

Modeling Methods in Clustering Analysis for Time Series Data

Naglaa A. Morad

Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Cairo, Egypt

Email: Naglaa_Morad@yahoo.com

How to cite this paper: Morad, N.A. (2020) Modeling Methods in Clustering Analysis for Time Series Data. *Open Journal of Statistics*, 10, 565-580.

<https://doi.org/10.4236/ojs.2020.103034>

Received: May 15, 2020

Accepted: June 27, 2020

Published: June 30, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper is concerned about studying modeling-based methods in cluster analysis to classify data elements into clusters and thus dealing with time series in view of this classification to choose the appropriate mixed model. The mixture-model cluster analysis technique under different covariance structures of the component densities is presented. This model is used to capture the compactness, orientation, shape, and the volume of component clusters in one expert system to handle Gaussian high dimensional heterogeneous data set. To achieve flexibility in currently practiced cluster analysis techniques. The Expectation-Maximization (EM) algorithm is considered to estimate the parameter of the covariance matrix. To judge the goodness of the models, some criteria are used. These criteria are for the covariance matrix produced by the simulation. These models have not been tackled in previous studies. The results showed the superiority criterion ICOMP PEU to other criteria. This is in addition to the success of the model based on Gaussian clusters in the prediction by using covariance matrices used in this study. The study also found the possibility of determining the optimal number of clusters by choosing the number of clusters corresponding to lower values for the different criteria used in the study.

Keywords

Gaussian Mixture Model-Based Clustering (GMMC), The Expectation-Maximization (EM) Algorithm, AIC, SBC, ICOMP PEU

1. Introduction

The clustering analysis is one of the statistical methods that deal with the division and classification of variables data elements into several homogeneous groups that are homogeneous within one group (cluster) and are different from

other groups (other clusters). Cluster analysis is defined as, a set of methods for constructing a (hopefully) sensible and informative classification of an initially unclassified set of data, using the variable values observed on each individual. All such methods essentially try to imitate what the eye-brain system does so well in two dimensions (Everitt and Skrondal [1]). Because of this characteristic of cluster analysis, it has been used in many applied fields. It is used to divide and classify data into aggregates, which help to properly select appropriate statistical analysis of these data as a decision-making tool. The objective of this statistical method is to divide the data matrix containing the number of (n) of the samples and (p) of the variables into a homogeneous number of partial groups (k) by assembling homogeneous and convergent sample items in clusters. Thereafter, criteria and measures must be used to distinguish between the different cluster results to reach two main points: the similarity of the data elements within the different clusters and the optimal number of clusters. This is done through the use of the legal functions, known as the standards of validity and legal performance of the cluster. In this paper, one of the most important hybrid models based on clusters, the Gaussian mixed model-based clustering is used. The hybrid models based on clusters are able to predict accurately if the appropriate variance model is chosen. It is applied through the use of four heterogeneity models. The covariance matrix of the Gaussian mixed model is unknown. So to estimate these parameters we need to maximize the log-likelihood function of. Direct maximization of the log-likelihood function is complicated, so the maximum likelihood estimator (MLE) of a finite mixture model is usually obtained via the EM algorithm (Dempster *et al.* [2]).

Banfield and Raftery [3] proposed a model-based clustering method based on constraining these geometric features of components using the eigenvalue decomposition of the covariance matrix.

Different constraints on the covariance matrix provides different models that are applicable to different data structures, which is another advantage of model-based clustering. In 1995, Celeux and Govaert [4] classified these models in three main families of models: spherical, diagonal and general families. They have given the definitions and derivations of all 14 available models, along with the covariance matrix update equations based on these models to be used in the EM algorithm. However, only nine of those have a closed form solution to the covariance update equation, which is evaluated in the M-step of the EM algorithm.

Later in 2016, Chi *et al.*, [5] showed that the population likelihood function has bad local maxima even in the special case of equally-weighted mixtures of well-separated and spherical Gaussians. They proved that the log-likelihood value of these bad local maxima can be arbitrarily worse than that of any global optimum. Also, they showed that the EM algorithm with random initialization will converge to bad critical points with probability at least. They further establish that the first-order variant of EM will not converge to strict saddle points almost

surely, indicating that the poor performance of the first-order method can be attributed to the existence of bad local maxima rather than bad saddle points.

Cluster analysis is used in various fields of science. Tóth *et al.*, [6] described gamma-ray bursts (GRBs) using clustering. They analyzed the Final BATSE Catalog using Gaussian-mixture-models-based clustering methods for six variables (durations, peak flux, total fluency and spectral hardness ratios) that contain information on clustering.

In 2000, Bozdogan [7] studied the basic idea of Akaike’s [8] information criterion (AIC). Then, he presented some recent developments on a new entropic or information complexity (ICOMP) criterion of Bozdogan [9] for model selection.

The main contribution of the present paper is to propose the mixture-model cluster analysis technique under different covariance structures of the component densities. To determine the optimal number of clusters by selecting the number of clusters corresponding to the lowest values for the different criteria. Four models for covariance structures that have not been applied in previous studies are studied using three criteria of the complexity of information.

This paper is organized as follows: Section one is the introduction and section two the Gaussian Mixture Model-based Clustering (GMMC) is discussed. In section three, the Expectation-Maximization (EM) algorithm is introduced. The Model Selection Criteria are introduced in section four. Finally, sections five and six contain the Numerical Results, and the Conclusion, respectively (Table 1).

2. The Gaussian Mixture Model-Based Clustering (GMMC)

The Gaussian mixture model is a powerful clustering algorithm used in cluster analysis. It is the most widely used clustering method of this kind, is the one based on learning a mixture of Gaussians. It assumes that there are a certain number of Gaussian distributions, and each of these distributions represents a cluster. Hence, a Gaussian Mixture Model tends to group the data points belonging to a single distribution together. Gaussian Mixture Models are probabilistic models and use the soft clustering approach for distributing the points in different clusters. It’s difficult to determine the right model parameters, Expectation-Maximization method is used to determine the model parameters.

