

# Fuzzy Weighted Gaussian Mixture Model for Feature Reduction

Charles.S

Assistant Professor  
Department of Computer Science  
St. Joseph's College, Tiruchirappalli

L. Arockiam

Associate Professor  
Department of Computer Science  
St. Joseph's College, Tiruchirappalli

## ABSTRACT

Feature reduction is one kind of pattern recognition and decision making technique, which can be achieved by using Fuzzy Weighted Gaussian Mixture Model (FWGMM) based on the Gaussian Mixture Model. This model helps to find relevant features by using Fuzzy ordered weighted average, which leads to determine the similarity of the density mixture. The salient feature of this approach is to find the relevant features simultaneously by employing fuzzy weighted approach. By applying Ordered Weighted Average (OWA), the feature weights are calculated and they are ordered using the membership values (oring criterion). Hence the feature weights are used as a regulator to determine the relevant features in feature reduction process. Maximum Ordered Weighted Average Likelihood (MOWAL) Framework adopts the Fuzzy Weighted – Gaussian Mixture Model (FW-GMM) for finding the component, which helps to discriminate the relevance of the features and improve the accuracy of density mixture.

**Keywords :** GMM, OWA, FW-GMM

## 1. INTRODUCTION

In feature reduction, a number of techniques are proposed to select the relevant features from the datasets. In this paper, the features are selected based on weight  $w_i \in [0,1]$  using Maximum Weighted Likelihood framework. It is one of the promising techniques to identify the features which are relevant. In review of literature, the common weight finding mechanism is not available for selecting the relevance of features in GMM. So the features are selected using Ordered Weighted Average (OWA) approach by applying fuzzy decision technique. Fuzzy decision making process is an efficient method compared with other estimation approaches under incomplete or uncertain information. The OWA based ordered weighting can be applied to weigh the features by using membership values. Hence, Gaussian Mixture Model also plays a vital role in the form clustering initialization and it is trained by the Expectation Maximization Technique. The process stops until the relative log-likelihood is obtained by applying the preset threshold. The density of component Mixture is improved by applying fuzzy weighted approach and thereby identifying relevant features and reducing the feature subset.

## 2. REVIEW OF LITERATURE

In feature reduction, various approaches are employed for reducing the features by using the filter, wrapper and hybrid approach. One of the approaches is called Semi Supervised which identifies the features. In this approach, information theory analysis was employed and symmetry cross entropy

distance measure was used to measure the difference of two random variables. The average symmetry cross entropy was used to measure the difference in degree of a multi-class problem [1]. Various methods were employed for the classification problem and seven feature selection techniques were used for evaluating imbalanced data sets. The receiver operating characteristic and area under the precision-recall curve metric can be used for finding the average performance for all classes. The likelihood metric was used to predict the performance of the model. The results revealed that very small number of features are selected using this approach for prediction [2]. The wrapper approach was applied to evaluate the feature subset. And the performance was good compared to the earlier approaches but, the searching strategy of finding the feature subset stops with local maxima [3,4]. The feature saliency was measured in the form of relevance by applying the unequal weights. The likelihood criterion optimized the feature reduction by using weights in Expectation Maximization algorithm [5]. The weights were introduced in the Maximum Weighted Likelihood framework, which yielded the component mixture and discriminated the features based on the relevance in the feature space [6]. The OWA optimization can be used into a mixed integer programming problem with monotonic weights, which were employed for higher dimensions [7].

## 3. MOTIVATION

### 3.1 Weighted likelihood Estimation (WLE)

The different weights are assigned to different samples, which are merged as relevant information into WLE function. The WLE for implication on  $\alpha$  is defined as,

$$WL(\alpha) = L_1(x_1, x_2, \dots, x_n; \alpha)^{\lambda_1} L_1(x_1, x_2, \dots, x_n; \alpha)^{\lambda_2} \quad (1)$$

where  $\lambda_1$  and  $\lambda_2$  are weights and they are used for finding the relevance of the likelihood. The non-negative weights are used for the experiment which is optimum.  $L_1(x_1, x_2, \dots, x_n; \alpha)$  can be used instead of  $L_2(x_1, x_2, \dots, x_n; \beta)$ , which defines  $WL(\alpha)$ . This value is used as a marginal distribution of X's and Y's. The WL value depends on the X's and it doesn't depend on Y's distributions. The value of X's and Y's are not independent and the weights of the likelihood function are modelled as the dependent one, which are not expressed in the marginals. Maximum Weighted Likelihood Estimator is obtained from likelihood function by maximizing the weights  $\lambda_1$  and  $\lambda_2$ .

