Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

Comparison of Some Methods for Estimating Mixture of Linear Regression Models with Application

Urdak Ibrahim Kareem
Dept. of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq
Iraqurdak124ibrahim@gmail.com

Fadhaa Mezher Hashim
Dept. of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq
fali197477@gmail.com

Received: 25/1/2021 Accepted: 15/3/2021 Published: SEPTEMBER/ 2021

Abstract:

A mixture model is used to model data that come from more than one component. In recent years, it became an effective tool in drawing inferences about the complex data that we might come across in real life. Moreover, it can represent a tremendous confirmatory tool in classification observations based on similarities amongst them. In this paper, several mixture regression-based methods were conducted under the assumption that the data come from a finite number of components. A comparison of these methods has been made according to their results in estimating component parameters. Also, observation membership has been inferred and assessed for these methods. The results showed that the flexible mixture model outperformed the others in most simulation scenarios according to the integrated mean square error and integrated classification error.

Keywords: Mixture Model, EM algorithm, Linear Regression, Trimmed Maximum Likelihood, Laplace Distribution

1. Introduction:

Finite mixture models are considered as mathematical approaches for a wide range of statistical models that have been used to model a variety of random phenomena. The most common use of it can be seen in many statistical applications that have been conducted in medicine, genetics, economics, biology, and agriculture. One key advantage in these applications is the determination of points membership when no information is available (B.G.Lindsay 1995, D. Böhning 2000, G.J. McLachlan and D. Peel 2000)[12,3,14]. A mixture of linear regression was proposed for the first time by (R. Quandt and J. Ramsey,1978)[19] as a switching regression model in a general framework. Their idea was to estimate the parameters of the model by employing a method of moments. With the fast pace of technological revolution, data and phenomena become much complicated to be analyzed statistically. This can be seen in the scenarios in which the data cannot interpret one single model. One successful work was done by De Veaux(1989)[23] to tackle complicated situations through developing an EM algorithm approach to fit a two-component mixture of linear regression models.

Turner (2000)[22] applied a finite mixture model of linear regression to different types of data by the use of an EM algorithm. One year later, another problem was considered to determine the number of components investigated by A.J. Stomberg (2001)[8], who used a likelihood-derived method based on the complete data. Assigning a weight factor for each observation was proposed by (Markatou 2000)[13] and (Shen et al., 2004)[20] as a robust method for mixture models. The trimmed likelihood-based method has taken a good part in the estimation mixture of linear regressions through a well-developed framework by Filzmoser, Dimova, and Neytchev (2007)[17]. Other methods were proposed to employ this model as required classification methods to cluster data to several groups based on similarity amongst them by Gordaliza et al. (2010), Garc´ıa-Escudero, et al. (2009-2010)[5,6], Hennig (2002, 2003)[9,10], Mueller et al. (2005)[16]. A critical advance in researches involving mixture model was modifying the standard EM algorithm as an alternative for paying attention to the maximization of the function in estimating the parameters of a mixture model of linear regression, which was an excellent robust method proposed by (Bai, X. et al. 2012)[2]. Another robust method was developed by (Song W. et al., 2014)[21] that assumed Laplace distribution for the error terms and then used the EM algorithm to implement the estimation procedure considering the Laplace distribution as a mixture of normal distribution. This research included several sections. In the next section, we introduced the general form of a finite mixture of linear regression models followed by the section of estimation model parameters. The latter sections would described the simulation procedure and discuss the results of the methods.

2. Materials and Methods

Let X be a p-dimensional vector of independent variables, and Y be a scalar response variable. To look into the relationship between Y and X, we use the classical linear regression model. However, if the model has p components (p>2) and (X', Y) that have the probability π_j, where , $j = 1, 2, \dots, p$. We may use the mixture linear regression

$$Y_i = X_i^T \beta_j + \varepsilon_j , \qquad i = 1, 2, \dots, n;\ j = 1, 2, \dots, p \quad \dots\dots (1)$$

where ε_j is a random error and distributed as $N(0, \sigma^2{}_j)$, and $\sum_{j=1}^{p} \pi_j = 1$, for $j = 1, 2, \dots, p$.

