



Pairwise Fuzzy Ordered Weighted Average Algorithm-Gaussian Mixture Model for Feature Reduction

Charles.S, Arockiam.L

Assistant Professor, Department of Computer Science, St. Joseph's College, Tiruchirappalli
cs_bd@yahoo.com

Associate Professor, Department of Computer Science, St. Joseph's College, Tiruchirappalli
larockiam@yahoo.co.in

ABSTRACT

Feature reduction is a kind of dimensionality reduction of feature space. There are a number of approaches are used to identify the significant features but they are not using the weighting approach. The weighing approach is quite useful for obtaining the significant features and removing the insignificant and irrelevant features using OWA formulation. The aim of this approach is to obtain the significant features and removing insignificant features by using the pairwise approach. This approach is helpful to find the weights of pairwise features at the same time, which leads to remove the insignificant features from the feature space using OWA. The significance of the OWA formulation is that, the paired features are identified in priori and their sum of weights are equal to 1. OWA criterion is introduced to obtain the significant features that are useful for predicting the accuracy of the cluster in GMM.

Indexing terms

Gaussian Mixture Model, OWA, Pairwise Feature Reduction.

Academic Discipline and Sub-Disciplines

Computer Science

SUBJECT CLASSIFICATION

Data MINING, Feature Reduction

Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 6, No 1

editor@cirworld.com

www.cirworld.com, member.cirworld.com



1. Introduction

Feature reduction is a challenging problem of finding the significant features in data mining analysis. There are a number of selection techniques existing in feature selection, extraction and reduction. In real world datasets, the insignificant and irrelevant features exist, which are not useful for data mining process. This problem is addressed using the pairwise feature selection method, which finds the significant features in feature space. Model based clustering is one of the approaches for finding the significant features in a “one-in-all-out” manner. The paired features are selected, if at least one pair of clusters is separable by this feature. The insignificant and irrelevant features are removed, if it does not separate any of the clusters. Pairwise Mixture Model is able to take into account for finding the significant features in the model based clustering. In this approach, the new pairwise penalty is employed to find the interdependence between states and between observations.

This penalty based method is used to penalize the difference between all pairs of cluster centers for each feature and reduce the centroids of non-separable clusters. The cluster centroids associated with the observations and other insignificant information are removed from the model. In this context, the feature reduction is employed in the pairwise penalty approach for finding the significant features using Gaussian Mixture Model by removing insignificant features from the feature space. Model based clustering helps to form the Gaussian mixtures of different shapes of clusters and it helps to improve the accuracy of clustering.

2. Motivation

The feature selection in clustering is one of the issues, also known as subspace clustering. Friedman et.al [1] proposed a technique for finding the subset of features using hierarchical clustering, which uncovers the cluster structures. Tadesse et al. [2] introduced a Bayesian term for feature reduction that searches for models, which finds different clusters and subsets of features.

Rafter et al. [3] also employed a regularization approach for identifying the relevant features and removing the irrelevant and insignificant features of the model, which were included or excluded from the model.

Pan et al. [4] proposed a Gaussian Mixture Model to impose a penalty on the cluster means. The means of all clusters were summed up to zero, where the method eradicates the insignificant features in clusters and their means shrunk to zero. In this approach the Bayesian Information Criterion (BIC) was used to identify the insignificant features from the feature space. The drawback of this approach was that, BIC identified a number of non zero components of maximum likelihood and some of the insignificant features were also clustered as non zero component. This Model is less accurate due to the insignificant features in winning component mixture.

Wang et al. [5] introduced two methods to find the insignificant features from the feature space such as Adaptive L_∞ norm Penalized Gaussian Mixture Model (ALP-GMM) and Adaptive Hierarchically Penalized Gaussian Mixture Model (AHP-GMM). In this study, if the feature is significant for clustering but its weight was small that feature was lightly penalized. Whereas the feature was insignificant for clustering and its weight was large, hence the feature was heavily penalized. Some insignificant features were also selected by GMM, which affects the accuracy of the model.

Jian Guo et.al. [6] proposed a new feature reduction method for obtaining the significant features from the feature space. The pairwise fusion penalty criterion was introduced to differentiate the pairs of the clusters for each feature and shrunk the centroids of non separable clusters. All the clusters were associated with features that are fused, that feature was insignificant and removed from GMM. This method did not identify the insignificant features in priori. So, this approach was not quite useful for identifying the significant features in priori and accuracy of the model was less.

Sen et al. [7] proposed a convex based method to penalize the pairwise L_∞ norm regression/classification coefficients for obtaining the pairwise features in the feature space by simultaneous feature selection. In this study, the analysis was made with synthetic and realworld datasets, and some of the significant features were not obtained from the feature space in convex approach.

3. Preliminaries

3.1 Pairwise Mixture Model

Let $y = \{y_1, y_2, \dots, y_n\}$ is a set of N observed data ($y_n \in \mathbb{R}$), $x = \{x_1, x_2, \dots, x_n\}$ is the classification of y -data into a finite set classes ($\Omega = \{1, 2, \dots, k\}$). In the probabilistic classic model, the data y_n are the realization of mutually independent random variables (Y_n) with the same mixture distribution.

$$fmm(y_n) = \sum_{k=1}^K \pi_k f_k(y_n) \quad (1)$$

Each $f_k(\cdot) = p(\cdot | k)$ is a class k conditional probability density function and priors $\pi_k = p(x_n = k)$ are such that $\sum_{k=1}^K \pi_k = 1$. In this model, the observed data are distributed into k classes; it is called as mixture components.

Two sets of N observed data point $y^1 = \{y_n^1\}_{n \in [1, N]}$ and $y^2 = \{y_n^2\}_{n \in [1, N]}$ and corresponding classification are $x^1 = \{x_n^1\}_{n \in [1, N]}$ and $x^2 = \{x_n^2\}_{n \in [1, N]}$ Modelled into K and L classes. $\forall n \in [1, N]$, $x_n^1 \in \Omega^1 = \{1, 2, \dots, K\}$ and $x_n^2 \in \Omega^2 = \{1, 2, \dots, L\}$, $y_n = (y_n^1, y_n^2)$ and $x_n = (x_n^1, x_n^2)$. The pairwise data y_n are the realizations of mutually independent vectors Y_n .



$$fmm(y_n) = \sum_{i=1}^K \sum_{j=1}^L \gamma_{k,l} f_{k,l}(y_n^1, y_n^2) \tag{2}$$

Where $\gamma_{k,l}^{(\ell-1)} = p(x_n^1 = k, x_n^2 = l)$ is a joint prior with $\sum_{i=1}^K \sum_{j=1}^L \gamma_{k,l} f_{k,l} = 1$. The sets of Pairwise Model Mixture parameters are used in this model $\theta = \{\theta_{k,l}\}_{(kl) \in \Omega^1 \times \Omega^2}$ with $\theta_{k,l} = (\gamma_{k,l}, \mu_{k,l}^{(1)}, \sigma_{k,l}^{(1)}, \mu_{k,l}^{(2)}, \sigma_{k,l}^{(2)}, \rho_{k,l})$. Here, y_n is used to estimate one classification x_n only. It is the product of two independent Mixture Models defined by

$$fimm(y_n) = \sum_{i=1}^K \pi_k^{(1)} f_k^{(1)}(y_n^1) \sum_{j=1}^L \pi_l^{(2)} f_l^{(2)}(y_n^2) \tag{3}$$

3.2 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 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.[8,10,11]

4. PFOWA-GMM algorithm

The algorithm for the proposed PFOWA is given below.