The eq (1) can be used for the analysis to find out the WL estimation in terms of  $\alpha$ .

$$WL(\alpha) = \lambda_1 \ln L_1 (x_1, x_2, \dots, x_n; \alpha) + \lambda_2 \ln L_2 (y_1, y_2, \dots, y_n; \alpha) \quad (2)$$

$$\frac{\partial \ln WL(\alpha)}{\partial \alpha} = \frac{\lambda_1}{2\sigma_1^2} \sum_{i=1}^n t_i (x_i - \alpha t_i) + \frac{\lambda_2}{2\sigma_2^2} \sum_{i=1}^n t_i (y_i - \alpha t_i) \quad (3)$$

$$\text{Where } \alpha \text{ is } \bar{\alpha} = \frac{\lambda_1}{\lambda_1 + \lambda_2} \bar{\alpha} + \frac{\lambda_2}{\lambda_1 + \lambda_2} \bar{\beta}$$

Therefore the above equations are rewritten as  
 $\bar{\alpha} = \lambda_1 \bar{\alpha} + \lambda_2 \bar{\beta}$  Where  $\lambda_1 + \lambda_2 = 1$

### 3.2 Maximum Weighted Likelihood (MWL) Learning Framework

MWL estimation framework has been formulated with N observations. The following values  $x_1, x_2, \dots, x_N$ , are used for the forming the Mixture Model.

$$P(X_t | \theta^*) = \sum_{j=1}^{k^*} \alpha_j^* P(X_t | \theta_j^*) \quad (4)$$

$$\sum_{j=1}^{k^*} \alpha_j^* = 1 \text{ and } \forall j \leq k^* \alpha_j^* > 0$$

In eq (4),  $x_t$  ( $1 \leq t \leq N$ ) is a column vector of d-dimensional features, i.e  $x_t = [x_{1t}, \dots, x_{dt}]^T$ , and  $\theta^* = \{\alpha_j^*, \theta_j^*\}_{j=1}^{k^*}$ .  $\theta^*$  denotes the parameter set of the  $j^{\text{th}}$  probability density function  $p(x_t | \theta_j^*)$  in the Gaussian mixture model,  $k^*$  is called as cluster number,  $\alpha_j^*$  denoted as the mixer portion of  $j^{\text{th}}$  component.  $\theta^*$  is calculated by using N observations.

$$\hat{\theta} ml = \text{argmax}\{X_N | \theta\} \quad (5)$$

Hence,  $X_N = \{x_1, x_2, \dots, x_N\}$ , and  $\hat{\theta} ml = \{\alpha_j, \theta_j\}_j^k = 1$  is a maximum likelihood estimate of  $\theta^*$ . The number of components denoted by k and ML can be evaluated by EM algorithm. MWL framework finds the unequal weights and it's applied into conventional maximum likelihood

### 3.3 MOWAL Framework

In this frame work, OWA averages are used and features are ordered based on the origing criterion in the feature space. In this context, framework finds the ordered weights and it is applied into conventional maximum likelihood. This framework does not increase candidate model complexity due to the regularization approach. The weighted likelihood functions :

$$\begin{aligned} Q(\theta, X_N) &= \frac{1}{N} \sum_{i=1}^N \log P(X_i | \theta) = \frac{1}{N\zeta} \sum_{i=1}^N \sum_{j=1}^k g(j | X_i, \theta) \log P(X_i | \theta) \\ &= \frac{1}{N\zeta} \sum_{i=1}^N M(\theta, X_i) \end{aligned} \quad (6)$$

$$\begin{aligned} M(\theta, X_t) &= \sum_{j=1}^k g(j | X_t, \theta) \log[\alpha_j P(X_t | \theta)] \\ &\quad - \sum_{j=1}^k g(j | X_t, \theta) \log h(j | X_t, \theta) \end{aligned}$$

$$h(j | X_t, \theta) = \frac{\alpha_j P(X_t | \theta_j)}{P(X_t | \theta)} \quad (7)$$

where the posterior probability of  $j^{\text{th}}$  component holds the value of  $x_t$ , in the mixture, k value is an estimation of  $k^*$  with  $k \geq k^*$  and  $\zeta$  is a constant value. The  $g(j | x_t, \theta)$ 's are used in the weight functions and the following constraints are satisfied

$$\begin{aligned} \forall t, j, \sum_{j=1}^k g(j | X_t, \theta) &= \\ \zeta; \lim_{h(j | X_t, \theta) \rightarrow 0} g(j | X_t, \theta) \log h(j | X_t, \theta) &= 0 \end{aligned}$$