Hence,

$$f(x, \beta) = \sum_{j=1}^{p} \pi_j \phi(y; x'\beta_j, \sigma_j^2), \dots\dots (2)$$

where $\sum_{j=1}^{p} \pi_j \phi(y; x'\beta_j, \sigma_j^2)$ is a density function of normal distribution. To Eq(2), we define an indicator vector of observations' memberships as follow:

$Z_{ij} = \{1 \quad if\ ith\ observation\ (X_i, Y_i)\ is\ from\ jth\ component\ \ 0 \quad otherwise.$

The incomplete log-likelihood function can be written in the following form:

$$l(x) = \sum_{i=1}^{n} log \sum_{j=1}^{p} \pi_j \phi(y_j; x'\beta_j, \sigma_j^2), \qquad \dots (3)$$

where $\theta = (\beta, w, \sigma^2)^T$

And the complete log-likelihood function is written as

$$l(x) = \sum_{i=1}^{n} \sum_{j=1}^{p} z_{ij} log\ [\pi_j \phi(y_j; x'\beta_j, \sigma_j^2)] \qquad \dots (4)$$

Note that finding the estimators of eq. (4) is quite tricky analytically. Therefore, many methods have been found to tackle such issues. In this research, a comparison of three critical methods would be conducted to show the best performance.

3. Estimation Methods

Several methods are being discussed in terms of their frameworks in estimating the mixture of linear regressions. In this study, we are interested in applying three important methods in this study.

3.1 EM Algorithm as Flexible Fitting of Finite Mixture Models

One classical method that is widely used with the mixture model is the EM algorithm. It is an iterative method designed to find the solution to the maximum likelihood (Lindsay, B. G., 1995)[12]. The E and M steps of the EM algorithm can be summarized as follow:

E-step: In this step, we calculate the function $Q(\theta, \theta^{(r)})$ Which represents the expectation of the log-likelihood function.

$$Q(\theta,\theta^{(r)}) = E(l(\theta|X,Y,Z)) = \sum_{i=1}^{n}\sum_{j=1}^{p} \hat{z}_{ij}{}^{(r)}\left[log\ (\pi_j) + log\ \left(\emptyset_j(y_j; x_i{}'B_j,\sigma_j^2)\right)\right], \ldots (5)$$

where

$$\hat{z}_{ij}{}^{(r)} = \frac{\pi_j^r \emptyset_j(y_j; x_i{}'B_j,\sigma_j^2)}{\sum_{j=1}^{p} \pi_j^r \emptyset_j(y_j; x_i{}'B_j,\sigma_j^2)} \ldots (6)$$

Here, $\hat{z}_{ij}{}^{(r)}$ is the estimated probability of the i^{th} the object comes from the j^{th} component of the mixture model after r iteration.

M_step: This is an updating step of the estimator $\theta^{(r+1)}$ that maximizes the Q-function by letting $\frac{\partial\, Q(\theta,\theta^{(r)})}{\partial\theta} = 0$ and solve it for θ. We then get:

$$\hat{\pi}_j^{(r+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}{}^{(r)}}{n} \quad (j = 1, \ldots, p) \ldots (7)$$

$$\hat{\beta}_j^{(r+1)} = (X'W_jX)^{-1} X'W_jY \quad (j = 1,.., p), \ldots (8)$$

where X is a matrix which has n rows and $p+1$ columns, and W_j is a matrix of $n \times n$ elements, say $w_{ij}^{(r)}$, that is $w_{ij}^{(r)} = diag\ (\hat{z}_{ij}{}^{(r)})$.

$$\hat{\sigma}_j^{(r+1)} = \sqrt{\frac{\sum_{i=1}^{n} \hat{z}_{ij}{}^{(r)}(y_i - x_i{}'\hat{\beta}_j^{(r+1)})^2}{\sum_{j=1}^{n} \hat{z}_{ij}{}^{(r)}}} \quad (j = 1, \ldots, p) \ldots \ldots (9)$$

These two steps are repeated until the convergence criterion is achieved.

3.2 Laplace Distribution Method for Mixture Model

This method was proposed by Weixing Song, Weixin Yao and Yanru Xing (2014)[21]. It assumes a Laplace distribution with mean 0 and scale parameter $\frac{1}{\sqrt{2}}$ as a distribution for σ_j's. To this end, we set G={(X_i,Y_i,Z_{ij})} to be the full observable data, $i = 1, 2, \ldots, n; j = 1, 2, \ldots p$. Therefore, the function of the complete log-likelihood can be written as

$$log\ log\ L(\theta; G) = \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\ log\ log\ \frac{\pi_j}{\sqrt{2}\,\sigma_j}\ exp\ exp\left(-\frac{\sqrt{2}\,|Y_i - X_i'\beta_j|}{\sigma_j}\right) \ldots \ldots (10)$$

