6.1 INTRODUCTION

Mining of time series data is most important in many practical situations arising at places like medical diagnostics, pollutant concentrate analysis, portfolio management, meteorological studies etc. The approaches to clustering time series can be classified as model based or model free. Model based approaches assume some form of underline generating process, estimating the model parameters and then perform classification based on the sample information. The clustering and classification together is needed for efficient and effective mining applications. This combined process of mining often utilized the unsupervised learning algorithms. In unsupervised learning algorithm, the first step is to obtain the number of classes and cluster the whole data into those classes. Then the learning algorithm is developed by utilizing the model building techniques for proper identification of the data set to its associated class.

Several learning algorithms have been developed using mixture of ARMA models (Yimin Xiong and Dit-Yan Yeung (2004)). But in all these models, it is customary to consider that the attributes associated with the data set are independent and each attribute can be separately treated for analysis. But in many domains at places like industry, medicine, environment, economics etc., the associated variables are correlated and the pattern can be well characterized by utilizing delayed cause-effect dependencies. This has motivated to develop the learning algorithms using AR(1) model in chapter 3 and 4 and learning algorithm based on AR(p) model in
In chapter 5, it is assumed that the number of classes of the whole data which is characterized by mixture of bivariate autoregressive process of order $p$ is known.

But in meteorological studies to understand the dynamics of weather it is needed to analyze the time series data of the regions with the feature vector consisting of Temperature and Humidity. Hence, in these studies the whole data can be characterized by considering mixture of bivariate AR($p$) process. But the number of components in this mixture is unknown as a priori and it requires the regions are to be divided based on the data set of the bivariate feature vector. Hence in this chapter, an unsupervised learning algorithm is developed and analyzed using bivariate AR($p$) model and K-means algorithm. K-means algorithm is used to identify the number of regions according to feature vector $\begin{bmatrix} \text{Temperature} \\ \text{Humidity} \end{bmatrix}$.

Using the EM algorithm, the model parameters are estimated. The initial estimates of the parameters are obtained with the K-means algorithm and least squares estimates. The unsupervised learning algorithm is developed by utilizing the conditional maximum likelihood of each class. The performance of the developed algorithm is studied through computing the performance measures like sensitivity, specificity, precision, recall, F-measure and misclassification rate. A comparative study of the developed algorithm with existing univariate autoregressive process of order $p$ algorithm is also discussed. This algorithm includes the results of chapter 4 when $p=1$. 

126
6.2 K-MEANS ALGORITHM

Let the time series are generated by M different bivariate AR(p) models which corresponds to M clusters of interest with weights $P(o_1), P(o_2), ..., P(o_M)$. Let $P(Z_i|o_k, \Phi_k)$ denote the conditional likelihood function of $k^{th}$ cluster with $\Phi_k$ as set of parameters for the model. Then the conditional likelihood function of mixture of finite bivariate AR process of order $p$ can be expressed in the form

$$P(Z_i|\Theta) = \sum_{k=1}^{M} P(Z_i|o_k, \Phi_k) P(o_k) \quad \text{(6.2.1)}$$

where, $\Theta = \{\Phi_1, \Phi_2, ..., \Phi_M, P(o_1), (o_2), ... P(o_M)\}$ represents the set of model parameters for mixture model.

A bivariate time series $Z_t$ is assigned to cluster $o_k$ with posterior probability $P(o_k|Z_t)$ such that $\sum_{k=1}^{M} P(o_k) = 1$. This model includes mixture of univariate autoregressive process of order $p$ model when the model parameters $\phi^{(i)}_{ij}, \phi^{(j)}_{jj}$ and $\phi^{(j)}_{ij}$ are zero for $k=1,2,...,M$. If $M=1$, then this process becomes Bivariate AR(p) process. The feature vector representing $\Phi = (\phi^{(i)}_{11}, \phi^{(i)}_{12}, \phi^{(j)}_{21}, \phi^{(j)}_{22})$; $j=1,2,...,p$ for each bivariate time series is obtained through using least square method of estimation (Douglas C. Montgomery, Lynwood A. Johnson and John S. Gardiner (1990)).