In a case where $X \in \mathbb{R}^{(n \times p)}$ are given (p dimensional data of size n), would be interested in estimating the number of clusters K . Assuming the observations x_{ij} ($i = 1, \dots, n, j = 1, \dots, p$) are assumed to be drawn from the following mixture K distribution, each corresponding to a different cluster:

Table 1. Nomenclatures of used parameters.

parameters	nomenclatures	parameters	nomenclatures
π_k	mixing proportion	λ_k	Scalar controlling the volume of the ellipsoid
θ_k	vector of unknown parameters	A_k	diagonal matrix
Σ_k	covariance matrix	D_k	orthogonal matrix
μ_k	mean vector		

$$f(x; \pi, \theta) = \sum_{k=1}^K \pi_k g_k(x; \theta_k)$$

Here π_1, \dots, π_K are the mixing proportions that satisfy $\pi_k > 0$ and $\sum_{k=1}^K \pi_k = 1$. θ_k is the vector of unknown parameters of the k th component, and π_k represents the probability that an observation belongs to the k th component. The Gaussian mixture model assumes that the components of the mixture are the multivariate normal distribution, thus the density function becomes:

$$f(x; \pi, \mu, \Sigma) = \sum_{k=1}^K \pi_k g_k(x; \mu_k, \Sigma_k)$$

The mixture components (*i.e.* clusters) are ellipsoids centered at μ_k with other geometric features, such as volume, shape, and orientation, determined by the covariance matrix Σ_k . (Titterington *et al.* [10]).

In this case, the component densities g_k are given by:

$$g_k(x; \mu_k, \Sigma_k) = (2\pi)^{-\frac{p}{2}} |\Sigma_k|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x - \mu_k) \Sigma_k^{-1} (x - \mu_k)\right\}$$

Parsimonious parameterizations of the covariance matrices can be obtained by using the eigenvalue decomposition of the covariance matrix. The eigenvalue decomposition of the k th covariance matrix is given as:

$$\Sigma_k = \lambda_k D_k A_k D_k^T > k$$

where: λ_k is a scalar controlling the volume of the ellipsoid.

A_k is a diagonal matrix specifying the shape of the density contours with $\det(A_k) = 1$.

D_k is an orthogonal matrix which determines the orientation of the corresponding ellipsoid (Banfield and Raftery [3] and Celeux and Govaert [4]).

In one dimension, there are just two models: *E* for equal variance and *V* for varying variance. In the multivariate setting, the volume, shape, and orientation of the covariance can be constrained to be equal or variable across groups. Thus, 14 possible models with different geometric characteristics can be specified. **Table 2** reports all such models with the corresponding distribution structure type, volume, shape, orientation, and associated model names. See (Erar [11], Gupta and Bhatia [12], Chi *et al.*, [5], Scrucca *et al.*, [13], Malsiner-Walli *et al.*, [14] and Tóth, *et al.*, [6]).

Approaching the clustering problem from this probabilistic standpoint reduces the whole problem to the parameter estimation of a mixture density. The unknown parameters of the Gaussian mixture density, are the mixing proportions, π_k , the mean vectors, μ_k , and the covariance matrices, Σ_k . Therefore, to estimate these parameters, we need to maximize the log-likelihood given by:

$$\log L(\theta | x) = \sum_{i=1}^n \log \left[\sum_{k=1}^K \pi_k g_k(x_i | \mu_k, \Sigma_k) \right]$$

The estimates of the mixing proportion, π_k , the mean vector μ_k , and the covariance matrix Σ_k for the k th population are given as:

Table 2. Parameterizations of the covariance matrix and the corresponding geometric features.

	Model	Covariance	Distribution	Volume	Shape	Orientation
1	EII	λI	Spherical	Equal	Equal	-
2	VII	$\lambda_k I$	Spherical	Variable	Equal	-
3	EEI	λA	Diagonal	Equal	Equal	Coordinate axes
4	VEI	$\lambda_k A$	Diagonal	Variable	Equal	Coordinate axes
5	EVI	λA_k	Diagonal	Equal	Variable	Coordinate axes
6	VVI	$\lambda_k A_k$	Diagonal	Variable	Variable	Coordinate axes
7	EEE	λDAD^T	Ellipsoidal	Equal	Equal	Equal
8	EVE	$\lambda D A_k D^T$	Ellipsoidal	Equal	Variable	Equal
9	VEE	$\lambda_k DAD^T$	Ellipsoidal	Variable	Equal	Equal
10	VVE	$\lambda_k D A_k D^T$	Ellipsoidal	Variable	Variable	Equal
11	EEV	$\lambda D_k A D_k^T$	Ellipsoidal	Equal	Equal	Variable
12	VEV	$\lambda_k D_k A D_k^T$	Ellipsoidal	Variable	Equal	Variable
13	EVV	$\lambda D_k A_k D_k^T$	Ellipsoidal	Equal	Variable	Variable
14	VVV	$\lambda_k D_k A_k D_k^T$	Ellipsoidal	Variable	Variable	Variable

$$\hat{\pi}_k = \frac{1}{n} \sum_{i=1}^n I_k(\hat{\mathcal{Y}}_i)$$

$$\hat{\mu}_k = \frac{1}{\hat{\pi}_k n} \sum_{i=1}^n \chi_i I_k(\hat{\mathcal{Y}}_i)$$

$$\hat{\Sigma}_k = \frac{1}{\hat{\pi}_k n} \sum_{i=1}^n \left[(\chi_i - \hat{\mu}_k)' (\chi_i - \hat{\mu}_k) \right] I_k(\hat{\mathcal{Y}}_i)$$

where: $I_k(\hat{\mathcal{Y}}_i) = \begin{cases} 1 & \hat{\mathcal{Y}}_i = k \\ 0 & \hat{\mathcal{Y}}_i \neq k \end{cases}$.