Procedure Feature Reduction (F,β,γ,T)

Step 1 : $F : X \rightarrow [0, 1]$

$$F = \mu_1/x_1 + \mu_2/x_2 \dots + \mu_n/x_n$$

Step 2: Calculate Weights: $F_r \in F$

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

Step 3: Feature reduction and Filtering

$$R' \leftarrow F - (F_r | \text{Rank}_r < \beta, F_r \in F)$$

$$R'' = [(U_{w_i} - L_{w_i}) + (M_{w_i} - L_{w_i})] / n + L_{w_i}$$

Step 4: Initialization of parameters at $\ell=0$

Step 5: Compute Estimation

for $\ell=1$ to ℓ do

5.1 Compute posteriori probabilities

$$\xi_n^{(\ell-1)}(k, l) = \frac{\gamma_{k,l}^{(\ell-1)} f_{k,l}^{(\ell-1)}(y_n^1, y_n^2)}{\sum_{i=1}^K \sum_{j=1}^L \gamma_{k,l}^{(\ell-1)} f_{k,l}^{(\ell-1)}(y_n^1, y_n^2)}$$

5.2 Compute priori probabilities

$$\gamma_{k,j}^{(\ell-1)} = \frac{1}{N} \sum_{n=1}^N \xi_n^{(\ell-1)}(k, l)$$

5.3 Compute Data Driven Parameters

$$\mu_{k,j}^{(\ell)} = \frac{\sum_{n=1}^N \xi_n^{(\ell-1)}(k, l) y_n}{\sum_{n=1}^N \xi_n^{(\ell-1)}(k, l)}$$

$$\Gamma_{k,j}^{(\ell)} = \frac{\sum_{n=1}^N \xi_n^{(\ell-1)}(k, l) (y_n - \mu_{k,l}^{(\ell)}) (y_n - \mu_{k,j}^{(\ell)})}{\sum_{n=1}^N \xi_n^{(\ell-1)}(k, l)}$$

Step 6: Classification from (θ^ℓ)

Step 7: $R'' \leftarrow$ Projection of R

Step 8: $R \leftarrow R''$

5 Experimental Results

5.1. Analysis with Wine Dataset

Wine dataset contains the 178 data points and 13 features are used in this experiment. The dataset is divided into training and test set for checking the accuracy of the selected features. Orness criterion is used to select the pairwise features and the selected pairs are utilized for forming the Gaussian Mixture.

Table 1. Wine Model parameters used to draw mixtures

(k,l)	$\gamma_{k,l}$	$\mu_{k,l}$	$\sigma_{k,l}$	$\mu_{k,l}$	$\sigma_{k,l}$	$\rho_{k,l}$
(F6, F13)	1.91	2.11	0.63	630.55	314.91	0.21
(F5 F6)	97.62	97.91	14.28	2.11	0.63	0.50
(F3 F5)	1.12	2.33	0.28	97.91	14.28	0.08
(F2 F3)	0.05	1.91	1.12	2.33	0.28	0.16
(F10, F13)	229.47	4.11	2.32	630.55	314.91	0.32
(F10, F12)	-0.70	4.11	2.32	2.39	0.71	-0.43

Table 1 shows the various pairwise mixture values of wine dataset. Here, pairwise features are selected based on orness criterion. The weighted sum of pairwise features is equal to 1. The selected mean vectors are $\mu_{6,13} = (2.11, 630.55)$, $\mu_{5,6} = (97.91, 2.11)$, $\mu_{3,5} = (2.33, 97.91)$, $\mu_{10,13} = (4.11, 630.55)$ and covariance matrixes $\Gamma_{6,13} = \begin{pmatrix} 1.91 & 0 \\ 0 & 1.91 \end{pmatrix}$, $\Gamma_{5,6} = \begin{pmatrix} 97.62 & 0 \\ 0 & 97.62 \end{pmatrix}$, $\Gamma_{3,5} = \begin{pmatrix} 1.21 & 0 \\ 0 & 1.21 \end{pmatrix}$, $\Gamma_{10,13} = \begin{pmatrix} 229.47 & 0 \\ 0 & 229.47 \end{pmatrix}$ which corresponds to correlation $\rho = 0.21$, $\rho = 0.50$, $\rho = 0.08$, $\rho = 0.32$. These values are used to form the pairwise mixtures of features of wine dataset. The pairwise features are identified in prior by using OWA formulation. The selected features are highly significant and relevant for grouping the components in Gaussian Mixture clustering. Each component is validated by using data driven parameters and their values are tabulated in table 1

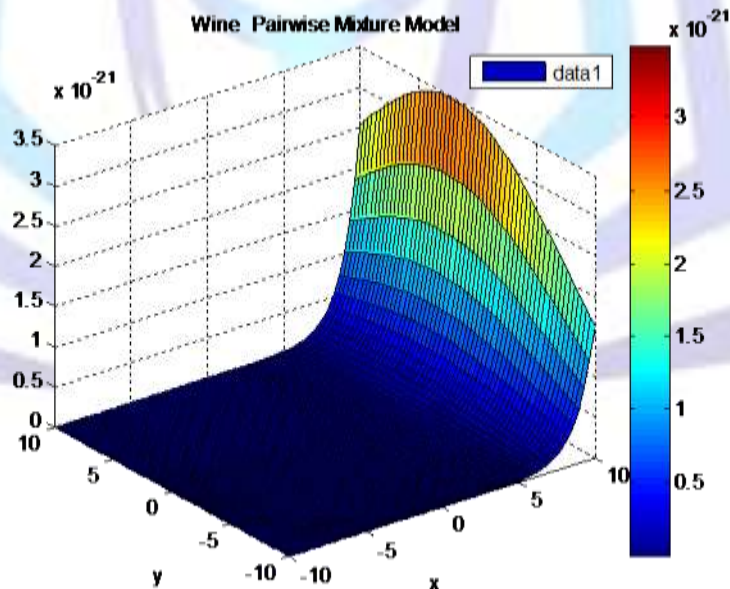


Figure 1 Wine-Pairwise Mixture Model with Parameters

Figure 1 pictorially represented the wine pairwise mixture model; it is plotted in 3 dimensional area. For Instance, the four pairwise features are selected, which are based on pairwise mean values such as $\mu_{6,13} = (2.11, 630.55)$, $\mu_{5,6} = (97.91, 2.11)$, $\mu_{3,5} = (2.33, 97.91)$, and $\mu_{10,13} = (4.11, 630.55)$. The blue color denotes the data points related to the pairwise features selected by the algorithm. Color bar denotes the posterior probability of component mixture and its density ranging from 0 to 3.5. Here, data matrix X (4-by-3), where 4 is the number of observations and 3 is the dimension of the data. X-axis denotes the data matrix of wine dataset and Y-axis denotes vector-y containing the values of the probability density function.