The K-means algorithm is a faster method to perform clustering of time series data. In K-means algorithm, continuous reassignments of objects into different clusters are done so that within cluster distance is minimized. For utilizing the K-
means algorithm, each time series is reduced to the object vector
\[ \Phi_i = \left\{ (\phi_1^{(i)}, \phi_2^{(i)}, \phi_3^{(i)}), \ldots, (\phi_{2p}^{(i)}) \right\}, \]
the parameters of bivariate AR(p) which are obtained using ordinary least square estimates of the model. If \( C_k \) is the center of the \( k \)th cluster then K-means algorithm attempts to minimize the objective function

\[ F = \sum_{t=1}^{N} \sum_{k=1}^{K} (\Phi_{tk} - C_k)^2 \]

The K-means algorithm also requires initial number of classes, which can be obtained by plotting a bivariate scatter surface for all time series of training data. Based on the number of surfaces visualized, the initial value of \( M \) is obtained.

The K-means algorithm for obtaining the number of clusters and classes is as follows:

Step 1: Identify the value of \( K \).

Step 2: Initialize \( K \) cluster centers.

Step 3: Decide the class memberships of \( N \) objects by assigning them to nearest Cluster center.

Step 4: Re-estimate the \( K \) cluster centers by assuming the membership found above as correct.

Step 5: If none of the \( N \) objects changed membership in last iteration, then exit otherwise go to step 3.

For classifying the training time series data into \( M \) clusters the residual variances of each class, correlation coefficient between the attributes in each class and model parameters \( \phi_1^{(j)}, \phi_2^{(j)}, \phi_3^{(j)}, \phi_{2p}^{(j)} \) \( j = 1, 2, \ldots, p \) within each class are obtained using clustered time series of that class and taking average over the number of time series in that class.
6.3 ESTIMATION OF PARAMETERS THROUGH EM ALGORITHM

In this section, the estimation of parameters using Expectation-Maximization algorithm is presented. Here, without loss of generality, it is assumed that the time series values are standardized by taking each individual observation from their mean value and hence the mean function of each time series is considered as zero. Therefore the problem of estimation is to obtain the estimates of the parameters \( \sigma^2_x, \sigma^2_y, \rho, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}; j=1,2,3,...,p \). For estimating the parameters through EM algorithm, one can use likelihood function of the sample.

Let \( D = \{Z_1, Z_2, ..., Z_N\} \) are \( N \) time series and assuming that these are conditionally independent under the given model, the likelihood function of realization is

\[
P(D | \Theta) = \prod_{i=1}^{N} P(Z_i | \Theta) = \prod_{i=1}^{N} \sum_{k=1}^{M} P(Z_i | \omega_k, \Phi_k) P(\omega_k)
\]

The log likelihood function of the realization \( D \) is

\[
\ln P(D | \Theta) = \sum_{i=1}^{N} \ln \left[ \sum_{k=1}^{M} P(Z_i | \omega_k, \Phi_k) P(\omega_k) \right]
\]

In EM algorithm, the first step is to obtain the expectation of log likelihood function. Given the observed data set \( D \) and current parameter estimate \( \Theta(r) \), the expected value of log likelihood for \( M \) labels of time series each having \( n \) points of observations can be expressed as
\[ Q(\Theta | \Theta (r)) = E(\ln P(D | \Theta)) \]

\[
= \sum_{i=1}^{N} \sum_{k=1}^{M} P(\omega_k | Z_i, \Theta (r)) \ln P(Z_i | \omega_k, \Phi_k) + \sum_{i=1}^{N} \sum_{k=1}^{M} P(\omega_k | Z_i, \Theta (r)) \ln P(\omega_k)
\]

(6.3.1)

where, the posterior probabilities \( P(\omega_k | Z_i, \Theta) \) are computed using the Bayes rule as

\[
P(\omega_k | Z_i, \Theta) = \frac{P(Z_i | \omega_k, \Phi_k) P(\omega_k)}{\sum_{q=1}^{M} P(Z_i | \omega_q, \Phi_q) P(\omega_q)}, \quad t = 1, 2, \ldots, N \text{ and } k = 1, 2, \ldots, M.
\]

(6.3.2)

This implies, \( Q(\Theta | \Theta (r)) = \)

\[
+ \sum_{i=1}^{N} \sum_{k=1}^{M} P(\omega_k | Z_i, \Theta (r)) \ln \left\{ \frac{P(Z_i | \omega_k, \Phi_k) P(\omega_k)}{\sum_{q=1}^{M} P(Z_i | \omega_q, \Phi_q) P(\omega_q)} \right\}
\]

(6.3.3)

In the M- step of EM algorithm, one has to find estimates of parameters by maximizing \( Q(\Theta | \Theta (r)) \).