This estimation requires the non-linear optimization of the mixture likelihood for high-dimensional data sets. However, there are no closed-form solutions to $\frac{\partial}{\partial \theta} \log L(\hat{\theta} | x) = 0$ for any mixture density; so the likelihood has to be numerically maximized. For this numerical optimization, the Expectation-Maximization (EM) algorithm of Dempster *et al.* [2] is used, which treats the data as incomplete and the group labels y_i as missing.

3. The Expectation-Maximization (EM) Algorithm

The expectation-maximization (EM) algorithm is an iterative procedure used to find maximum likelihood estimates when data are incomplete or are treated as being incomplete. The consummate citation for the EM algorithm is the famous paper by Dempster *et al.* [2]. In EM algorithm, E and M steps are iterated until convergence is reached. The EM algorithm is based on the “complete-data”; *i.e.*,

the observed data plus the missing data. In E-step, the expected value of the complete-data log-likelihood, say Q , is computed; in the M-step, Q is maximized with respect to the model parameters. The EM algorithm is easy to implement and a numerically stable algorithm that has reliable global convergence under fairly general conditions. However, the likelihood surface in mixture models tends to have multiple modes. So initialization of EM is crucial because it usually produces sensible results when started from reasonable starting values (Wu [15]). In this approach, hierarchical clusters are obtained by recursively merging the two clusters that provide the smallest decrease in the classification likelihood for the Gaussian mixture model (Banfield and Raftery [3], Xu *et al.* [16]).

The EM algorithm is an iterative procedure consisting of two alternating steps, given some starting values for all parameters ($\hat{\pi}_k$, $\hat{\mu}_k$ and $\hat{\Sigma}_k$). The algorithm can be summarized as follows at iteration $(t + 1)$:

1) In the E-step, the posterior probability, \hat{T}_{ik} of the i th observation belonging to the k th component is estimated, given the current parameter estimates.

$$\hat{T}_{ik} = \frac{\hat{\pi}_k^{(t)} g_k(x_i | \hat{\mu}_k^{(t)}, \hat{\Sigma}_k^{(t)})}{\sum_{k=1}^K \hat{\pi}_k^{(t)} g_k(x_i | \hat{\mu}_k^{(t)}, \hat{\Sigma}_k^{(t)})}$$

2) In the M-step, the parameter estimates of π_k , μ_k and Σ_k are updated given the estimated posterior probabilities, using the update equations

$$\begin{aligned} \hat{\pi}_k^{(t+1)} &= \frac{1}{n} \sum_{i=1}^n \hat{T}_{ik} \\ \hat{\mu}_k^{(t+1)} &= \frac{1}{n \hat{\pi}_k^{(t+1)}} \sum_{i=1}^n x_i \hat{T}_{ik} \\ \hat{\Sigma}_k^{(t+1)} &= \frac{1}{n \hat{\pi}_k^{(t+1)}} \sum_{i=1}^n \hat{T}_{ik} (x_i - \hat{\mu}_k^{(t+1)})' (x_i - \hat{\mu}_k^{(t+1)}) \end{aligned}$$

3) Iterate the first two steps until convergence.

The EM algorithm requires two issues to be addressed; determining the number of components, K , and initialization of the parameters.

4. The Model Selection Criteria

After estimating the parameters for the covariance matrix, the next step of determining the optimal cluster structure is selecting the best model. Despite the vast number of different model selection criteria in the literature, Schwarz's Bayesian Criteria (SBC) (Schwarz [17]) is no doubt the most widely used in the model-based clustering. Besides these criteria, two other selection criteria are used. Namely AIC (Akaike [8]) and the information complexity (ICOMP) criterion (Bozdogan [18]). When using any information criterion to perform model selection, the model corresponding to the lowest score as providing the best balance between good fit and parsimony is chosen. Using the likelihood function, the AIC and SBC functions of the Gaussian mixture model can be defined as follows:

$$\text{AIC} = -2 \log L(\hat{\theta} | x) + 2m$$

$$\text{SBC} = -2 \log L(\hat{\theta} | x) + m \log(n)$$

where: $L(\hat{\theta} | x)$ is the likelihood function.

m is the number of independent parameters to be estimated.

$\hat{\theta}$ is the maximum likelihood estimate for parameter θ .

ICOMP, originally introduced by Bozdogan [7] [9] [18] [19], is a logical extension of AIC and SBC, based on the structural complexity of an element or set of random vectors via the generalization of the information-based covariance complexity index of Van Emden [20]. ICOMP penalizes the lack-of-fit of a model with twice the negative of the maximized log-likelihood, following the same procedure of AIC and SBC. However, in ICOMP, a combination of lack-of-parsimony and profusion-of-complexity are also simultaneously penalized by a scalar complexity measure, C , of the model covariance matrix; while in AIC and SBC, only the lack of parsimony is penalized in terms of the number of parameters. In general, ICOMP is defined by using the likelihood function, the AIC and SBC functions of the Gaussian mixture model can be defined as follows:

$$\text{ICOMP} = -2 \log L(\hat{\theta} | x) + 2C(\widehat{\text{Cov}}(\theta))$$

where: $L(\hat{\theta} | x)$ is the likelihood function.

C is a real-valued complexity measure.

$\widehat{\text{Cov}}(\theta)$ is the estimated model covariance matrix.

The covariance matrix is estimated by the estimated inverse Fisher information matrix (IFIM), $\hat{\mathcal{F}}^{-1}$ is given by:

$$\hat{\mathcal{F}}^{-1} = \left\{ -E \left[\frac{\partial^2 \log L(\hat{\theta})}{\partial \theta \partial \theta'} \right] \right\}^{-1}$$

That is to say, IFIM is the negative expectation of the matrix of the second partial derivatives of the maximized log-likelihood of the fitted model, evaluated at the maximum likelihood estimators $\hat{\theta}$.