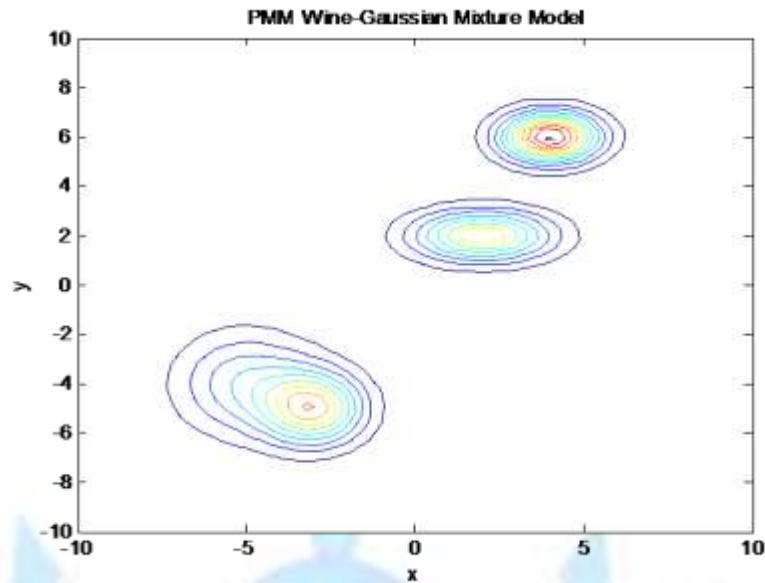


Figure 2 Pairwise Mixture Model - Wine Gaussian Mixture Component

The highly significant and relevant features are selected from the wine feature space, which are pictorially represented as components (2-dimensional space x and y) in Figure 2. The cluster centers are formed based on the pairwise features selected by the algorithm. RPEM accurately cluster the data points by using Gaussian Mixture Model structures and component parameters are estimated.

Table 2 Accuracy of Test Sets for Each Algorithms

Dataset	Method	Model Order (Mean ±Std)	Error Rate (Mean ±Std)
Wine d=13 N=178 K*=3	IRRFs-RPEM	3.1±0.5	0.0509±0.0248
	FOWA[8]	2.9±0.4	0.0424±0.0234
	FWOWA[9]	2.6±0.25	0.0418±0.0220
	PFOWA	2.7±0.32	0.0421±0.0320

Table 2 shows the results of wine dataset based on the model and sampling. The accuracy of PFOWA is compared with FWOWA, FOWA and IRRFS-RPEM, and its model and sampling error rate index values are shown in figure 3 and 4.

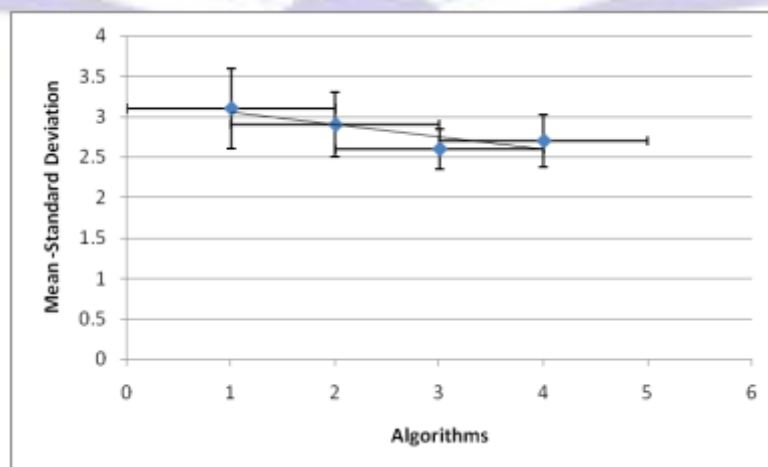


Figure 3. Accuracy of Model (Wine)

Figure 3 shows the results of accuracy of the wine GMM model. Here, x-axis denotes the algorithms such as 1-IRRFS-RPEM, 2-FOWA, 3-FWOWA and 4-PFOWA and y-axis denotes mean-standard deviation values of the model. Each algorithm modelled is based on different techniques and values graphically present in the form error rate index. When pairwise approach (2.7 ± 0.32) is compared with FWOWA (2.6 ± 0.25) accuracy value is less and the accuracy value is high when compared with FOWA (2.9 ± 0.4). The PFOWA model clustering accuracy is high compared with FOWA model; it is due to the pairwise clustering of wine dataset.

Figure 4 pictorially represents the accuracy of the samples in the form of the error rate index. The following algorithms are listed as 1-IRRFS-RPEM, 2-FOWA, 3-FWOWA and 4-PFOWA. The x-axis and y-axis contain the error rate index value i.e. Mean and standard deviation value. PFOWA accuracy (0.0421 ± 0.0320) value is high compared with FOWA (0.0424 ± 0.0234) and its accuracy is less compared with FWOWA (0.0418 ± 0.0220). It is evident that the PFOWA algorithm performs better compared with FOWA. It's due to the nature of joint prior probability of the model.

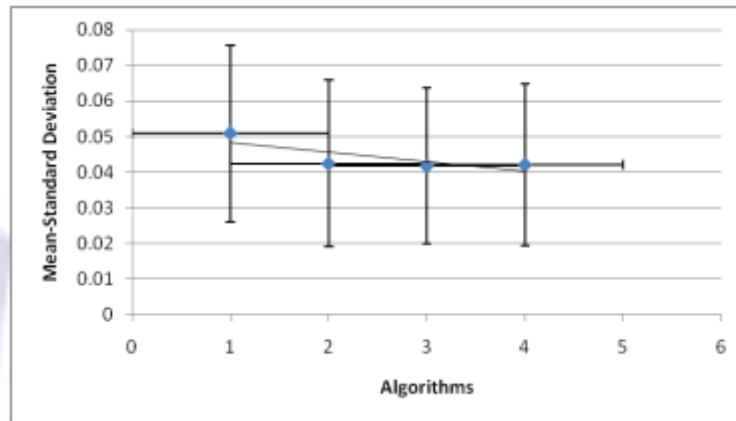


Figure 4. Accuracy of Sample (Wine).