The updated equation \( P(\hat{\omega}_k) \) of the model is

\[
P(\hat{\omega}_k) = \frac{1}{N} \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta (r)) \quad k = 1, 2, \ldots, M
\]

(6.3.4)

For estimating \( \sigma^2_{\epsilon_{11}}, \sigma^2_{\epsilon_{22}}, \phi^{(j)}_{11k}, \phi^{(j)}_{22k}, \phi^{(j)}_{12k}, \phi^{(j)}_{21k}, \) \( j = 1, 2, \ldots, p \) and \( \rho_k \) the updated equations respectively are
\[
\sum_{i=1}^{N} P(\omega_i | Z_t, \Theta(r)) \left[ (n-p) - \frac{1}{1-\rho_k^2} \sum_{u=p+1}^{n} \left( \frac{X_{i,u} - \sum_{j=1}^{p} \phi_{11k}^{(j)} X_{t,u-(p-j)}}{\sigma_{eq_k}} - \frac{\sum_{j=1}^{p} \phi_{12k}^{(j)} Y_{t,u-(p-j)}}{\sigma_{eq_k}} \right)^2 \right] = 0
\]

(6.3.5)

\[
\sum_{i=1}^{N} P(\omega_i | Z_t, \Theta(r)) \left[ (n-p) - \frac{1}{1-\rho_k^2} \sum_{u=p+1}^{n} \left( \frac{Y_{i,u} - \sum_{j=1}^{p} \phi_{11k}^{(j)} X_{t,u-(p-j)}}{\sigma_{eq_k}} - \frac{\sum_{j=1}^{p} \phi_{12k}^{(j)} Y_{t,u-(p-j)}}{\sigma_{eq_k}} \right)^2 \right] = 0
\]

(6.3.6)

\[
\sum_{i=1}^{N} P(\omega_i | Z_t, \Theta(r)) \sum_{u=p+1}^{n} \left[ \frac{1}{\sigma_{eq_k}} X_{t,u-(p-j)} Y_{t,u} - \frac{\rho_k}{\sigma_{eq_k}} X_{t,u-(p-j)} Y_{t,u} \right] = 0
\]

(6.3.7)

\[
\sum_{i=1}^{N} P(\omega_i | Z_t, \Theta(r)) \sum_{u=p+1}^{n} \left[ \frac{1}{\sigma_{eq_k}} X_{t,u-(p-j)} Y_{t,u} - \frac{\rho_k}{\sigma_{eq_k}} X_{t,u-(p-j)} Y_{t,u} \right] = 0
\]

(6.3.8)
\[
\sum_{j=1}^{N} P(\omega_k | Z_t, \Theta(r)) \sum_{\omega=1}^{n} \left\{ \frac{\rho_{k}}{\sigma_{\alpha_k}} X_{t,u-(p-j+1)} Y_{t,u} - \frac{1}{\sigma_{\alpha_k}} Y_{t,u-(p-j+1)} \right\} = 0
\]

(6.3.9)

\[
\sum_{j=1}^{N} P(\omega_k | Z_t, \Theta(r)) \sum_{\omega=1}^{n} \left\{ \frac{\rho_{k}}{\sigma_{\alpha_k}} X_{t,u-(p-j+1)} Y_{t,u} - \frac{1}{\sigma_{\alpha_k}} Y_{t,u-(p-j+1)} \right\} = 0
\]

(6.3.10)

\[
\sum_{j=1}^{N} P(\omega_k | Z_t, \Theta(r)) \left\{ (n-p) \rho_k - \frac{\rho_k}{(1-\rho_k)} \sum_{\omega=1}^{n} \left\{ \frac{X_{t,u} - \sum_{j=1}^{p} \phi_{11k} X_{t,u-(p-j+1)} - \sum_{j=1}^{p} \phi_{22k} Y_{t,u-(p-j+1)}}{\sigma_{\alpha_k}^2} \right\} \right\}^2
\]

\[
+ \frac{\left[ Y_{t,u} - \sum_{j=1}^{p} \phi_{11k} X_{t,u-(p-j+1)} - \sum_{j=1}^{p} \phi_{22k} Y_{t,u-(p-j+1)} \right]^2}{\sigma_{\alpha_k}^2}
\]

\[
\left( 1+\rho_k^2 \right) \sum_{\omega=1}^{n} \left\{ \frac{X_{t,u} - \sum_{j=1}^{p} \phi_{11k} X_{t,u-(p-j+1)} - \sum_{j=1}^{p} \phi_{22k} Y_{t,u-(p-j+1)}}{\sigma_{\alpha_k}} \right\}^2
\]

\[
\left( 1-\rho_k^2 \right) \sum_{\omega=1}^{n} \left\{ \frac{Y_{t,u} - \sum_{j=1}^{p} \phi_{11k} X_{t,u-(p-j+1)} - \sum_{j=1}^{p} \phi_{22k} Y_{t,u-(p-j+1)}}{\sigma_{\alpha_k}} \right\}^2 = 0
\]