For a multivariate normal model, the general form of ICOMP is defined as:

$$\text{ICOMP}_{\text{PEU}}(\hat{\mathcal{F}}^{-1}) = -2 \log L(\hat{\theta} | x) + m + \log(n) C_1(\hat{\mathcal{F}}^{-1})$$

where:

$$C_1(\hat{\mathcal{F}}^{-1}) = \frac{S}{2} \log \left[\frac{\text{tr}(\hat{\mathcal{F}}^{-1})}{s} \right] - \frac{1}{2} \log |\hat{\mathcal{F}}^{-1}|$$

$$s = \dim(\hat{\mathcal{F}}^{-1}) = \text{rank}(\hat{\mathcal{F}}^{-1})$$

For all the above criteria, the decision rule is to select the model that gives the minimum score for the loss function.

5. The Numerical Results

All results were obtained by using MATLAB.

The Gaussian mixture-model based clustering is applied, which implements the EM algorithm for inference, to four simulated data sets. The maximum number of clusters is taken $K_{\max} = 6$ for all examples. The convergence criteria of the EM algorithm are set to $\text{see} = 10^{-6}$ and a maximum of 1000 iterations is allowed. After confirming the validity of mathematical equations and the program, four models of covariance matrix were applied. These models are:

Model: EVV with the covariance matrix $(\lambda D_k A_k D_k^T)$.

Model: VII with the covariance matrix $(\lambda_k I)$.

Model: VEE with the covariance matrix $(\lambda_k D A D^T)$.

Model: VVE with the covariance matrix $(\lambda_k D A_k D^T)$.

These models have been selected due to their distinguishing features: They represent different cases of the covariance matrix. Where the models [EVV] [VEE] and [VVE] belong to the General Family (Celeux and Govaert [4]). While the model [VII] belongs to the spherical family. In all models, the AIC, SBC and ICOMPPEU parameters were calculated. The optimal number of clusters has been determined by reaching the lowest values. The values of the complexity criteria were as follows:

1) Model: EVV with the covariance matrix $(\lambda D_k A_k D_k^T)$ (Figure 1 & Figure 2)

From Table 3, the optimal number of clusters was determined. It was determined by achieving the lowest values of the criteria at the same time. It was found to fit the number of clusters of two clusters.

Given below in Table 4 the parameter values estimated for the best simulation.

For the selected model, GMMC identifies the cluster labels with a miss classification rate of 1%. The miss classification rate is calculated as follows:

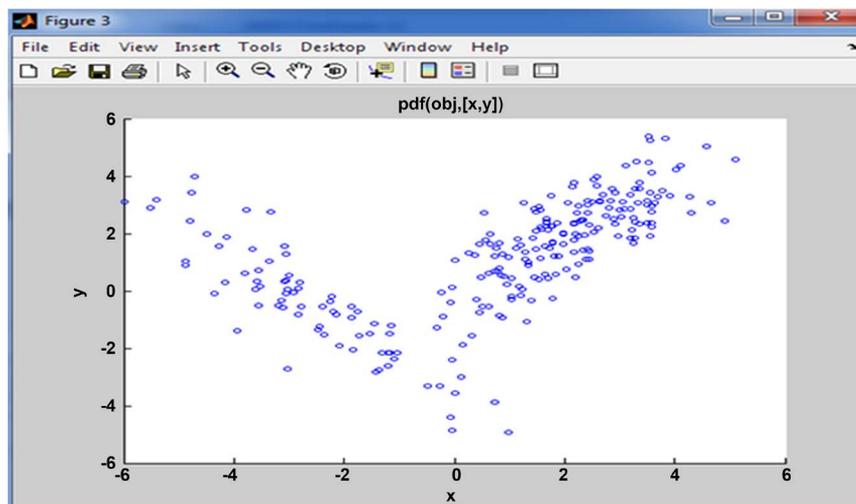


Figure 1. Scatterplot of the actual dataset labeled by groups (Model EVV).

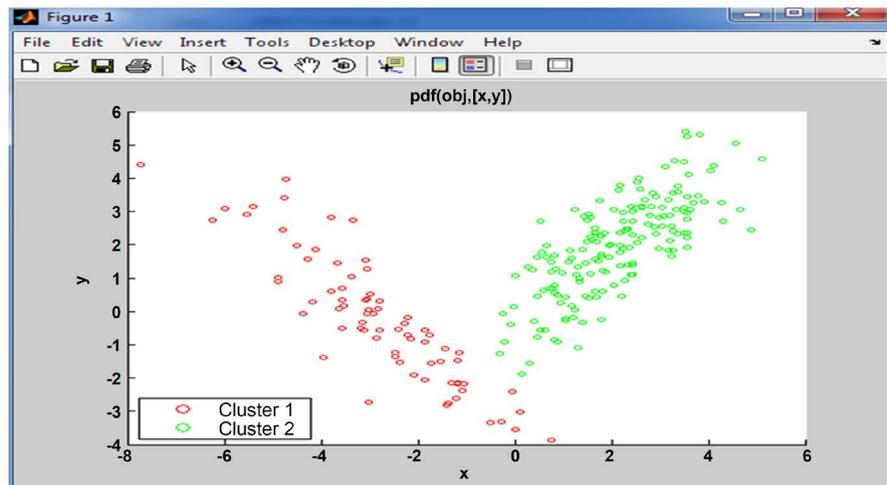


Figure 2. Scatterplot of the estimated dataset labeled by groups (Model EVV).

Table 3. Values of the criteria for selecting the model to reach the best simulation for the model (EVV) for the number of clusters $k = 1, \dots, 6$.

No. of clusters	AIC	SBC	ICOMPPEU
1	2188.1	2188.3	2174.1
2	1837.8	1844.3	1826.8
3	1839.7	1852.3	1828.7
4	1842	1860.8	1831
5	1843.4	1868.4	1832.4
6	1853.2	1884.5	1842.2

Table 4. The resulting confusion matrix for model (EVV).