The weight functions are constructed using the following equations.

$$g(j | X_t, \theta) = (1 + \epsilon_t) I(j | X_t, \theta) - h(j | X_t, \theta) \quad (8)$$

with

$$I(j | X_t, \theta) = \begin{cases} 1 & \text{if } j = c \equiv \text{argmax}_{1 \leq i \leq k} h(i | X_t, \theta) \\ 0 & \text{otherwise,} \end{cases}$$

In the above equation,  $\epsilon_t$  is a small positive quantity. The weight can be calculated by using the  $x_t$  and  $g(c | x_t, \theta)$  is a positive weight assigned to the log-likelihood of the winning component. The maximum value  $h(j | x_t, \theta)$  is obtained in the component and it's updated to select the  $x_t$ . The remaining components are considered as rival components, that are penalized due to the negative weight. The maximum weighted likelihood is estimated by updating the details of  $\theta_{MWL} = \text{argmax}_{\theta} \{Q(\theta, X_N)\}$ .

### 3.4 Feature Relevance

The features are selected from the feature space based on relevance. The relevance of feature is identified by applying feature weights. If the feature is more relevant, the weight is large. John G. Kohavi introduces the three types of relevance such as irrelevant, weakly relevant and strongly relevant. The highly relevant features are selected from the feature space, which is used for the data mining process. A feature X is weakly relevant which is not useful for the data mining process [8].

### 3.5 Weighting Vs. Reduction

Fu proposed a feature reduction algorithm, which assigns binary weights to features and identifies the optimal features by applying continuous weighting to the features in the feature space. The speed and performance can be improved by employing the weights for the feature reduction process. Feature reduction algorithms can reduce the dimensionality through weighting. The irrelevant features are removed in the learning phase. [9,10,11].

### 3.6 Fuzzy sets

A fuzzy set F containing X is characterized by the membership function, where X is a nonempty set.

$$\mu_F : X \rightarrow [0, 1]$$

where  $\mu_A(x)$  is called as the membership degree of an element x in fuzzy set F for each  $x \in X$ .

$$F = \{(x, \mu_F(x)) | x \in X\}$$

Where A is a set of tuples. Then F(x) can be written as  $\mu_A(x)$ . The fuzzy (sub) sets in X is called as F(X).

If  $X = \{x_1, \dots, x_n\}$  is called as finite set and F is a fuzzy set in X and F can be rewritten as

$$F = \mu_1/x_1 + \dots + \mu_n/x_n$$

Where  $\mu_i/x_i$ ,  $i = 1, n$  denotes that  $\mu_i$  is a membership grade of  $x_i$  in F and plus sign is used as union of membership.

### 3.7 Ordered Weighted Average (OWA)

In fuzzy set, aggregating multiple numeric criteria technique is used to design overall objective function based on weighted sum. Hence the preference of weights can be introduced, which is called as Ordered Weighted Average (OWA). The OWA aggregations in which ordered the weights based on relevance. This OWA aggregation is differing from other weighting approach as maximum, minimum and average [11] weights can be calculated. The main feature of this approach recognizes the patterns and decides the patterns with its decision making capability, whether it's relevant or not. This operator helps us to recognize and decide the patterns from the maximum, arithmetic mean and minimum values. A predefined degree orness finds the optimistic to pessimistic value [12, 13, 14]. A maximum entropy approach is used to find the weight in a constrained non-linear optimization problem with a predefined degree of orness, which is an objective for the model [15, 16, 17]. A Minimum variance method is adopted for finding the minimum variability weights of the variables identified by this approach [18]. MinMax approach has been adopted for calculating the maximum difference of two adjacent weights by using Linear programming [19, 20, 21, 22,23].

The equation for assigning the weights is given below.

$$W(a_1, \dots, a_N) = \sum_{i=1}^N w_i a_{\sigma(i)}$$

where,  $\sigma$  corresponds to a permutation of  $a_i$ , which are ordered from the largest one to the lowest one. So,  $a_1$  will be the largest of the  $a_i$  and  $a_N$  will be the lowest of the  $a_i$ . In this definition,  $p_i$  corresponds to the weight of the  $i^{\text{th}}$  data after ordering them. In this way, weights are introduced and the degree of relevance of the data is expressed as low, high and central data.

