$$\begin{aligned}\left[\frac{(x_i - m_{ki})^2 w_i^2}{\lambda^2} + \frac{(x_i - m_{k'i})^2}{2\sigma^2}\right] &= 2y_i^2 + \frac{[(x_i - m_{ki}) + (m_{ki} - m_{k'i})]^2}{2\sigma^2} \\ &= \left(2 + \frac{\rho^2}{w_i^2}\right) y_i^2 + \frac{(x_i - m_{ki}) c_{kk'i}}{\sigma^2} + \frac{c_{kk'i}^2}{2\sigma^2} \\ &= \left(2 + \frac{\rho^2}{w_i^2}\right) y_i^2 + \frac{\sqrt{2}\lambda y_i c_{kk'i}}{\sigma^2 w_i} + \frac{c_{kk'i}^2}{2\sigma^2}.\end{aligned}$$

Therefore, using the result of J (Eq. (A.3)) we have

$$J_{ki1} = \frac{1}{\sqrt{2\pi}\sigma} \frac{\sqrt{2}\lambda}{w_i} \frac{1}{\left(1 + \frac{\rho^2}{w_i^2}\right)^{1/2}} \sqrt{\pi} = \frac{\rho}{(w_i^2 + \rho^2)^{1/2}},$$

where $\alpha = 1 + (\rho^2/w_i^2)$, $\beta = 0$, and $\gamma = 0$. Similarly, the expressions for J_{ki2}, $J_{kk'i1}$ and $J_{kk'i2}$ are obtained as follows:

$$J_{ki2} = \frac{1}{\sqrt{2\pi}\sigma} \frac{\sqrt{2}\lambda}{w_i} \frac{1}{\left(2 + \frac{\rho^2}{w_i^2}\right)^{1/2}} \sqrt{\pi} = \frac{\rho}{(2w_i^2 + \rho^2)^{1/2}},$$

where $\alpha = 2 + (\rho^2/w_i^2)$, $\beta = 0$, and $\gamma = 0$.

$$\begin{aligned}J_{kk'i1} &= \frac{\rho \exp\left(- (c_{kk'i}^2/(2\sigma^2(1 + (\rho^2/w_i^2))))\right)}{\left(1 + \frac{\rho^2}{w_i^2}\right)^{1/2} w_i} \\ &= \frac{\rho \exp\left(- (c_{kk'i}^2 w_i^2/(2\sigma^2(\rho^2 + w_i^2)))\right)}{(\rho^2 + w_i^2)^{1/2}},\end{aligned}$$

where $\alpha = 1 + (\rho^2/w_i^2)$, $\beta = \sqrt{2}\lambda c_{kk'i}/\sigma^2 w_i$ and $\gamma = c_{kk'i}^2/2\sigma^2$.

$$J_{kk'i2} = \frac{\rho \exp\left(- (c_{kk'i}^2/(\sigma^2(2 + (\rho^2/w_i^2))))\right)}{\left(2 + \frac{\rho^2}{w_i^2}\right)^{1/2} w_i} = \frac{\rho \exp\left(- (c_{kk'i}^2 w_i^2/(\sigma^2(\rho^2 + 2w_i^2)))\right)}{(\rho^2 + w_i^2)^{1/2}},$$

where $\alpha = 2 + (\rho^2/w_i^2)$, $\beta = \sqrt{2}\lambda c_{kk'i}/\sigma^2 w_i$, and $\gamma = c_{kk'i}^2/2\sigma^2$. Therefore, from Eq. (A.2) we have,

$$\mathscr{E}\left(\sum_k \mu_k(1-\mu_k)\alpha_k\right) = \rho^n \sum_k \alpha_k P_k \left[\prod_i \left(\frac{1}{\rho^2 + w_i^2}\right)^{1/2} - \prod_i \left(\frac{1}{\rho^2 + 2w_i^2}\right)^{1/2}\right] + \rho^n \sum_k \sum_{k' \neq k} \alpha_k P_{k'} \left[\prod_i \frac{\exp\left(- c_{kk'i}^2 w_i^2/(2\sigma^2(\rho^2 + w_i^2))\right)}{(\rho^2 + w_i^2)^{1/2}} - \prod_i \frac{\exp\left(- (c_{kk'i}^2 w_i^2/(\sigma^2(\rho^2 + 2w_i^2)))\right)}{(\rho^2 + 2w_i^2)^{1/2}}\right].$$

References

[1] S.K. Pal, J. Basak, R.K. De, Fuzzy feature evaluation index and connectionist realization, Information Sciences, accepted.

[2] S.K. Pal, P.K. Pramanik, Fuzzy measures in determining seed points in clustering, Pattern Recognition Letters 4 (1986) 159–164.