Model	Input	Parameters	Output	
EVV	$n = 250$ $n_1 = 175$ $n_2 = 75$ $K = 2$	$\lambda = 2$	No. of simulations = 100 $COV_1 = \begin{bmatrix} 1.2671 & 1.2559 \\ 1.2559 & 2.0334 \end{bmatrix}$ $COV_2 = \begin{bmatrix} 3.738 & -3.6801 \\ -3.6801 & 4.5328 \end{bmatrix}$ $\pi_k = [0.6997 \quad 0.3003]$ $\mu_k = \begin{bmatrix} 1.9466 \\ 1.8859 \end{bmatrix}, \begin{bmatrix} -2.8553 \\ -0.1221 \end{bmatrix}$	
		$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$		
		$A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$		
		$COV_1 = \begin{bmatrix} 1.2929 & 1.2483 \\ 1.2483 & 2.000 \end{bmatrix}$		
		$COV_2 = \begin{bmatrix} 4.7071 & -4.6268 \\ -4.6268 & 5.4142 \end{bmatrix}$		
Predicted				
		1	2	Total
Actual	1	174	1	175
	2	0	75	75
	Total	174	76	250

$$\left(1 - \frac{a_{ii} + a_{jj}}{\Sigma}\right) \times 100 = \left(1 - \frac{174 + 75}{250}\right) \times 100 = \left(1 - \frac{249}{250}\right) \times 100 = (1 - 0.99) \times 100 = 1$$

2) Model: VII with the covariance matrix ($\lambda_k I$) (Figure 3 & Figure 4)

Using Table 5, the optimal number of clusters was two clusters. GMMC achieves a miss classification rate of 2% for the model (VII). The resulting confusion matrix is shown in Table 6.

3) Model: VEE with the covariance matrix ($\lambda_k DAD^T$) (Figure 5 & Figure 6)

From the results in Table 7, it was found that the optimal number of clusters is three, so the number of clusters was increased. To achieve greater clarity, the sample size was 500 instead of 250 and was divided into three groups as follows (Table 8).

For this model, the miss classification rate was 15%.

4) Model: VVE with the covariance matrix ($\lambda_k DA_k D^T$) (Figure 7 & Figure 8)

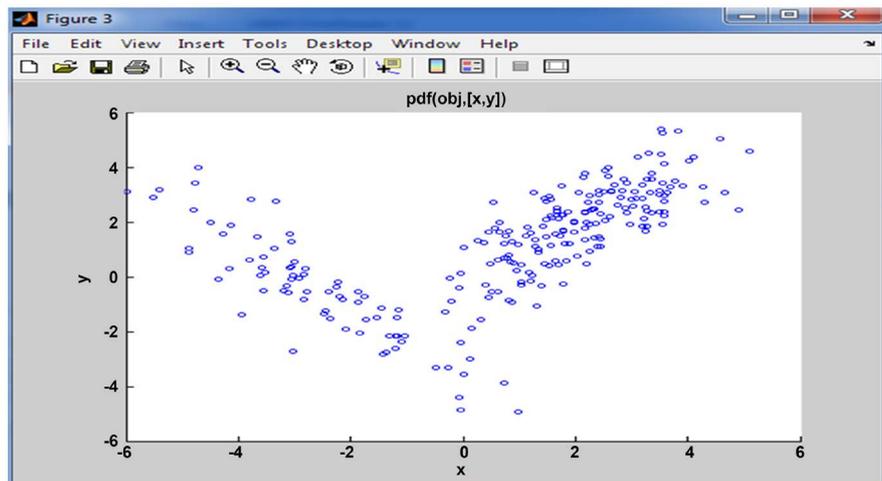


Figure 3. Scatterplot of the actual dataset labeled by groups (Model VII).

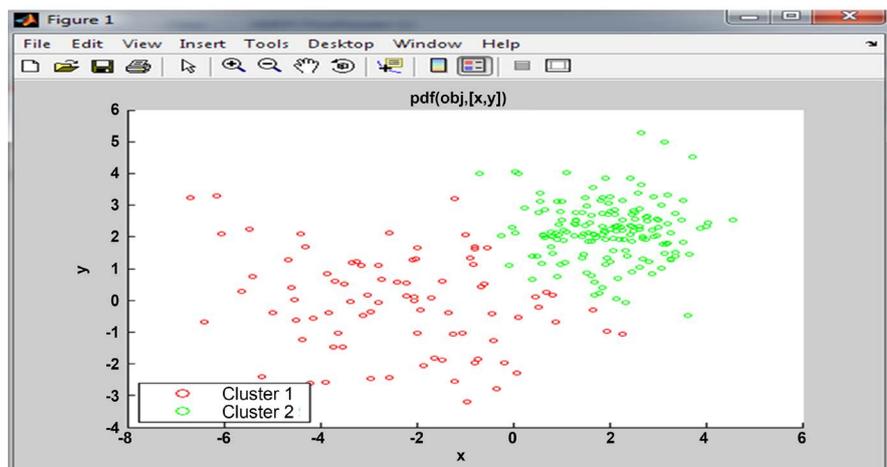


Figure 4. Scatterplot of the estimated dataset labeled by groups (Model VII).

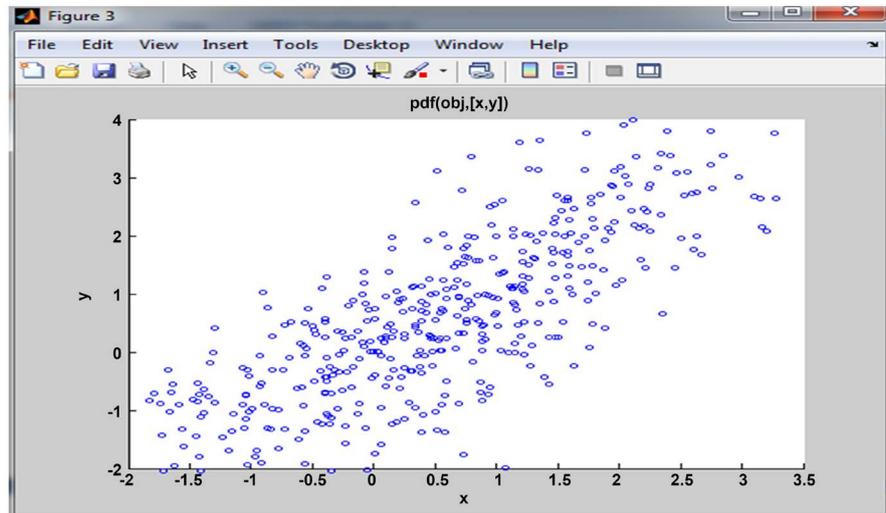


Figure 5. Scatterplot of the actual dataset labeled by groups (Model VEE).