5.2. Analysis with Ionosphere Dataset

In this experiment, ionosphere dataset which has 351 data points with 32 features are used for finding the pairwise feature reduct. Each subset is obtained by using orness criterion and selected pairwise features for constructing the GMM Model. The selected subsets are evaluated, and their accuracy is tested by RPEM algorithm.

Table 3. Ionosphere model parameters used to draw mixtures

(k,l)	$\gamma_{k,l}$	$\mu_{k,l}$	$\sigma_{k,l}$	$\mu_{k,l}$	$\sigma_{k,l}$	$\rho_{k,l}$
(F25, F34)	0.03	0.40	0.58	0.01	0.47	0.11
(F1 F18)	0.01	0.89	0.31	0.00	0.50	0.08
(F22 F28)	0.12	0.01	0.52	-0.07	0.55	0.41
(F25 F18)	-0.05	0.00	0.50	0.40	0.58	-0.18
(F25 F22)	-0.02	0.01	0.52	0.40	0.58	-0.08
(F34, F27)	0.02	0.54	0.52	0.01	0.47	0.08
(F34, F30)	0.09	-0.03	0.51	0.01	0.47	0.39
(F8, F18)	0.02	0.12	0.52	0.00	0.50	0.07
(F27, F22)	-0.01	0.01	0.52	0.54	0.52	-0.03
(F3, F18)	0.04	0.64	0.50	0.00	0.50	0.17
(F18, F22)	0.09	0.00	0.50	0.01	0.52	0.36
(F18, F19)	0.01	0.00	0.50	0.36	0.63	0.02

The pairwise mixture ionosphere values are tabulated in table 3. Each mixture orness criterion value is summed as 1. These mixtures are highly significant and relevant features are selected and insignificant features are removed by using OWA formulation with Joint probability. The mean vectors are $\mu_{25,34} = (0.4, 0.1)$, $\mu_{22,28} = (0.01, -0.07)$, $\mu_{34,30} = (-0.03, 0.01)$, $\mu_{34,27} = (0.54, 0.01)$ and covariance matrixes $\Gamma_{25,34} = \begin{pmatrix} 0.03 & 0 \\ 0 & 0.03 \end{pmatrix}$, $\Gamma_{22,28} = \begin{pmatrix} 0.12 & 0 \\ 0 & 0.12 \end{pmatrix}$, $\Gamma_{34,30} = \begin{pmatrix} 0.09 & 0 \\ 0 & 0.09 \end{pmatrix}$, $\Gamma_{34,27} = \begin{pmatrix} 0.02 & 0 \\ 0 & 0.02 \end{pmatrix}$ which corresponds to correlation $\rho = 0.11$, $\rho = 0.41$, $\rho = 0.39$, $\rho = 0.08$. The selected pairwise mixtures are used to form the Gaussian mixture which is shown in Figure 5.6. The data points are highly significant and relevant, which are obtained from the Ionosphere feature space.

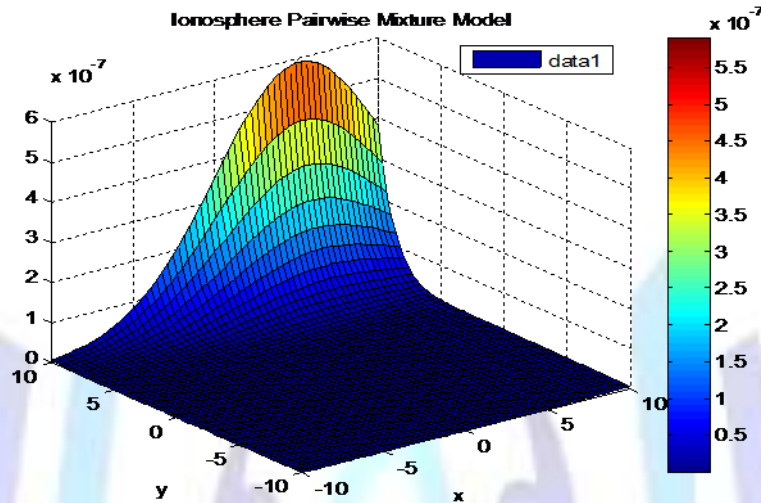


Figure 5 Ionosphere-Pairwise Mixture Model with Parameters

Figure 5 shows the ionosphere pairwise mixture model. For Instance, the pairwise features are selected, which are based on pairwise mean values such as $\mu_{25,34} = (0.4, 0.1)$, $\mu_{22,28} = (0.01, -0.07)$, $\mu_{34,30} = (-0.03, 0.01)$, and $\mu_{34,27} = (0.54, 0.01)$. The data points denoted as blue color are used to form pairwise mixtures. The posterior probability of component mixture is represented in a color bar; density of component mixture is ranging from 0 to 6. Here, X-axis denotes the data matrix of ionosphere dataset and Y-axis denotes containing the values of the probability density function values.

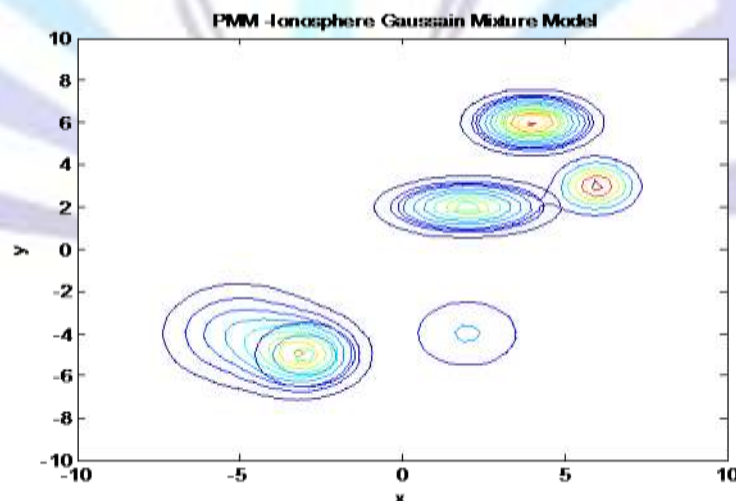


Figure 6. Gaussian mixture Distribution of Ionosphere Dataset

The highly significant and relevant features are pictorially represented in the form component mixture (2-dimensional space x and y) in Figure 6. The cluster structures are formed using significant and relevant features obtained from the ionosphere feature space. Various data points are clustered as Gaussian Mixture Model structures and component parameters are estimated.