The OWA operator has been mapped with n dimension  $F: \mathbb{R}^n \rightarrow \mathbb{R}$  that has been associated with n vector.

$$w = (w_1, w_2, \dots, w_n)^T$$

where,  $w_i \in [0,1], 1 \leq i \leq n$ .

$$\sum_{i=1}^n w_i = w_1 + \dots + w_n = 1$$

$$F(a_1, \dots, a_n) = \sum_{j=1}^n w_j b_j$$

where,  $b_j$  is the  $j^{\text{th}}$  largest element of the bag  $\langle a_1, \dots, a_n \rangle$

The Weighting function can be calculated by using the following cases:

Case 1: Finding the Maximum

$$w^* = (1, 0, \dots, 0)^T$$

$$\text{Max}(a_1, \dots, a_n) = \max \{ a_1, \dots, a_n \}$$

Case 2: Finding the Minimum

$$w^* = (0, 0, \dots, 1)^T$$

$$\text{Min}(a_1, \dots, a_n) = \min \{ a_1, \dots, a_n \}$$

Case 3: Finding the Average

$$w_A = (1/n, \dots, 1/n)^T$$

$$F(a_1, \dots, a_n) = \frac{a_1 + \dots + a_n}{n}$$

The significance of OWA operator is the re-ordering step, where the aggregate  $a_i$  is not related with the weight  $w_i$ . But weight is related with a particular ordered position of aggregate.

### 3.8 Rival Penalization Expectation Maximization (RPEM)

RPEM technique is a black box behind the Gaussian Mixture Model for feature reduction. In this process, the clustering process can be performed iteratively by learning. This technique scans all data points and finds a near optimal number of subsets. The proposed feature reduction approach yields the relevant and non-redundant features from the subset. Here the partitioning and feature reduction can be performed iteratively in several epochs.

#### Proposed Algorithm

Procedure Feature Reduction (F,  $\beta$ ,  $\gamma$ , T)

Input (F,  $\beta$ ,  $\gamma$ , T)

Output R

1.  $F : X \rightarrow [0, 1]$   
 $F = \mu_1/x_1 + \dots + \mu_n/x_n$
2. Calculate Weights:  $F_r \in F$   
 $W' = \frac{1}{n-1} \sum_{i=1}^n (n-i) * w_i$   
/\* Feature reduction and Filtering\*/  
 $R' \leftarrow F - (F_r | \text{Rank}_r < \beta, F_r \in F)$
3.  $R'' = [(U_{w_i} - L_{w_i}) + (M_{w_i} - L_{w_i})] / n + L_{w_i}$
4.  $R'' \leftarrow$  Projection of R
5.  $R \leftarrow R''$

The above algorithm has been formulated for finding the weights

$$g(j|X_t, \theta) = I(j|X_t, \theta) + h(j|X_t, \theta), j = 1 \dots k_{\text{max}}$$

In each epoch, the weights can be calculated as winning component  $I(j|X_t, \theta) = 1$ , where j is considered as winning component. This weighted approach has been adopted in RPEM for finding the relevant component mixture and discarding irrelevant, redundant components in this process. RPEM algorithm finds the relevance weights and mixture by applying the following equation.

$$\text{orness}(w) = \frac{1}{n-1} \sum_{i=1}^n (n-i) * w_i$$

The algorithm can be modelled as mathematical problem

$$\begin{aligned} \text{Maximize } \text{orness}(w) &= \frac{1}{n-1} \sum_{i=1}^n (n-i) * w_i \\ \text{Subject to } \text{owa} &= \sum_{i=1}^n w_i b_i = 0 \quad 0 \leq 0 \leq 1 \\ w_i &= w_1 + \dots + w_n = 1, 0 \leq w_i, i=1 \dots n \end{aligned}$$

Each mixture is weighted and ordered winning components are identified  $R^m \rightarrow R^m$  such that  $\Theta(\mathbf{y}) = (\theta_1(\mathbf{y}), \theta_2(\mathbf{y}), \dots, \theta_n(\mathbf{y}))$ , where  $\theta_1(\mathbf{y}) \geq \theta_2(\mathbf{y}) \geq \dots \geq \theta_n(\mathbf{y})$ . It exists in each epoch  $\tau$  of set  $I$  such that  $\theta_i(\mathbf{y}) = y \tau(i)$  for  $i = 1, \dots, n$ . In this feature reduction process, the smaller weights are discarded and highest weights are used for finding the relevant features from  $n$  observations.