(6.3.11)
Solving the equations (6.3.4), (6.3.5), (6.3.6), (6.3.7), (6.3.8), (6.3.9), (6.3.10) and (6.3.11) simultaneously and iteratively the refined estimates of model parameters $P(\omega_k), \sigma^2_{e_k}, \sigma^2_{\epsilon_k}, \phi^{(l)}_{11k}, \phi^{(l)}_{12k}, \phi^{(l)}_{21k}$ and $\phi^{(l)}_{22k}, j = 1, 2, 3, \ldots, p$ and $\rho_k$ can be obtained.

**Expectation Maximization Algorithm**

Step 1: Find the initial parameters using equations (6.4.1) to (6.4.3) given in section (6.4).

Step 2: Obtain revised estimates of the parameters

$$P(\omega_k), \sigma^2_{e_k}, \sigma^2_{\epsilon_k}, \phi^{(l)}_{11k}, \phi^{(l)}_{12k}, \phi^{(l)}_{21k}, \phi^{(l)}_{22k}, j = 1, 2, 3, \ldots, p$$

and $\rho_k$ using equations (6.3.4), (6.3.5), (6.3.6), (6.3.7), (6.3.8), (6.3.9), (6.3.10) and (6.3.11).

Step 3: Repeat the process until the parameters do not change or the difference in successive computations is within given threshold value.

Step 4: Write final estimates of parameters

$$P(\omega_k), \sigma^2_{e_k}, \sigma^2_{\epsilon_k}, \phi^{(l)}_{11k}, \phi^{(l)}_{12k}, \phi^{(l)}_{21k}, \phi^{(l)}_{22k}, j = 1, 2, 3, \ldots, p$$ and $\rho_k$.

**6.4 INITIAlISATION OF PARAMETERS**

To initialize the parameters $\sigma^2_{e_k}, \sigma^2_{\epsilon_k}$ and $\rho_k$ using K-means algorithm one can obtain the number of classes (clusters) of the training data set. After classifying the time series data into $M$ clusters, the residual variance of each attribute for each class, the correlation coefficient between the attributes in each class and the model parameters with in the class are obtained using the clustered time series of that class.

Thus the initial estimates of $\sigma^2_{e_k}$ and $\sigma^2_{\epsilon_k}$ can be taken as
\[ \hat{\sigma}_{\varepsilon_k}^2 = \frac{1}{(n-p)G_k} \sum_{g=1}^{G_k} \sum_{u=2}^{n} \varepsilon_{\tau_{gu}}^2 \quad \text{and} \quad \hat{\sigma}_{\varepsilon_k}^2 = \frac{1}{(n-p)G_k} \sum_{g=1}^{G_k} \sum_{u=2}^{n} \varepsilon_{\tau_{gu}}^2, \quad \text{for} \quad k=1,2,\ldots,M \]

(6.4.1)

where, \( G_k \) is the number of time series in the \( k \)th cluster.

The initial estimate of \( \rho_k \), the correlation coefficient between \( X_t \) and \( Y_t \), is given by

\[ \hat{\rho}_k = \frac{1}{G_k} \sum_{g=1}^{G_k} \frac{\text{Cov}(x)(X_t,Y_t)}{\sqrt{\text{Var}(x)(X_t)\text{Var}(x)(Y_t)}}, \quad k=1,2,\ldots,M \]

(6.4.2)

where, \( G_k \) is the number of time series in the \( k \)th cluster.