The selected pairs are test sets and they are evaluated for the accuracy using RPEM. The ionosphere test results are listed in the table 4. The pairwise features are selected and obtained features form the Gaussian Mixture. The model order and sampling values reveal the accuracy of the PFOWA, which is less than compared with FWOVA and greater than FOWA

Table 4. Accuracy of Ionosphere Test Set for Each Algorithms

Dataset	Method	Model Order (Mean ±Std)	Error Rate (Mean ±Std)
ionosphere d=32 N=351 K*=2	IRRFs-RPEM	2.5±0.5	0.2121±0.0273
	FOWA[8]	2.4±0.5	0.2100±0.0260
	FWOWA[9]	2.3±0.4	0.2095±0.0245
	PFOWA	2.375±0.475	0.2087±0.0250

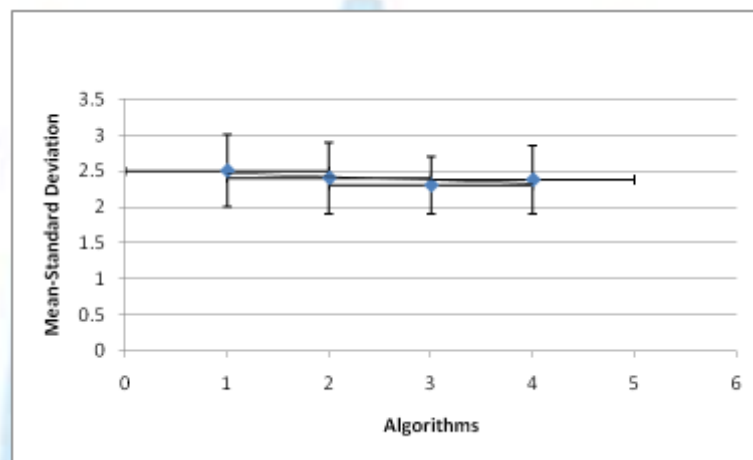


Figure 7. Accuracy of Model (Ionosphere)

The mean and standard deviation value of four different algorithms are plotted (y-axis) in the Figure 7. This figure reveals that the accuracy of the PFOWA (4) model is compared with other algorithms IRRFS-RPEM(1), FOWA(2) and FWOVA(3) plotted in x-axis, The FWOVA accuracy is the best compared with PFOWA. FOWA model is less accurate compared with PFOWA.

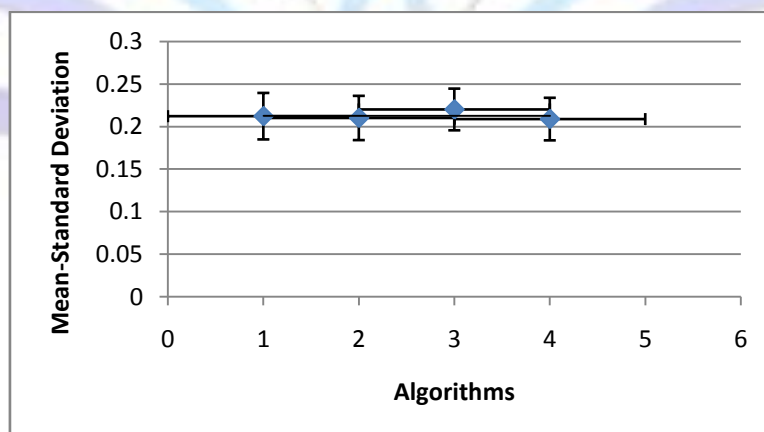


Figure 8. Accuracy of Sampling (Ionosphere)

Figure 8 reveals the accuracy of the ionosphere sampling. The error rate index value is less in sample FWOVA (3) compared with PFOVA (4), FOWA(2) accuracy is the best compared with IRRFS-RPEM(1), which plotted in x-axis and y-axis denotes the mean-standard deviation value.

5.3. Analysis with Wdbc dataset

In Wdbc dataset, 569 data points with two classes are used in this experiment. Pairwise model and sampling are formed to build the Gaussian mixture. The clusters are formed based on the pairwise features using orness criterion.

Table 5 Wdbc Model Parameters used to draw the mixture

(k,l)	$\gamma_{k,l}$	$\mu_{k,l}$	$\sigma_{k,l}$	$\mu_{k,l}$	$\sigma_{k,l}$	$\rho_{k,l}$
(F2,F4)	85.30	14.13	3.52	91.97	24.30	1.00
(F3,F5)	485.14	19.29	4.30	654.89	351.91	0.32
(F6,F7)	0.04	0.10	0.01	0.10	0.05	0.66
(F7,F11)	0.06	0.10	0.05	0.06	0.01	0.57
(F12,F13)	0.03	0.41	0.28	1.22	0.55	0.21
(F15,F16)	0.01	40.34	45.49	0.01	0.00	0.08
(F18,F19)	0.12	0.03	0.03	0.01	0.01	0.77
(F17,F22)	0.02	0.03	0.02	16.27	4.83	0.20
(F27,F29)	0.01	0.25	0.16	0.11	0.07	0.80
(F29,F30)	0.00	0.11	0.07	0.29	0.06	0.50

The pairwise features are obtained from the experiment and it is tabulated in table 5. The orness criterion value is a measure to find the pairwise mixture and their weighted sum is 1. Highly significant and relevant features are obtained and insignificant features are eradicated by using OVA formulation with Joint probability. The mean vectors are $\mu_{2,4} = (14.13, 91.97)$, $\mu_{6,7} = (0.1, 0.1)$, $\mu_{7,11} = (0.1, 0.06)$, $\mu_{18,19} = (0.03, 0.01)$, $\mu_{27,29} = (0.25, 0.11)$ and covariance matrixes $\Gamma_{2,4} = \begin{pmatrix} 85.30 & 0 \\ 0 & 85.30 \end{pmatrix}$, $\Gamma_{6,7} = \begin{pmatrix} 0.04 & 0 \\ 0 & 0.04 \end{pmatrix}$, $\Gamma_{7,11} = \begin{pmatrix} 0.06 & 0 \\ 0 & 0.06 \end{pmatrix}$, $\Gamma_{18,19} = \begin{pmatrix} 0.12 & 0 \\ 0 & 0.12 \end{pmatrix}$, $\Gamma_{27,29} = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$ Which corresponds to correlation $\rho = 1$, $\rho = 0.66$, $\rho = 0.57$, $\rho = 0.77$, $\rho = 0.8$.