From the point of view of both (Andrews and Mallows, 1974)[1], the random variable that follows Laplace distribution is a mixture of two random variables of which one is distributed as a normal distribution and the other as an exponential distribution. Denote M_i, coupled with (X_i, Y_i), to be latent scale variable and let $V = \{X_i, Y_i, M_i, Z_{ij}\}$, where $i = 1, 2, \ldots, n; j = 1, 2, \ldots, p$. The complete log-likelihood function then is written as:

$$log\ log\ L(\theta;V) = \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\ log\ log\ \pi_j \frac{M_i}{\sqrt{\pi}\,\sigma_j} exp\ exp\left(-\frac{M_i^2\left(Y_i - X_i'\beta_j\right)^2}{\sigma_j^2}\right)\frac{1}{M_i^3} exp\ exp\left(-\frac{1}{2M_i^2}\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\ log\ log\left[\pi_j \frac{M_i}{\sqrt{\pi}\,\sigma_j} exp\ exp\left(-\frac{M_i^2\left(Y_i - X_i'\beta_j\right)^2}{\sigma_j^2}\right)\frac{1}{M_i^3} exp\ exp\left(-\frac{1}{2M_i^2}\right)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\ log\ log\ \pi_j + \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\frac{M_i^2\left(Y_i - X_i'\beta_j\right)^2}{\sigma_j^2} + \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\ log\ log\ \frac{1}{M_i^2} + \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\left(-\frac{1}{2M_i^2}\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\ log\ log\ \pi_j - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\ log\ log\ \pi_j\sigma_j^2 - \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\frac{M_i^2(Y_i - X_i'B)^2}{\sigma_j^2}$$

$$- \sum_{i=1}^{n}\sum_{j=1}^{p} Z_{ij}\ log\ log\ M_i^2 - \sum_{i=1}^{n}\sum_{j=1}^{p} \frac{Z_{ij}}{M_i^2} \ldots . (1.11)$$

To this end, applying EM algorithm involves calculating the conditional expectation of $[L(\theta;V)|\ H,\theta^{(0)}]$ where $\theta^{(0)} = \left(\beta_1^{(0)}, \sigma_1^{2(0)}, \pi_1^{(0)}, \ldots, \beta_p^{(0)}, \sigma_p^{2(0)}, \pi_p^{(0)}\right)$, where $\theta^{(0)}$ is the initial values that of which the algorithm starts. What is significant about the last two terms in (11) is the fact that they are free of the unknown regression parameters so that they can be ignored in this study. Therefore, to calculate E[L(θ; V)| $H, \theta^{(0)}$], we just need to evaluate

$log\ log\ L(\theta;V) = = \sum_{i=1}^{n} \sum_{j=1}^{p} Z_{ij}\ log\ log\ \pi_j - \frac{1}{2}\sum_{i=1}^{n} \sum_{j=1}^{p} Z_{ij}\ log\ log\ \pi_j\sigma_j^2 - \sum_{i=1}^{n} \sum_{j=1}^{p} Z_{ij} \frac{M_i^2(Y_i - X_i'\beta)^2}{\sigma_j^2}$.

Imply,

$$\tau_{ij} = E[H, \theta^{(0)}], \quad \varrho_{ij} = E[H, \theta^{(0)}, Z_{ij} = 1] \dots\dots (12)$$

The latter equation implies

$$\tau_{ij}^{(1)} = \frac{\pi_j^{(0)}\sigma_j^{-1(0)}\ exp(-\sqrt{2}\ |Y_i - X_i'\beta_j^{(0)}|/\sigma_j^{2(0)})}{\sum_{l=1}^{p} \pi_l^{(0)}\sigma_l^{-1(0)}\ exp(-\sqrt{2}\ |Y_i - X_i'\beta_l^{(0)}|/\sigma_l^{2(0)})} \dots\dots (13)$$

Phillips (2002) defined a thread to calculate ϱ_{ij}

$$\varrho_{ij}^{(1)} = \frac{\sigma_j^{(0)}}{\sqrt{2}\ |Y_i - X_i'\beta_j^{(0)}|} \dots\dots (14)$$

Concerning π_j, β_j and σ_j^2, the following term has to be maximized

$$\sum_{i=1}^{n} \sum_{j=1}^{p} \tau_{ij}\ log\ log\ \pi_j - \frac{1}{2}\sum_{i=1}^{n} \sum_{j=1}^{p} \tau_{ij}\ log\ log\ \sigma_j^2 - \sum_{j=1}^{n} \sum_{i=1}^{p} \frac{\tau_{ij}\varrho_{ij}(Y_i - X_i'\beta)^2}{\sigma_j^2} \dots\dots (15)$$

To maximize (15), we use the EM algorithm, as explained in the next section.

3.2.1 Applying EM algorithm under the assumption of Laplace distribution

The EM algorithm can be used in the following two steps:

1- Choose an initial value for $\theta = (\beta_1, \sigma_1^2, \pi_1, \dots\dots, \beta_p, \sigma_p^2, \pi_p)$.