Table 5. Values of the criteria for selecting the model to reach the best simulation for the model (VII) for the number of clusters $k = 1, \dots, 6$.

No. of clusters	AIC	SBC	ICOMPPEU
1	2034.3	2034.5	2018.8
2	1873.5	1904.8	1859.8
3	1878.1	1890.8	1864.4
4	1877.7	1896	1863.5
5	1882.4	1907.4	1868.7
6	1888.74	1893.8	1873.7

Table 6. The resulting confusion matrix for model (VII).

Model	Input	Parameters	Output	
VII	$n = 250$	$\lambda_k = 1, 2$	No. of simulations = 100	
		$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$		
$\lambda_k I$	$n_1 = 175$	$A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$	$COV_1 = \begin{bmatrix} 1.0460 & -0.0659 \\ -0.0659 & 0.8913 \end{bmatrix}$	
		$n_2 = 75$	$COV_2 = \begin{bmatrix} 2.4341 & -0.5134 \\ -0.5134 & 1.9715 \end{bmatrix}$	
	$K = 2$	$COV_1 = \begin{bmatrix} 1.2929 & 1.2483 \\ 1.2483 & 2.000 \end{bmatrix}$	$\pi_k = [0.7040 \quad 0.2960]$	
		$COV_2 = \begin{bmatrix} 4.7071 & -4.6268 \\ -4.6268 & 5.4142 \end{bmatrix}$	$\mu_k = \begin{bmatrix} 1.9139 & -2.7437 \\ 2.0450 & -0.1582 \end{bmatrix}$	
		$\pi_k = [0.7 \quad 0.3]$		
		$\mu_k = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \end{bmatrix}$		
Predicted				
		1	2	Total
Actual	1	175	0	175
	2	3	72	75
	Total	178	72	250

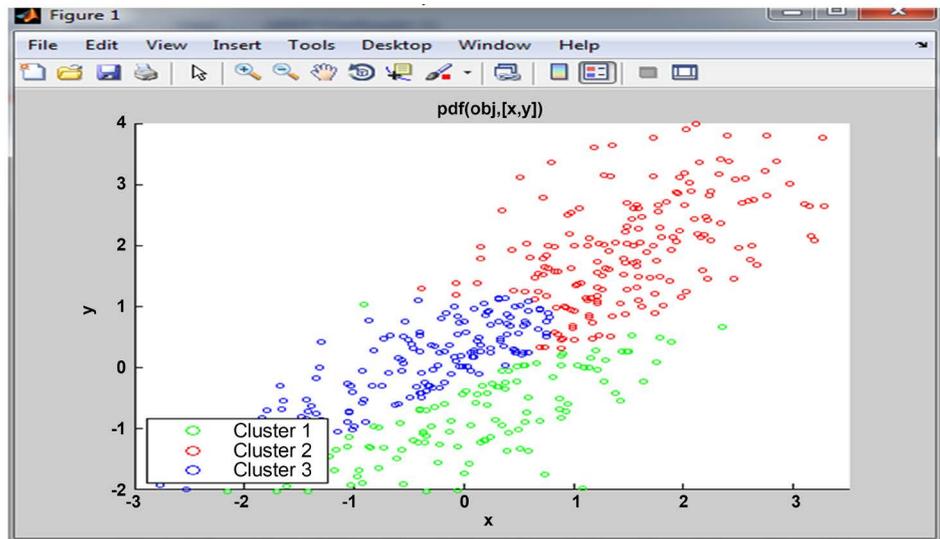


Figure 6. Scatterplot of the estimated dataset labeled by groups (Model VEE).

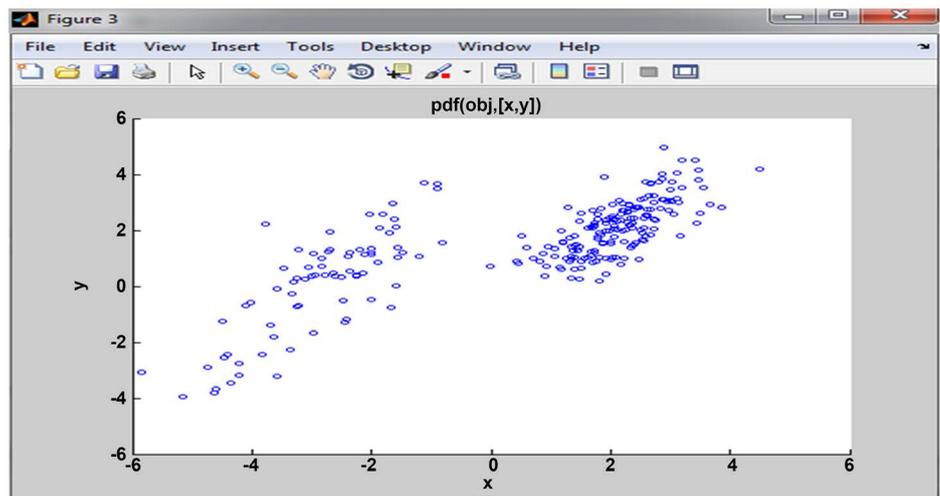


Figure 7. Scatterplot of the actual dataset labeled by groups (Model VVE).