#### 4. EXPERIMENTAL RESULTS

In this proposed approach,  $k_{\max}$  is set as  $1/N$ , and  $\beta_j$ 's and  $\gamma_i$ 's to 0, which is equivalent to setting each  $\alpha_j$  to  $1/N$  and  $w_1$  to 0.5. The remaining parameters are randomly initialized by the approach. The learning rates are  $\eta = 10^{-5}$ ,  $\eta\beta = 10^{-4}$ . The initial centers of the clusters ( $\mu_j$ 's) are randomly chosen from the data points.

**Experiment 1:** In wine dataset, 178 data points with three classes are forming the Gaussian mixture of three components and weights namely  $w_1 = 1$ ,  $w_2 = 0.7$  and  $w_3 = 0.6933$ . The feature weights of the three dimensions converge close to 1. The algorithm has accurately silhouetted the data points by using Fuzzy Weighed-Gaussian Mixture model structures in the first three dimensions, and component parameters are estimated.

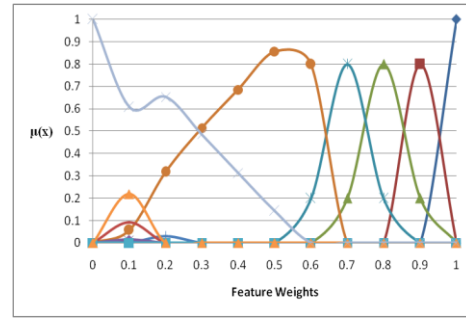
$$0.4 * N\left[\begin{pmatrix} 13 \\ 2 \end{pmatrix} \begin{pmatrix} -0.2 & 0 \\ 0 & -0.2 \end{pmatrix}\right] + 0.3 * N\left[\begin{pmatrix} 2 \\ 2 \end{pmatrix} \begin{pmatrix} 0.4 & 0 \\ 0 & 0.4 \end{pmatrix}\right] + 0.3 * N\left[\begin{pmatrix} 2 \\ 2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\right]$$

In this experiment, this approach selects the relevant features by reducing the feature space. In Table 1, the features are selected and it is weighted according to the membership values. In the first epoch, each data points are independent features from standard normal distribution.

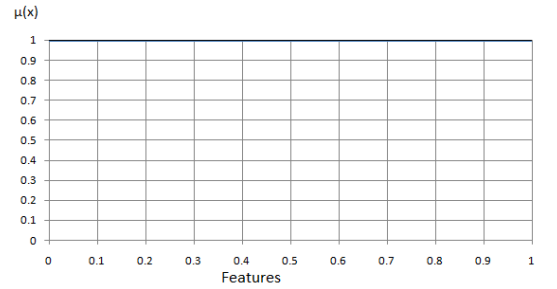
**Table1: Feature Selection based on Weights in Wine dataset**

Epoch	Selected features
1	F1,F2,F3,F4,F5,F6,F7,F8,F10,F12,F13
2	F1,F2,F3,F4,F5,F6,F13
3	F1,F2,F3,F4,F5,F6,F13

In the second epoch, the number of features are reduced due to the weights calculated by the algorithm depending on the relevance of the feature and irrelevant features are removed. Hence, the features are ranked in descending order and three components are selected. In three intermediate runs, the components mixture are formed and the features are reduced by fuzzy filtering.



**Fig 1.  $\mu(x)$  vs. Wine Feature Weights**



**Fig 2. Orness values of wine Features (Window type)**

Fig.1 shows the various weights of features against the membership values. There is a compensative connective with a high degree of satisfaction of oring criteria. Fig.2 shows that the oring criterion reveals full compensation. The orness ( $w$ ) is always denoted as a unit interval. It reveals that the nearer  $w$  is an oring criterion value, the nearer its value is to one. An OWA operator with top most weights ( $\text{orness}(w) \geq 0.5$ ) can be used for finding the good mixture model.

**Experiment 2:** In Ionosphere dataset, 351 data points with two classes are forming the Gaussian mixture of two components. The feature weights  $w_1$  is 1  $w_2$  is 0.828 and its value nearer to 1. In this approach, only the weight value nearer to 1 and a component has been added to improve the accuracy, namely orness ( $w$ )  $\leq 0.5$ . The algorithm clusters the data points by using Gaussian mixture model in the three dimensions and component parameters are estimated.