2- In the E-step, once have reached $(r+1)^{th}$ iteration, $\tau_{ij}^{(r+1)}$ and $\varrho_{ij}^{(r+1)}$ should be calculated from equations (13) and (14) after replacing the initial values $\theta^{(0)}$ with $\theta^{(r)}$.

3- In the M-step, we maximize (15) using the following formulas:

$$\pi_j^{(r+1)} = \frac{1}{n}\sum_{i=1}^{n} \tau_{ij}^{(r+1)} \dots\dots (16)$$

$$\beta_j^{(r+1)} = \left(\sum_{i=1}^{n} \tau_{ij}^{(r+1)}\varrho_{ij}^{(r+1)} X_i X'\right)^{-1}\left(\sum_{i=1}^{n} \tau_{ij}^{(r+1)}\varrho_{ij}^{(r+1)} X_i Y_i\right), \dots\dots (17)$$

And

$$\sigma_j^{(r+1)} = \sqrt{\frac{2\sum_{i=1}^{n} \tau_{ij}{}^{(r+1)} \varrho_{ij}{}^{(r+1)} \left(Y_i - X_i' \beta_j^{(r+1)}\right)^2}{\sum_{i=1}^{n} \tau_{ij}{}^{(r+1)}}} \dots\dots (18)$$

These steps should be repeated until the convergence is reached.

To this end, the minor absolute deviation (LAD) is adopted in the EM algorithm to make it a robust method in estimating the model parameters.

3.3 Trimmed Maximum Likelihood Method

The definition of weighted trimmed likelihood estimator (WTLE) is mentioned in (Hadi and Lucefio, 1997)[7] and (Vandev and Neykov, 1998)[23] in the following formula:

$$\widehat{\theta}_{WTLE} = argmin_{\theta \in \Theta^p} \sum_{i=1}^{h} t_{u(i)}\, l\left(y_{u(i)}; \theta\right), \quad \dots\dots (20)$$

where $l\left(y_{u(i+1)}; \theta\right) \geq l\left(y_{u(i)}; \theta\right)$, θ is the vector of all parameters and $l(y_i; \theta) = -logf(y_i; \theta)$, and $y_i \in R^q$, for i=1,2,…,n are independently identically distributed with the density $f(y|x_i, \theta)$, which depends on an unknown parameter $\theta \in \Theta^q \subset R^q$, $U = (u(1), u(2), \dots, u(n))$ is the corresponding modification of the indices that depends on θ. Here, the parameter of trimming the high leverage points is denoted by h, and the weights $t_{u(i)} \geq 0$ for i=1,2,…,n are increasing function of $l\left(y_{u(i)}; \theta\right)$ such that $t_{v(h)} > 0$. Note that removing h high leverage points would be highly unlikely to occur if the fitted model was true, reflecting the underlying trimming in (20).

With that, the combinational procedure of the weighted trimmed likelihood estimator (WTLE) can be emphasized by

$$min_{\theta \in \Theta^p} \sum_{i=1}^{h} t_{u(i)} l\left(y_{u(i)}; \theta\right) = min_{\theta \in \Theta^p}\, min_{I \in I_h} \sum_{i \in I}^{h} t_{u(i)} l(y_i; \theta)$$

$$= min_{I \in I_h} min_{\theta \in \Theta^p} \sum_{i \in I}^{h} t_{u(i)} l(y_i; \theta) \dots\dots (21)$$

The set of all combinational subsets is denoted by $I_h$and equal to $\left(\frac{n}{h}\right)$ Which have to be fitted by the MLE, and the partition that results in the smallest negative log-likelihood would be WTLE.

3.3.1 Applying EM algorithm through Trimmed Maximum Likelihood Method

To apply TLE to solve Eq. (4), we need to formulate the Q function.

$$Q\left(\Theta; \Theta^{(l)}\right) = \sum_{i=1}^{n} \sum_{j=1}^{p} \tau_j(y_i; x_i, \Theta^{(l)}) \{log\ log\ \pi_j + log\ log\ \phi_j\left(y_i; x_i, \theta_j\right)\} \dots (22)$$

where $\tau_j\left(y_i; x_i, \Theta^{(l)}\right) = \pi_j^{(l)} \phi_j\left(y_i; x_i; \theta_j{}^{(l)}\right) / \sum_{j=1}^{p} \pi_j^{(l)} \phi_j\left(y_i; x_i; \theta_j{}^{(l)}\right)$