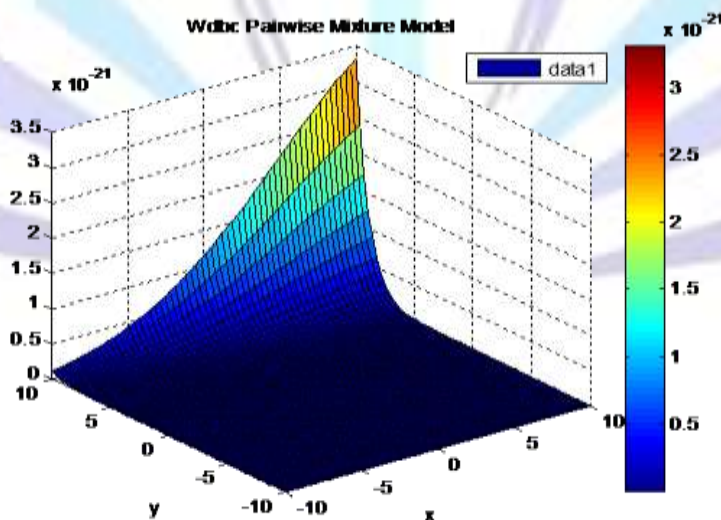


Figure 9 Wdbc-Pairwise Mixture Model with Parameters

The pairwise mixtures are used to form the Gaussian mixture which is shown in Figure 9. The data points are highly significant and relevant such as $\mu_{2,4} = (14.13, 91.97)$, $\mu_{6,7} = (0.1, 0.1)$, $\mu_{7,11} = (0.1, 0.06)$, $\mu_{18,19} = (0.03, 0.01)$, $\mu_{27,29} = (0.25, 0.11)$, which are obtained from the Wdbc feature space. The blue color is denoted as the data points., ie. pairwise features are used to form the winning component mixture. The Color bar denotes that the posteriori probability

value lies between 0 to 3.5 (Density of component mixture). X-axis denotes the various data points and y-axis denotes the probability density function values.

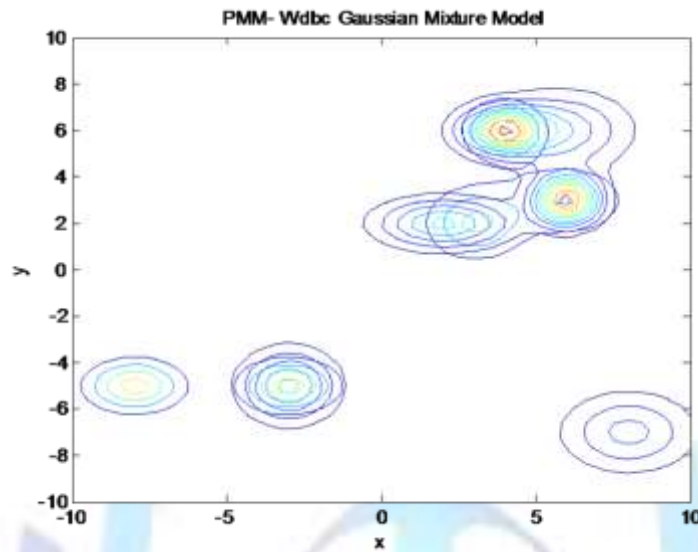


Figure 10. Gaussian Mixture Distribution of Wdbc Dataset

Wdbc dataset Gaussian Component Mixture Distribution is presented in Figure 10. Various cluster structures are formed using significant and relevant features from the feature space using orness criterion and posteriori probability. The significant and relevant features are sampled and insignificant features are removed from the Wdbc feature space. The component parameters are estimated and values are evaluated using RPEM.

Table 6 Accuracy of wdbc test set for each algorithm

Dataset	Method	Model Order (Mean ±Std)	Error Rate (Mean ±Std)
wdbc	IRRFs-RPEM	Fixed at 2	0.0897±0.0308
d=30	FOWA[8]	1.97±0.2	0.0776±0.0268
N=569	FWOWA[9]	1.85±0.3	0.0625±0.0250
K*=2	PFOWA	1.90±0.3	0.0770±0.0325

Table 6 shows the various accuracy values of different algorithms such as IRRFS-RPEM, FOWA, FWOWA and PFOWA. The PFOWA algorithm accuracy is less compared with FWOWA and the accuracy is high compared with FOWA due to the pairwise approach.

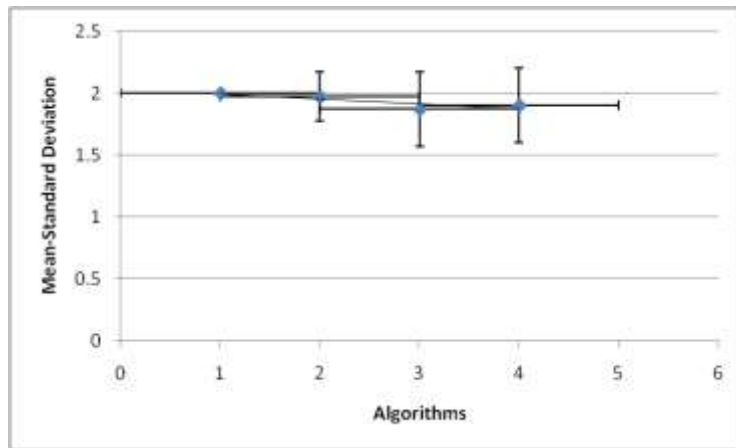


Figure 11. Accuracy of Model (Wdbc)

Figure 11 reveals the accuracy of the PFOWA (4) model compared with other algorithms IRRFS-RPEM(1), FOWA(2) and FWOWA(3) plotted in x-axis. FOWA model accuracy is less compared with PFOWA. FWOWA model accuracy is the best compared with PFOWA.

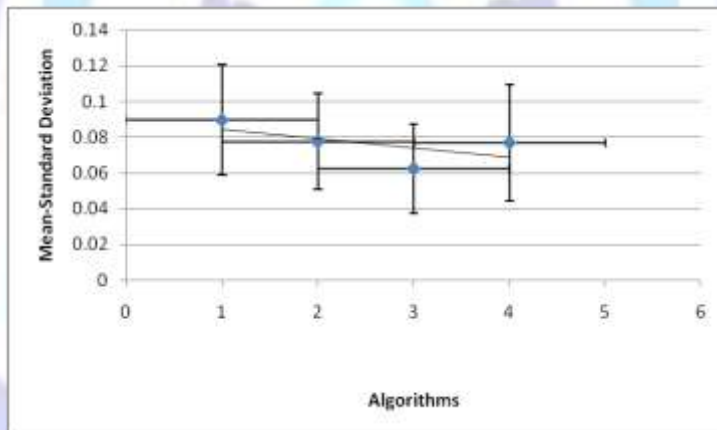


Figure 12. Accuracy of Sample (Wdbc)

The accuracy values of the wdbc sample with four different algorithms are pictorially represented in figure 12. X-axis denotes the various algorithms used in this experiment such as IRRFS-RPEM(1), FOWA(2), FWOWA(3) and PFOWA. Y-axis denotes the mean and standard deviation values of the algorithms. The results reveal that the accuracy of the PFOWA performs better than FOWA.