**Table 2: Feature Selection based on Weights in Ionosphere dataset**

Epoch	Selected features
1	F1,F18,F19,F22,F25,F34
2	F1,F2,F18,F20,F22, F24,F27,F34
3	F1,F2,F18,F19, F20,F21.F22,F25,F28,F32,F34
4	F1,F2,F6,F7,F12,F16,F18,F20,F22,F23,F26,F27,F30,F34
5	F1,F2,F8,F18,F20,F22,F26,F27,F30,F34
6	F1,F3,F18,F21,F22,F25,F29,F33,F34
7	F1,F2,F18,F19,F21,F32,F33,F34

In this filtering, six relevant features are selected by this approach in first epoch. In the second epoch, the number of features are increased as 8. In the third epoch, the eleven features are selected, in the fourth epoch, 14 features are selected. In the fifth epoch, the number of features is reduced to 10, In the sixth epoch, nine features are selected and in the seventh epoch eight features are selected by the algorithm. The above features are ranked according to the threshold value  $\geq 0.5$ . Only two components are selected and they are highly relevant to the class variables.

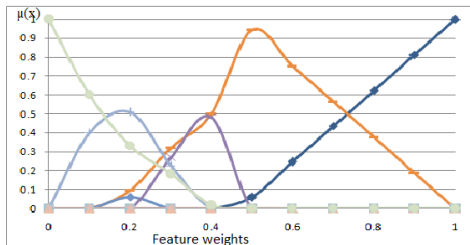


Fig 3.  $\mu(x)$  vs. Ionosphere feature Weights

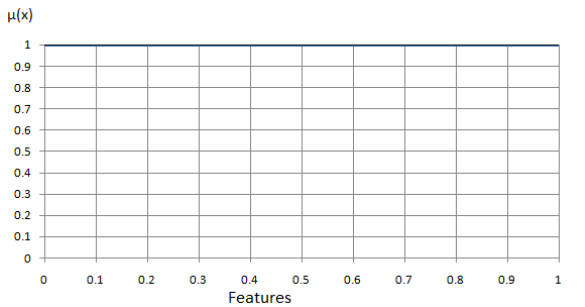


Fig 4. Orness values of Ionosphere Features (Window type)

Fig 3 and 4 show that, the oring criterion reveals the compensation connective of weights and its membership values, where as the weights nearer to 1 is selected for the component mixture. But the oring criterion nearer to the threshold value is also selected for the refinement of Gaussian mixture.

**Experiment 3:** In Wdbc dataset, 569 data points with two classes are used to form the Gaussian mixture of three components with weights such as  $w_1 = 0.7833$ ,  $w_2 = 0.6666$  and  $w_3 = 0.6333$ . The feature weights of the three dimensions converge close to 0.8. The weight value is nearer to 0.8 and the model has been built with Gaussian mixture. The algorithm clusters the various data points related to the centroid.

In this experiment, several features are selected which are relevant.. Other features are selected by this algorithm in eleven epochs. The relevant features are ranked according to the threshold value  $\geq 0.5$ . Only three components are selected and they are highly relevant to the class variables.

**Table 3 : Feature Selection based on Weights in Wdbc dataset**

Epoch	Selected features
1	F1,F8,F15,F18,F22,F23,F29,F31,F37,F38,F39,F43,F45,F47,F48,F50
2	F1,F21,F23,F28,F29,F33,F38,F42,F44,F47,F48
3	F1,F17,F18,F21,F29,F31,F32,F48
4	F1,F3,F13,F15,F18,F19,F20,F21,F22,F23,F40,F45

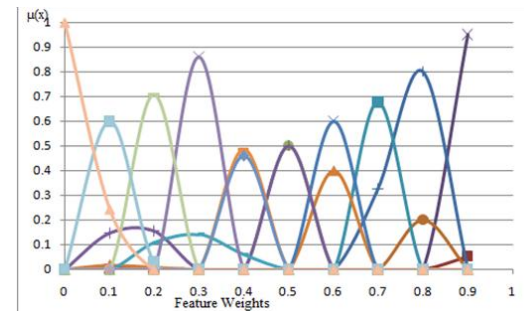


Fig 5.  $\mu(x)$  vs. wdbc feature Weights

Fig 5 shows that the oring criterion reveals the full compensation connective of weights and its membership values; whereas the weight nearer to a 0.8 is selected for the component mixture. But, the oring criterion above the criterion value is only selected for the mixture.

**Experiment 4:** In sonar dataset, 1000 data points with two classes are used to form Gaussian mixture of two components with weights such as  $w_1 = 1$ ,  $w_2 = 0.8352$ . The feature weights of the two of the component dimensions converge close to 1, and one more component has been added for the accuracy of the Gaussian mixture below the threshold. In this sonar dataset, several features are selected which are relevant in this experiment in four epochs. Only highly relevant features are ranked according to the threshold value  $\geq 0.5$ . The components are selected and they are highly relevant to the class variables. One more element is added with mixture which is less than the threshold value.