The Q function represents the probability that the i^{th} observation belongs to j^{th} mixture component. Q function aims to minimize the $logL(\Theta)$, i.e., $Q(\Theta;\Theta^{(l)}) \leq logL(\Theta^{(l)})$ and $Q(\Theta;\Theta^{(l)}) = log\ log\ L(\Theta^{(l)})$. In the maximization step, the function $Q(\Theta;\Theta^{(l)})$ It is minimized concerning Θ at the iteration $(l+1)$, which yields the new estimator for Θ. What is necessary to pay attention to is that the above two steps should be repeated until the stopping condition has been met. At the final iteration, the prior probability π_j have to be updated by the formula

$$\pi_j^{(l+1)} = \frac{\sum_{i=1}^{n} \tau_j(y_i; x_i, \Theta^{(l)})}{n} \ldots (23)$$

and the expression for θ_j is maximized,

$$max_{\theta_1,\ldots,\theta_k} \sum_{i=1}^{n} \sum_{j=1}^{p} \tau_j(y_i; x_i, \Theta^{(l)})\ log\ log\ \varphi_j(y_i; x_i; \theta_j) \ldots (24)$$

considering the posterior probabilities $\tau_j(y_i; x_i, \Phi^{(l)})$ As prior calculated weights. Since θ_j $(for\ j = 1,\ldots,p)$ are distinct prior probability densities. Hence, (24) is maximized for every single component by

$$max_{\theta_j} \sum_{i=1}^{n} \tau_j(y_i; x_i, \Theta^{(l)})\ log\ log\ \varphi_j(y_i; x_i; \theta_j), \qquad for\ j = 1,..,p \ldots (25)$$

where n_j is the jth cluster sample size and $n_1 + n_2 + \cdots + n_k = n$. Indeed, object function (25) is a K-means algorithm that satisfies the convergence in a finite number of iterations.

4. Simulation

We applied the three methods to generate data in order to compare their results. The comparison was conducted based on mean square error (MSE_j) of the fitted model for each m generated sample (Wackerly et al.,2008)[24]. Afterward, we average all calculated MSE_j, as in the following formulas:

$$MSE_j = \frac{\sum_{j=1}^{n} (y_i - \hat{y}_i)^2}{n}, for\ j = 1, 2, \ldots, m$$

$$AMSE = \underline{MSE} = \frac{\sum_{j=1}^{m} MSE_j}{m}$$

We also used another measure which is called classification accuracy(CA) as it is proposed by (Kassambara 2017)[11] and take the average of all calculated CA's over m replications as it is shown below:

$$ACA = \frac{1}{m} \sum_{j=1}^{m} \frac{Number\ of\ truely\ classified\ items\ in\ j^{th} sample}{n}$$

If the ACA value is close to 1, that means the classification technique is adequate. Data was generated according to the below steps:

1-We set $\theta = (n, \pi_1, \sigma_1, \sigma_2, \beta_1, \beta_2)$, where $\beta_1 = (p_{11}, p_{12}, p_{13}), \beta_2 = (p_{21}, p_{22}, p_{23})$
2- We generate $X_1 \sim N(0,1)$, $X_2 \sim N(0,1)$
3- We then generate $Y_i \sim N(X\beta_j, {\sigma_j}^2)$, where $X = (X_1, X_2)$ depending on membership of observation i.

We choose three different scenarios for θ. All these values are chosen randomly without restriction.
Scenario 1: $\theta_1 = [70, 0.2, 0.8, 0.8, (7,5,4), (1,2,3)]$
Scenario 2: $\theta_2 = [100, 0.3, 1, 1, (5,3,2), (2,1.5,1)]$
Scenario 3: $\theta_3 = [120, 0.4, 1.2, 1.2, (9,7,5), (3,2,1)]$

Several R packages are available to apply these methods. In this study, we used *RobMixReg* and *flex mix* to analyze the generated data. These packages are designed to choose the initial values of EM algorithms randomly and get them updated iteration by iteration.

Note that for simplicity, we consider the case of having a two components mixture model. However, one can follow the same steps mentioned in this paper to consider a high number of components. We plotted the scatters of Y and X of each scenario for the three methods. The tables and the graphs of the results are shown in the next section.

5. Results and Discussion

After applying the three methods to the generated data and replicating the process for m=500 for the three scenarios, we end up with the mean of estimators and the value of AMSE and ACA. The results are summarized in the below tables and figures. It can be seen from Table 1 and Figure 1 that the flexbox method outperformed the others in terms of AMSE and ACA under all the considered scenarios. In the plots above, the empty circles are for component 1, and the filled circles are for component 2. Whereas the blue color refers to the true values of Y and the red color refers to the estimated values of Y according to the estimated parameters. From the tables, it can be seen that the mean of estimators over all the replications is close to the true values of the parameters that have been used in the simulations. The value of AMSE is smaller in the FlexMix method than the others, and the value of ACA is higher in the FlexMix method. From the figures, it can be seen that how the estimated response variable values are close to the real ones that we generated under the three scenarios.