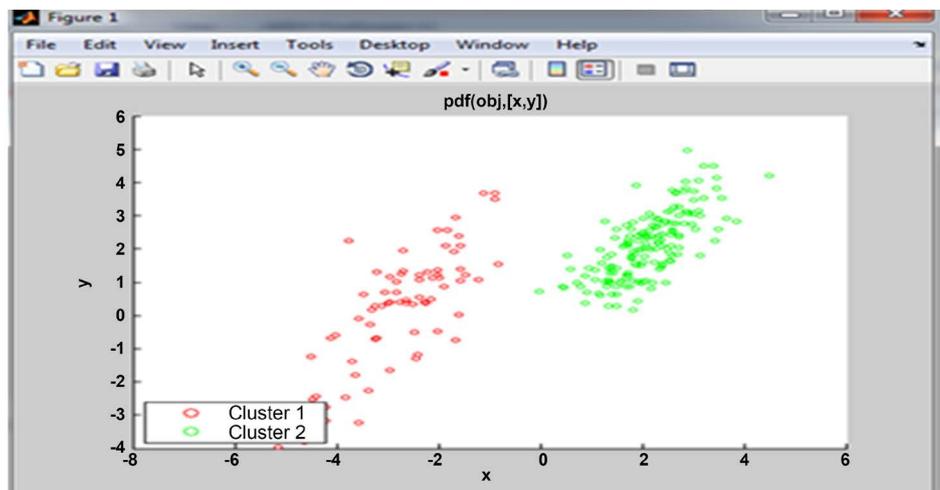


Figure 8. Scatterplot of the estimated dataset labeled by groups (Model VVE).

Table 7. Values of the criteria for selecting the model to reach the best simulation for the model (VEE) for the number of clusters $k = 1, \dots, 6$.

No. of clusters	AIC	SBC	ICOMPPEU
1	2043.4	2043.7	2028.1
2	1738.4	1744.9	1727.6
3	1736.1	1748.7	1725.2
4	1742.7	1761.5	1731.8
5	1752.6	1777.7	1741.8
6	1752.7	1784	1741.9

Table 8. The resulting confusion matrix for model (VEE).

Model	Input	Parameters	Output			
VEE $\lambda_k DAD^T$	$n = 500$ $n_1 = 150$ $n_2 = 200$ $n_3 = 150$ $K = 3$	$\lambda_k = 1, 1.5, 3$	No. of simulations = 100 $COV_1 = \begin{bmatrix} 0.8652 & 0.6431 \\ 0.6431 & 0.8904 \end{bmatrix}$ $COV_2 = \begin{bmatrix} 1.7439 & 2.1360 \\ 2.1360 & 3.2356 \end{bmatrix}$ $COV_3 = \begin{bmatrix} 1.2078 & 0.8378 \\ 0.8378 & 1.7065 \end{bmatrix}$ $\pi_k = [0.4559 \quad 0.4162 \quad 0.1279]$ $\mu_k = \begin{bmatrix} 0.7230 \\ 1.2093 \end{bmatrix}, \begin{bmatrix} 0.3678 \\ 0.0200 \end{bmatrix}, \begin{bmatrix} 1.1760 \\ 0.3249 \end{bmatrix}$			
		$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$				
		$D = \begin{bmatrix} \cos\left(6 * \frac{\pi}{8}\right) & \sin\left(\frac{\pi}{8}\right) \\ -\sin\left(\frac{\pi}{8}\right) & \cos\left(\frac{\pi}{8}\right) \end{bmatrix}$				
		$COV_1 = \begin{bmatrix} 0.6464 & 0.6242 \\ 0.6242 & 1.000 \end{bmatrix}$				
		$COV_2 = \begin{bmatrix} 0.9697 & 0.9362 \\ 0.9362 & 1.500 \end{bmatrix}$				
		$COV_3 = \begin{bmatrix} 1.9393 & 1.8725 \\ 1.8725 & 3.000 \end{bmatrix}$				
		$\pi_k = [0.3 \quad 0.5 \quad 0.2]$				
		$\mu_{1,2,3} = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -0.5 \end{bmatrix}$				
		Predicted				
					1	2
Actual	1	126	24	0	150	
	2	30	170	0	200	
	3	20	1	129	150	
	Total	176	195	129	500	

The fit number of clusters for this model was two clusters (Table 9).

It was shown that the miss classification rate was 0% from the data in Table 10.

6. Conclusion

In this paper, the Gaussian mixture model-based clustering is used. The mixture models based on clusters are able to predict accurately if the appropriate covariance matrix, model is selected. It is applied by using four models:

Table 9. Values of the criteria for selecting the model to reach the best simulation for the model (VVE) for the number of clusters $k = 1, \dots, 6$.

No. of clusters	AIC	SBC	ICOMPPEU
1	1884.1	1884.3	1872.6
2	1523.1	1529.5	1516.1
3	1527.7	1540.4	1520.7
4	1523.8	1542.6	1516.8
5	1536.7	1561.8	1529.7
6	1534.7	1566	1527.7

Table 10. The resulting confusion matrix for model (VVE).

Model	Input	Parameters	Output		
VVE $\lambda_k D A_k D^T$	$n = 250$ $n_1 = 175$ $n_2 = 75$ $K = 2$	$\lambda_k = 1, 1.5$	No. of simulations = 100 $COV_1 = \begin{bmatrix} 0.6007 & 0.5619 \\ 0.5619 & 0.9121 \end{bmatrix}$ $COV_2 = \begin{bmatrix} 1.1635 & 1.8430 \\ 1.8430 & 4.1089 \end{bmatrix}$ $\pi_k = [0.7039 \quad 0.2961]$ $\mu_k = \begin{bmatrix} 2.0486 & -3.0599 \\ 2.0247 & -0.0947 \end{bmatrix}$		
		$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$			
		$A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$			
		$D = \begin{bmatrix} \cos\left(6 * \frac{\pi}{8}\right) & \sin\left(\frac{\pi}{8}\right) \\ -\sin\left(\frac{\pi}{8}\right) & \cos\left(\frac{\pi}{8}\right) \end{bmatrix}$			
		$COV_1 = \begin{bmatrix} 1.1980 & 1.6492 \\ 1.6492 & 3.0186 \end{bmatrix}$			
		$COV_2 = \begin{bmatrix} 1.4402 & 1.7089 \\ 1.7089 & 3.2965 \end{bmatrix}$			
		$\pi_k = [0.7 \quad 0.3]$			
		$\mu_k = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \end{bmatrix}$			
		Predicted			
					1
Actual	1	175	0	175	
	2	0	75	75	
	Total	175	75	250	

1) Model [EVV] ($\lambda D_k A_k D_k^T$) represents the case of equal volume, variable shape, and orientation. It is showed that the optimal number of clusters equals two. From the values of the complexity criteria in **Table 3**, it is noted that the ICOMPPEU criterion corresponds to the lowest value compared to the other two criteria and the miss classification rate was 1%.