5.4. Analysis with Sonar Dataset

In Sonar dataset, 1000 data points with two classes are used in this experiment. The pairwise features are obtained from the feature space, which are significant and relevant features. These feature pairs are tested and estimated by RPEM.

Table 7 shows the various feature pairs selected by the algorithm and their weighted values are summed as 1. The mean vectors are $\mu_{2,3} = (0.04, 0.04)$, $\mu_{4,5} = (0.05, 0.08)$, $\mu_{7,8} = (0.12, 0.13)$, $\mu_{9,10} = (0.18, 0.21)$, $\mu_{11,13} = (0.24, 0.27)$, $\mu_{14,16} = (0.30, 0.38)$, $\mu_{16,18} = (0.38, 0.45)$ and covariance matrixes $\Gamma_{2,3} = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$, $\Gamma_{4,5} = \begin{pmatrix} 0.02 & 0 \\ 0 & 0.02 \end{pmatrix}$, $\Gamma_{7,8} = \begin{pmatrix} 0.02 & 0 \\ 0 & 0.02 \end{pmatrix}$, $\Gamma_{9,10} = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$, $\Gamma_{11,13} = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$, $\Gamma_{14,16} = \begin{pmatrix} 0.03 & 0 \\ 0 & 0.03 \end{pmatrix}$, $\Gamma_{16,18} = \begin{pmatrix} 0.05 & 0 \\ 0 & 0.05 \end{pmatrix}$ Which corresponds to correlation $\rho = 0.78$, $\rho = 0.73$, $\rho = 0.68$, $\rho = 0.88$, $\rho = 0.65$, $\rho = 0.75$, $\rho = 0.77$. Highly relevant features are selected by this algorithm and only the winning components are selected to form the Gaussian mixture.

Table 7 Sonar Model Parameters used to draw the mixture

(k,l)	$\gamma_{k,l}$	$\mu_{k,l}$	$\sigma_{k,l}$	$\mu_{k,l}$	$\sigma_{k,l}$	$\rho_{k,l}$
(F2,F3)	0.01	0.04	0.03	0.04	0.04	0.78
(F4,F5)	0.02	0.05	0.05	0.08	0.06	0.73
(F7,F8)	0.02	0.12	0.06	0.13	0.09	0.68
(F9,F10)	0.01	0.18	0.12	0.21	0.13	0.88
(F11,F13)	0.01	0.24	0.13	0.27	0.14	0.65
(F14,F16)	0.03	0.30	0.16	0.38	0.23	0.75
(F16,F18)	0.05	0.38	0.23	0.45	0.26	0.77

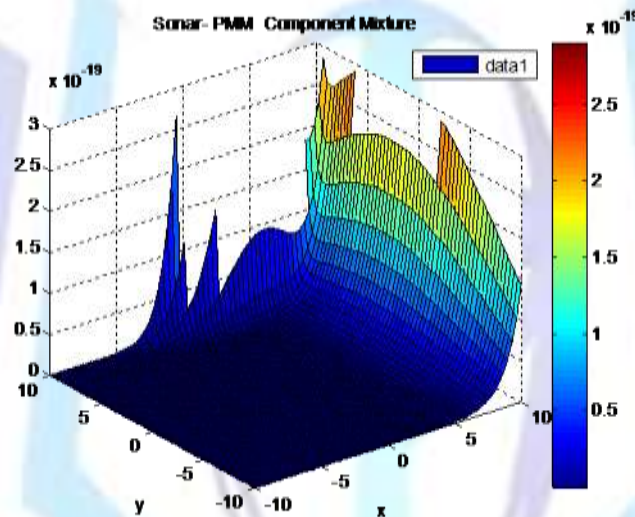


Figure 13 Sonar-Pairwise Mixture Model with Parameters

The Sonar pairwise mixture model is shown in the figure 13. The various pairwise features are selected for building the Gaussian Mixture Model. In this model, the blue color denotes the selected data points for the pairwise mixture in the winning component. The mixture components contain the highly significant and relevant feature pairs which are selected using the penalty. Posteriori probability is represented as the color bar in the Figure 13. The density of mixture value lies between 0 to 3.5.

In Figure 14, the highly weighted significant and relevant features are pictorially represented in the cluster structures. The x-axis in the graph denotes the features and the y-axis denotes the membership criterion values. Various cluster structures are formed based on pairwise feature selected by the RPEM . The significant and relevant features are sampled from feature space. The data points are clustered as Gaussian Mixture Model structures and parameters are estimated.

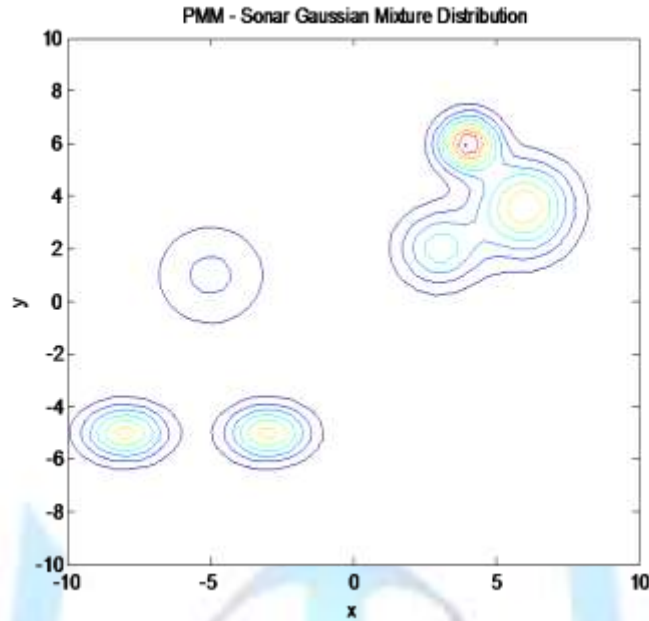


Figure 14. Gaussian Mixture Distribution of Sonar Dataset

Table 8 Accuracy of Sonar test set for each algorithms

Dataset	Method	Model Order (Mean ±Std)	Error Rate (Mean ±Std)
Sonar d=60 N=1000 K*=2	IRRFs-RPEM	2.7±0.7	0.3221±0.0333
	FOWA[8]	2.6±0.6	0.3120±0.0320
	FWOWA[9]	2.4±0.55	0.3110±0.0313
	PFOWA	2.5±0.35	0.3112±0.0308