Table 1: The mean of estimators for 500 replications with AMSE and ACE for data generated from θ_1.

	π_1	σ_1	σ_1	p_{11}	p_{12}	p_{13}	p_{21}	p_{22}	p_{23}	$AMSE$	ACA
FlexMIX	0.189	0.67	0.806	6.938	5.08	3.89	0.974	1.966	2.94	0.3411	0.943
MixTLE	0.172	0.58	0.73	6.958	5.056	3.87	0.980	1.967	2.93	0.3442	0.910
MixLP	0.199	1.44	0.90	6.314	4.988	3.76	0.995	1.931	2.92	0.3433	0.90

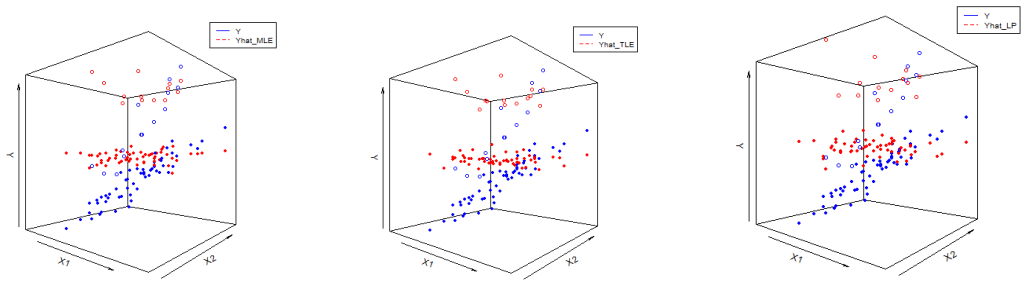

Figure 1: Represents the scatters of Y with respect to X_1 ,X_2 in blue color and $\widehat{Y}$ with respect to X_1 ,X_2in red color of the methods FlexMix, MixTLE, MixLP from left to right for data generated from scenario 1.

Table 2: The mean of estimators for 500 replications with AMSE and ACA for data generated from θ_2.

	π_1	σ_1	σ_1	p_{11}	p_{12}	p_{13}	p_{21}	p_{22}	p_{23}	$AMSE$	ACA
FlexMIX	0.276	0.910	0.995	4.887	3.125	2.022	2.010	1.452	1.025	0.239	0.948
MixTLE	0.293	0.644	0.905	4.807	3.169	2.030	2.044	1.442	1.034	0.269	0.932
MixLP	0.334	1.018	1.150	4.582	3.171	1.999	2.070	1.479	1.022	0.247	0.917

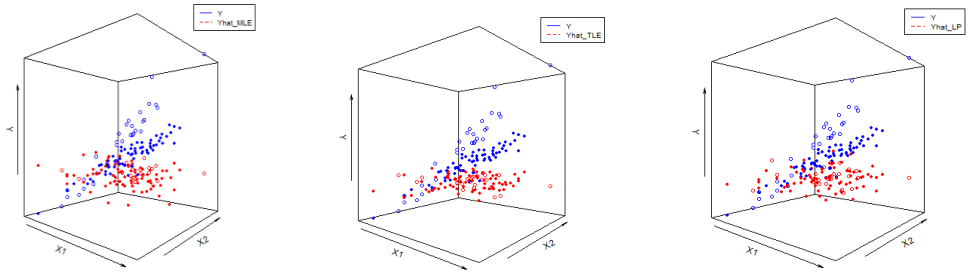

Figure 2: Represents the scatters of Y concerning X_1 ,X_2 in blue color and $\widehat{Y}$ concerning X_1 ,X_2in the red color of FlexMix, MixTLE, MixLP from left to right for data generated from scenario 2.

Table 3: The mean of estimators for 500 replications with AMSE and ACA for data generated from θ_3.