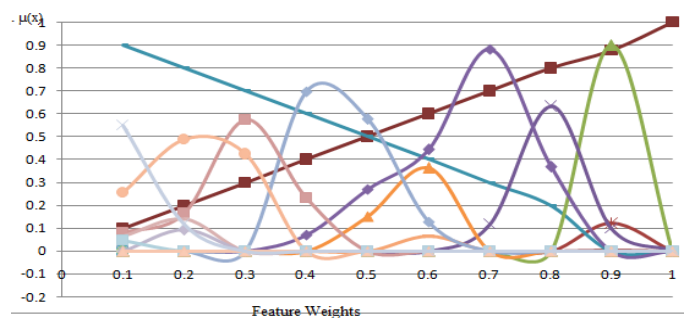


Fig 6.  $\mu(x)$  .vs. Sonar feature Weights

**Table 4: Feature Selection based on Weights in sonar dataset**

Epoch	Selected features
1	F2,F3,F4,F5,F7,F8,F9,F10,F11,F13,F14,F16,F18
2	F2,F3,F4,F5,F7,F8,F9,F10,F11,F12,F13,F14,F15,F16
3	F2, F4, F6, F7, F12, F13, F15, F16.
4	F6,F7,F10,F11,F12,F13,F14,F15,F16,F17,F19,F25,F29
5	F6,F7,F10,F11,F12,F13,F14,F15,F16,F17,F18,F19,F20,F25,F29,F30
6	F2,F4,F6,F7,F10,F11,F12,F13,F15,F16,F17,F18,F19,F20,F27,F29
7	F6,F7,F10,F11,F12,F13,F15,F1,F20,F25,F26,F29
8	F2,F3,F4,F5,F7,F8,F9,F10,F11,F14,F15,F16,F18
9	F6,F7,F11,F12,F13,F15,F17,F18,F19,F20,F25,F27,F29,F30
10	F4,F5,F7,F8,F9,F10,F11,F13,F14,F15,F16,F18
11	F1,F2,F3,F4,F5,F6,F7,F8,F9,F10,F11,F12,F13,F14,F15,F16,F17,F18

In this analysis, the performance is evaluated with real world datasets. The comparison has been made with existing algorithm and the relevance of the feature has been identified by fuzzy weighting. Here, the different weights are calculated by using orness criterion. The weights are ordered and selected based on the threshold value. The accuracy of the proposed approach is evaluated in the form error rate. The proposed approach has been evaluated by dividing the dataset as training and test set. The mean and standard deviation are measures for finding the accuracy of the cluster similarity. This experiment reveals that the error rate is reduced by using proposed the approach and the cluster assignments of data points are different from the class labels in the dataset

## 5. CONCLUSION

The proposed feature reduction approach determines the most relevant features by using fuzzy weight. The orness criterion measure is to inspect the effectiveness of the feature weight by setting the value  $f(w) = w$ ,  $w \in [0,1]$ , which is ordered and the lower level weights are discarded. The advantage of using fuzzy weighted Gaussian mixture is that the features are having strong relevance to form the density mixture. This new approach estimates the parameters and selects the highly weighted features for the density mixture. It reveals that the selected features' relevance is strong and accuracy of the model is high.

## ACKNOWLEDGEMENT

I owe my sincere gratitude to the Dr. Aros Research and Education (DARE) Group, Tiruchirappalli, Tamil Nadu, India for their constructive criticisms and productive suggestions to write this paper.

**Table 5 Accuracy of the test sets for each algorithm**

Dataset	Method	Model Order (Mean ±Std)	Error Rate (Mean ±Std)
wdbc d=30 N=569 K*=2	RPEM	1.7±0.4	0.2610±0.0781
	GMClusFW	5.7±0.3	0.1005±0.0349
	IRFS-RPEM	2.3±0.4	1.021±0.0546
	IRRFs-RPEM FW-RPEM	Fixed at 2 Fixed at 2	0.0897±0.0308 <b>0.0776±0.0268</b>
Sonar d=60 N=1000 K*=2	RPEM	2.3±0.8	0.4651±0.0532
	GMClusFW	1.0±0.0	0.5000±0.0000
	IRFS-RPEM	2.8±0.6	3.625±0.0394
	IRRFs-RPEM FW-RPEM	2.7±0.7 <b>2.6±0.6</b>	0.3221±0.0333 <b>0.3120±0.0320</b>
Wine d=13 N=178 K*=3	RPEM	2.5±0.7	0.0843±0.0261
	GMClusFW	3.3±1.4	0.0673±0.0286
	IRFS-RPEM	4.7±1.7	0.0492±0.0182
	IRRFs-RPEM FW-RPEM	3.1±0.5 <b>2.9±0.4</b>	0.0509±0.0248 <b>0.0424±0.0234</b>
Ionosphere d=32 N=351 K*=2	RPEM	1.8±0.5	0.4056±0.0121
	GMClusFW	3.2±0.6	0.2268±0.0386
	IRFS-RPEM	2.0±0.8	0.2921±0.0453
	IRRFs-RPEM FW-RPEM	2.5±0.5 <b>2.4±0.5</b>	0.2121±0.0273 <b>0.2100±0.0260</b>