2) Model [VII] ($\lambda_k I$) represents the case of variable volume, shape, and orientation. Also, in this model, the optimal number of clusters equals two and the ICOMPPEU criterion corresponds to the lowest value compared to the other two parameters (the values in **Table 5**). The miss classification rate was 2%.

3) Model [VEE] ($\lambda_k DAD^T$) represents the case of variable volume, equal shape, and direction. From **Table 7**, it is found that the optimal number of clusters is calculated by the number of clusters corresponding to the lowest values of the complexity of the information and found to be equal to three. The miss classification rate was 15%.

4) Model [VVE] ($\lambda_k DA_k D^T$) represents the case of variable volume, shape, and equal orientation. As the first and second model the optimal number of clusters equals two the ICOMPPEU criterion corresponds to the lowest value compared to the other two criteria (values are found in **Table 9**, while the miss classification rate was 0%.

The results showed that the ICOMPPEU criteria were superior to the rest of the criteria. In addition to the success of the Gauss model based on the clusters in the prediction using the covariance matrix. The study also determined the possibility of determining the optimal number of clusters by selecting the number of clusters corresponding to the lowest values of the different criteria.

For the number of clusters $k = 1, \dots, 6$, the three different selection criteria have chosen the VVE model for the number of clusters two to be the optimal model. For the selected model, the Gaussian Mixture Model-based Clustering (GMMC) diagnoses the cluster classification with a 0% miss classification rate.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Everitt, B. and Skrondal, A. (2010) The Cambridge Dictionary of Statistics. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511779633>
- [2] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**, 1-38. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- [3] Banfield, J.D. and Raftery, A.E. (1993) Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics*, **49**, 803-821. <https://doi.org/10.2307/2532201>
- [4] Celeux, G. and Govaert, G. (1995) Gaussian Parsimonious Clustering Models. *Pattern Recognition*, **28**, 781-793. [https://doi.org/10.1016/0031-3203\(94\)00125-6](https://doi.org/10.1016/0031-3203(94)00125-6)
- [5] Chi, J., Yuchen, Z., Sivaraman, B., Martin, J. and Michael, I. (2016) Local Maxima in the Likelihood of Gaussian Mixture Models: Structural Results and Algorithmic Consequences. *Advances in Neural Information Processing Systems*, Barcelona, 5-10 December 2016, 4116-4124.
- [6] Tóth, B.G., Rácz, I.I. and Horváth, I. (2019) Gaussian-Mixture-Model-Based Cluster Analysis of Gamma-Ray Bursts in the BATSE Catalog. *Monthly Notices of the Royal Astronomical Society*, **486**, 4823-4828. <https://doi.org/10.1093/mnras/stz1188>
- [7] Bozdogan, H. (2000) Akaike's Information Criterion and Recent Developments in Information Complexity. *Journal of Mathematical Psychology*, **44**, 62-91. <https://doi.org/10.1006/jmps.1999.1277>

- [8] Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrox, B. and Csaki, F., Eds., *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, 267-281.
- [9] Bozdogan, H. (1988) ICOMP: A New Model-Selection Criteria. In: Bock, H., Ed., *Classification and Related Methods of Data Analysis*, North-Holland, Amsterdam, 599-608.
- [10] Titterton, D., Smith, A. and Makov, U. (1985) *Statistical Analysis of Finite Mixture Distributions*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. Wiley, Hoboken.
- [11] Erar, B. (2011) Mixture model Cluster Analysis under Different Covariance Structures Using Information Complexity. Master's Thesis.
- [12] Gupta, S. and Bhatia, V. (2015) Gaussian Mixture Model Based Clustering Hierarchy Protocol in Wireless Sensor Network. *International Journal of Scientific Engineering and Research*, **3**, 2347-3878.
- [13] Scrucca, L., Fop, M., Murphy, T.B. and Raftery, A.E. (2016) Mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R Journal*, **8**, 205-233. <https://doi.org/10.32614/RJ-2016-021>
- [14] Malsiner-Walli, G., Frühwirth-Schnatter, S. and Grün, B. (2016) Model-Based Clustering Based on Sparse Finite Gaussian Mixtures. *Statistics and Computing*, **26**, 303-324. <https://doi.org/10.1007/s11222-014-9500-2>
- [15] Wu, C.J. (1983) On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, **11**, 95-103. <https://doi.org/10.1214/aos/1176346060>
- [16] Xu, J., Hsu, D. and Maleki, A. (2018) Benefits of Over-Parameterization with EM. *32nd Conference on Neural Information Processing Systems (NIPS 2018)*, Montréal, p. 35.
- [17] Schwarz, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, **6**, 461-464. <https://doi.org/10.1214/aos/1176344136>
- [18] Bozdogan, H. (1994) Mixture-Model Cluster Analysis Using Model Selection Criteria and a New Informational Measure of Complexity. In: Bozdogan, H., Ed., *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach*, Kluwer Academic Publishers, Dordrecht, Vol. 2, 69-113. https://doi.org/10.1007/978-94-011-0800-3_3
- [19] Bozdogan, H. (2000) On the Information-Based Measure of Covariance Complexity and Its Application to the Evaluation of Multivariate Linear Models. *Communications in Statistics—Theory and Methods*, **19**, 221-278. <https://doi.org/10.1080/03610929008830199>
- [20] Van Emden, M. (1971) An Analysis of Complexity. In: *Mathematical Centre Tracts*, Mathematisch Centrum, Amsterdam, 35.