	π_1	σ_1	σ_1	p_{11}	p_{12}	p_{13}	p_{21}	p_{22}	p_{23}	$AMSE$	ACA
FlexMIX mixture	0.372	1.172	1.209	9.057	6.950	4.998	2.982				
	2.028	1.036	0.138	0.970							
MixLP	0.385	0.953	1.073	9.111	6.931	4.993	3.023				
	2.033	1.020	0.156	0.947							
	0.397	1.309	1.358	9.035	6.960	4.981	2.964				
	2.000	1.067	0.151	0.928							

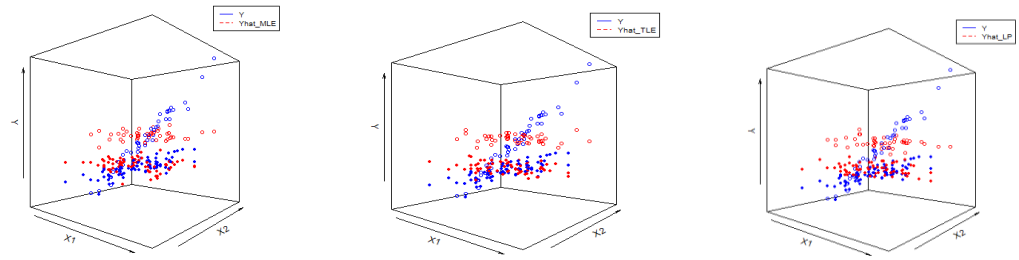

Figure 3: Represents the scatters of Y concerning X_1 ,X_2 in blue color and $\widehat{Y}$ concerning X_1 ,X_2in the red color of FlexMix, MixTLE, MixLP from left to right for data generated from scenario 3.

6. Conclusion

In this research, we compared three essential methods used in fitting the mixture of regression models. The methods were built under different procedures that can result in different results. The FlexMix was fitted through maximum likelihood and EM algorithm. The mixture was fitted by the use of trimmed likelihood after removing h of the high leverage values. MixLP was fitted under the assumption that the error follows Laplace distribution. The results of all methods are seen to be acceptable as far as accuracy is concerned. It is essential here to mention that the flexbox method showed relatively better results than the others. It can be seen that the FlexMix result in less AMSE than the others in all scenarios. The classification technique was accurate enough to classify observation into two groups, as shown in the graphs. The key advantage of the classification procedure is to help researchers in clustering data into several components based on similarities amongst them.

References

1.Andrews, D.F. and Mallows, C.L. (1974)," Scale mixtures of normal distributions", J.R.S.S.(B),Vol.36,No.1,pp.99-102.

2.Bai, X., Yao, W., & Boyer, J. E. (2012), "Robust fitting of mixture regression models", ***Computational Statistics & Data Analysis*****,Vol. 56, No. 7,pp. 2347-2359.**

3.Böhning, D. (2000), " Computer-Assisted Analysis of Mixtures and Applications: Meta-analysis", Disease Mapping and Others, Chapman & Hall, New York.

4.Frühwirth-Schnatter, S.(2006), ***"Finite mixture and Markov switching models"*****, Springer Series in Statistics.**

5.García-Escudero, L. A., Gordaliza, A., San Martin, R., Van Aelst , S., & Zamar, R. (2009), "Robust linear clustering", ***Journal of the Royal*** **Statistical Society:** ***Series B (Statistical Methodology)*****,Vol.71, No.1, pp. 301-318.**

6.García-Escudero, L. A., Gordaliza, A., Mayo-Íscar, A., & San Martín, R. (2010), "Robust clusterwise linear regression through trimming", ***Computational Statistics & Data Analysis*****, Vol. 54, No. 12, pp. 3057-3069.**

7.Hadi, A. S., & Luceño, A. (1997), "Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms", Computational Statistics & Data Analysis, Vol. 25, No. 3, pp. 251-272.

8.Hawkins, D. S., Allen, D. M., & Stromberg, A. J. (2001), "Determining the number of components in mixtures of linear models". ***Computational Statistics & Data Analysis*****,Vol. 38, No. 1,pp. 15-48.**

9.Hennig, C. (2002), "Fixed point clusters for linear regression: computation and comparison", ***Journal of classification*****, Vol. 19 , No. 2, pp. 249-276.**

10. Hennig, C. (2003), "Clusters, outliers, and regression: fixed. point clusters", ***Journal of Multivariate Analysis*****, Vol. 86, No. 1, pp. 183-212.**

11. Kassambara, A. (2017), ***Machine Learning Essentials: Practical Guide in R*****. STHDA**

12. Lindsay, B. G. (1995), " Mixture models: theory, geometry and applications", In ***NSF-CBMS regional conference series in probability and statistics*****, Institute of Mathematical Statistics and the American Statistical Association,pp. i-163.**

13. Markatou, M. (2000), "Mixture models, robustness, and the weighted likelihood methodology", ***Biometrics*****, Vol. 56, No. 2, pp. 483-486.**