## 6. REFERENCES

- [1] Ianisse Quinzán, José M. Sotoca, Filiberto Pla, "Clustering-based Feature Selection in Semi-supervised Problems", Ninth International Conference on Intelligent Systems Design and Applications, 2009.
- [2] Ghaderi, M.A.; Yazdani, N.; Moshiri, B.; Tayefeh Mahmoudi, M., "A new approach for text feature selection based on OWA operator", 5th International Symposium on Telecommunications (IST), IEEE, 2011.
- [3] Liu, H., Yu, L., Toward integrating feature selection algorithms for classification and clustering. IEEE Transactions on Knowledge and Data Engineering 17, 491–502, 2005.
- [4] Dash, M.C., Scheuermann, K., Liu, P.H.: Feature selection for clustering-A filter solution. In: Proceedings of 2nd IEEE International Conference on Data Mining, pp. 115–122. IEEE Computer Society Press, Los Alamitos, 2002
- [5] Dy, J., Brodley, C.: Feature selection for unsupervised learning. Journal of Machine Learning Research 5, 845–889, 2004.
- [6] Figueiredo, M.A.T., Jain, A.K., Law, M.H.C.: A feature selection wrapper for mixtures. LNCS, vol. 1642, pp. 229–237. Springer, Heidelberg, 2003.
- [7] Yager, R.R., Constrained OWA aggregation. Fuzzy Sets Syst. 81, 89–101, 1996.
- [8] Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Trans. Systems, Man and Cyber. 18, 183–190, 1988.
- [9] Fu .K.S, Sequential methods in pattern recognition and machine learning. New York: Academic Press, 1968 .
- [10] Yager, R.R., Kacprzyk, J.: The Ordered Weighted Averaging Operators: Theory and Applications. Kluwer AP, Dordrecht, 1997.

- [11] Yager, R.R., Kacprzyk, J.: *The Ordered Weighted Averaging Operators: Theory and Applications*. Kluwer AP, Dordrecht ,1997.
- [12] Ogryczak, W.: Inequality measures and equitable approaches to location problems. *Eur. J. Opnl. Res.* 122, 374–391 , 2000.
- [13] Fuller, R.: On obtaining OWA operator weights: a short survey of recent developments. In: *Proc. 5th IEEE Int. Conf. Comput. Cybernetics, Gammarth, Tunisia*,pp. 241–244 , 2007.
- [14] Xu, Z.: An overview of methods for determining OWA weights. *Int. J. Intell.Syst.* 20, 843–865, 2005.
- [15] Ogryczak, W.: Inequality measures and equitable approaches to location problems. *Eur. J. Opnl. Res.* 122, 374–391 ,2000 .
- [16] Xu, Z.: An overview of methods for determining OWA weights. *Int. J. Intell.Syst.* 20, 843–865,2005.
- [17] Fuller, R., Majlender, P.: An analytic approach for obtaining maximal entropy OWA operator weights. *Fuzzy Sets and Systems* 124, 53–57 , 2001.
- [18] Fuller, R., Majlender, P.: On obtaining minimal variability OWA operator weights. *Fuzzy Sets and Systems* 136, 203–215 , 2003.
- [19] Amin, G.R., Emrouznejad, A.: An extended minimax disparity to determine the OWA operator weights. *Comput. Ind. Eng.* 50, 312–316, 2006.
- [20] Wang, Y.-M., Luo, Y., Liu, X.: Two new models for determining OWA operator weights. *Comput. Ind. Eng.* 52, 203–209, 2007.
- [21] Liu, X.: The solution equivalence of min-max disparity and minimum variance problems for OWA operators. *Int. J. Approx. Reasoning* 45, 68–81, 2007.