In Table 8, the various algorithms are listed, which are used to test the sonar data set and their values are recorded in model and sampling error rate index. Each algorithm has its own specification and working principle is different. The PFOWA error rate index value is high compared with FWOWA in this experiment and error rate index value is less compared with FOWA and IRRFS-RPEM

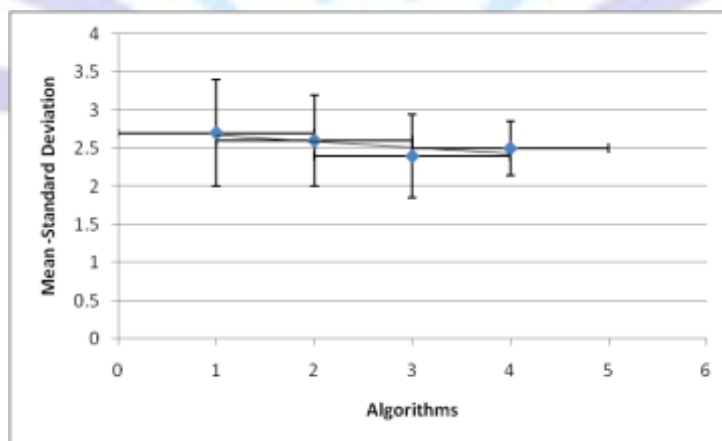


Figure 15. Accuracy of Model (Sonar) Mean–Standard Deviation

Figure 15 diagrammatically presents the various algorithms plotted on the x-axis and their mean and standard deviation values are plotted on the y-axis. PFOWA algorithm modeling accuracy is less compared with FWOWA.

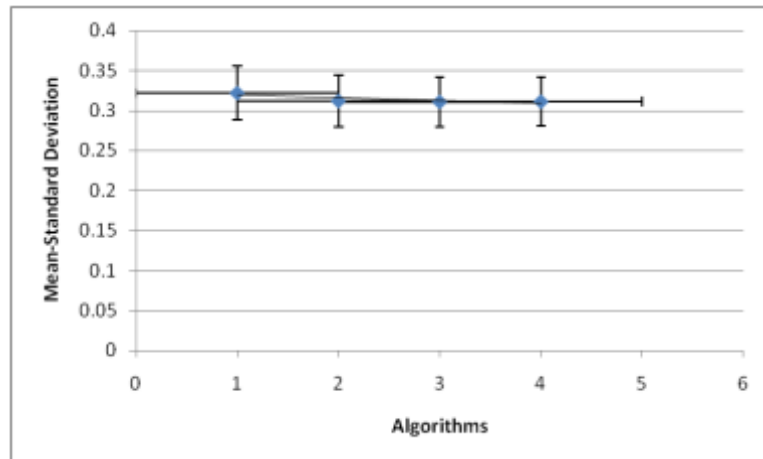


Figure 16. Accuracy of Sample (Sonar) Mean-Standard Deviation

Figure 16 reveals the accuracy of the sample of sonar dataset used for this experiment analysis. Various algorithms test values are plotted in x-axis and mean - standard deviation values are plotted. PFOWA accuracy is less compared with FWOWA, but FOWA accuracy is less compared with PFOWA.

6 Findings

Table 9 shows the proportions of the selected features by IRRFS-RPEM and FOWA in the real world dataset.

Table 9. Proportions of the Average Selected Features

Dataset	FWOWA (%)	PFOWA (%)
Wine	57.45	58.25
Ionosphere	30.65	31.75
Wdbc	46.12	47.25
Sonar	51.26	52.35

The relevant features are selected and their clustering accuracy is less compared with FWOWA. In this context, Wine, ionosphere, Wdbc and Sonar datasets, the accuracy get slightly decreased when compared to FWOWA. PFOWA finds the significant and relevant features in the dataset and its predictability power is high, when compared to FOWA. PFOWA predictive accuracy is less with a mean of 1.03% when compared to FWOWA.

9. Conclusion

Feature reduction techniques reduce the representation of dataset that is much smaller in volume in the feature space. The objective of this study is achieved through Pair wise Fuzzy Ordered Weighted Average approach, which identifies the significant and relevant features from the features space by removing the insignificant features. The strength of this approach, OWA operator and penalty method is combined for obtaining the most significant and relevant features in the real world datasets.

The low weighted features are identified by the algorithm and penalty is applied to enhance the accuracy of the cluster accuracy. In this penalty based method, the priori probability and posterior probability are incorporated for penalizing the low weighted features in the feature reduction process. PFOWA algorithm using RPEM is also added advantage, to estimate the parameter values and evaluate the features for constructing the cluster structures and component mixture in this approach. The analysis reveals that, there is an improvement in the accuracy by employing the penalty and OWA compared with FOWA. This experiment divulges that PFOWA efficiency is 1.03% less compared with FWOWA on real world datasets.



References

- [1] Friedman, J. and Meulman, J. (2004). Clustering objects on subsets of attributes . *Journal of the Royal Statistical Society, Series B* 66, 815–849.
- [2] Raftery, A. and Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association* 101, 168–178.
- [3] Raftery, A. and Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association* 101, 168–178.
- [4] Pan, W. and Shen, X. (2006). Penalized model-based clustering with application to variable selection. *Journal of Machine Learning Research* 8, 1145–1164.
- [5] Wang, S. and Zhu, J. (2007). Variable selection for model-based high dimensional clustering and its application to microarray data. *Biometrics* 64, 440–448.
- [6] Jian Guo, Elizaveta Levina, George Michailidis, and Ji Zhu, “Pairwise Variable Selection for High-Dimensional Model-Based Clustering”, *Biometrics* 66, 793–804,2010.
- [7] Sen Yang , Lei Yuan , Ying-Cheng Lai, Xiaotong Shen, Peter Wonka, Jieping Ye, “ Feature Grouping and Selection Over an Undirected Graph”, *KDD'12, ACM 978-1-4503-1462-6 /12/08*.
- [8] Charles.S Dr. L. Arockiam, “ Fuzzy Weighted Gaussian Mixture Model for Feature Reduction “, *International Journal of Computer Applications (0975 -8887) Volume 64 –No.18, February 2013*.
- [9] Charles.S Dr. L. Arockiam,” Fuzzy Weighted Ordered Weighted Average-Gaussian Mixture Model for Feature Reduction, *International Journal of Computers & Technology, Volume 4 No. 2,I, 2013*
- [10] Yager, R.R., Constrained OWA aggregation. *Fuzzy Sets Syst.* 81, 89–101 ,1996.
- [11] Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans. Systems, Man and Cyber.* 18, 183–190 ,1988