14. McLachlan, G.J., Peel, D. (2000), "Finite mixture models", New York.
15. McLachlan, G. J., & Krishnan, T (2007)," ***The EM algorithm and extensions*** **", John Wiley & Sons, Vol. 382.**
16. Müller, C. H., & Garlipp, T. (2005), "Simple consistent cluster methods based on redescending M-estimators with an application to edge identification in images", Journal of Multivariate Analysis , Vol.92 , No. 2,pp. 359-385.
17. Neykov, N., Filzmoser, P., Dimova, R., & Neytchev, P. (2007), "Robust fitting of mixtures using the trimmed likelihood estimator", ***Computational Statistics & Data Analysis*** **, Vol. 52, No. 1,pp. 299-308.**
18. Phillips, R. F. (2002)," Least absolute deviations estimation via the EM algorithm", ***Statistics and Computing*****, Vol. 12 No. 3,pp.281-285.**
19. Quandt, R. E., & Ramsey, J. B . (1978), "Estimating mixtures of normal distributions and switching regressions", ***Journal of the American statistical Association*** **,Vol. 73, No. 364 ,pp. 730-738.**
20. Shen, H. B., Yang, J., & Wang, S. T. (2004, September), "Outlier detecting in fuzzy switching regression models", In ***International Conference on Artificial Intelligence: Methodology, Systems, and Applications*****, Springer , Berlin, Heidelberg, pp. 208-215.**
21. Song, W., Yao, W., & Xing, Y. (2014), "Robust mixture regression model fitting by Laplace distribution". ***Computational Statistics & Data Analysis*****, Vol. 71, pp.128-137.**
22. Turner, T. R. (2000), "Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions", ***Journal of the Royal Statistical Society: Series C (Applied Statistics)*****, Vol. 49, No.3 ,pp. 371-384.**
23. Vandev, D. L., & Neykov*, N. M. (1998), "About regression estimators with high breakdown point", ***Statistics: A Journal of Theoretical and Applied Statistics*** **, Vol. 32 ,No. 2,pp. 111-129.**
24. Wackerly, Dennis; Mendenhall, William; Scheaffer, Richard L. (2008). ***Mathematical Statistics with Applications*** **(7 ed.). Belmont, CA, USA: Thomson Higher Education. ISBN 978-0-495-38508-0.**

مقارنة بعض طرق تقدير أنموذج الانحدار الخطي المختلط مع التطبيق

الباحث/ ئورده ك إبراهيم كريم
قسم الاحصاء ، كلية الإدارة والاقتصاد
جامعة بغداد ، بغداد ، العراق
Urdak124ibrahim@gmail.com

م.د. فضاء مزهر هاشم
قسم الاحصاء ، كلية الإدارة والاقتصاد
جامعة بغداد ، بغداد ، العراق
Fali197477@gmail.com

Received:25/1/2021 Accepted: 15/3/2021 Published: SEPTEMBER/ 2021

المستخلص:

يوظف الأنموذج المختلط لنمذجة المشاهدات التي تعود الى أكثر من انموذج(مركبة) بمعالم مختلفة. أن هذا النوع من النماذج أصبح من الطرائق المهمة في الاستدلال الاحصائي لكثير من البيانات ذات التركيب المعقد التي من الممكن ان نوجهها عندما نروم انجاز التحليل الاحصائي. أضافة الى ذلك ، يعتبر هذا الانموذج أسلوب متقن بشكل عالي في تصنيف البيانات الى عدد من المجاميع اعتمادا على بعض الخصائص المتشابهة بينها. أحد أنواع النماذج المختلطة هو انموذج الانحدار الخطي المختلط الذي يستعمل عندما تكون البيانات تمثل أكثر من خط انحدار واحد. في هذا البحث كان الهدف هو مقارنة ثلاث طرق مهمة يمكن استعمالها مع هكذا نوع من النماذج هي طريقة FlexMix, MixTLE, MixLP . حيث تم تطبيق هذا الطرق على بيانات محاكاة افتراضية لثلاث حالات مختلفة وتم التوصل الة مقدرات دقيقة بنسب عالية بالإضافة الى اثبات كفاءة هذه الطرق في تصنيف البيانات الى مجاميعها الافتراضية بدقة عالية الا ان طريقة FlexMix اثبتت كفاءتها في الحصول على مقدرات دقيقة بشكل أفضل من الطريقتين الأخريين. أجريت المقارنة بين الطرق الثلاث باستعمال مربع الخطأ التكاملي (IMSE) وخطأ التصنيف التكاملي (ICE).

المصطلحات الرئيسة للبحث/ الأنموذج المختلط ، خوارزمية تعظيم التوقع، الانحدار الخطي، الإمكان الأعظم المشذب، توزيع لابلاس

***البحث مستل من رسالة ماجستير**