For initializing parameters \( \phi_{11}^{(j)}, \phi_{21}^{(j)}, \phi_{21}^{(j)} \) and \( \phi_{22}^{(j)}, j=1,2,3,\ldots,p \), ordinary least squares estimates of the bivariate autoregressive process of order \( p \) are obtained for each time series data in each class and averaged over the number of time series in that class. The sample realization of bivariate autoregressive process of order \( p \) can be written in the following format

\[
\begin{bmatrix}
X_{t,p+1} & Y_{t,p+1} \\
X_{t,p+2} & Y_{t,p+2} \\
\vdots & \vdots \\
X_{t,n} & Y_{t,n}
\end{bmatrix}
= \begin{bmatrix}
X_{t,1} & Y_{t,1} & X_{t,2} & Y_{t,2} & \cdots & X_{t,p} & Y_{t,p} \\
X_{t,2} & Y_{t,2} & X_{t,3} & Y_{t,3} & \cdots & X_{t,p+1} & Y_{t,p+1} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
X_{t,n-p} & Y_{t,n-p} & X_{t,n-(p-1)} & Y_{t,n-(p-1)} & \cdots & X_{t,n-1} & Y_{t,n-1} \\
\end{bmatrix}
\begin{bmatrix}
\phi_{11}^{(1)} & \phi_{21}^{(1)} \\
\phi_{12}^{(1)} & \phi_{22}^{(1)} \\
\phi_{11}^{(2)} & \phi_{21}^{(2)} \\
\phi_{12}^{(2)} & \phi_{22}^{(2)} \\
\phi_{11}^{(p)} & \phi_{21}^{(p)} \\
\phi_{12}^{(p)} & \phi_{22}^{(p)} \\
\end{bmatrix}
\]

134
which can be represented as

\[ Y_t = X_t \Phi_t + \xi_t \]

where,

\[
Y_t = \begin{bmatrix} X_{t,p+1} & Y_{t,p+1} \\ X_{t,p+2} & Y_{t,p+2} \\ \vdots & \vdots \\ X_{t,n} & Y_{t,n} \end{bmatrix}, \quad X_t = \begin{bmatrix} X_{t,1} & Y_{t,1} & X_{t,2} & Y_{t,2} & \cdots & X_{t,p} & Y_{t,p} \\ X_{t,2} & Y_{t,2} & X_{t,3} & Y_{t,3} & \cdots & X_{t,p+1} & Y_{t,p+1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{t,n-p} & Y_{t,n-p} & X_{t,n-(p-1)} & Y_{t,n-(p-1)} & \cdots & X_{t,n-1} & Y_{t,n-1} \end{bmatrix}
\]

\[
\Phi_t = \begin{bmatrix} \phi_{1t}^{(1)} & \phi_{21t}^{(1)} \\ \phi_{12t}^{(1)} & \phi_{22t}^{(1)} \\ \phi_{1t}^{(2)} & \phi_{21t}^{(2)} \\ \phi_{12t}^{(2)} & \phi_{22t}^{(2)} \\ \phi_{1t}^{(p)} & \phi_{21t}^{(p)} \\ \phi_{12t}^{(p)} & \phi_{22t}^{(p)} \end{bmatrix}, \quad \xi_t = \begin{bmatrix} e_{X_{t,p+1}} & e_{Y_{t,p+1}} \\ e_{X_{t,p+2}} & e_{Y_{t,p+2}} \\ \vdots & \vdots \\ e_{X_{t,n}} & e_{Y_{t,n}} \end{bmatrix}
\]

Then the ordinary least square estimate of \( \Phi_t \) is

\[
\hat{\Phi}_t = (X_t^T X_t)^{-1} X_t^T Y_t \quad \text{for } t=1,2, \ldots, N \quad (6.4.3)
\]
6.5 UNSUPERVISED LEARNING ALGORITHM

Here, the unsupervised learning algorithm is presented for identifying the new time series data with one of the available clusters. The steps involved in this algorithm are as follows:

Step 1: Draw scatter surface diagram for training data set in order to obtain the initial number of clusters for using the K-means algorithm.

Step 2: Using K means algorithm obtain the refined number of classes and elements in each cluster.

Step 3: Obtain the initial estimates of model parameters using bivariate AR(p) process and the least square method of estimation given in section 6.4.

Step 4: Obtain the refined estimates of model parameters using updated equations of EM algorithm for bivariate AR(p) process given in section 6.3.

Step 5: For new time series bivariate data set, compute the conditional likelihood with model parameters of \( i^{th} \) class derived from step 4 over 1 to \( K \) and assign it to a class for which the component conditional likelihood is maximum.

6.6 EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

In this section, the utility of developed algorithm for clustering regions is demonstrated. After discussion with the people in Cyclone Warning Centre at Visakhapatnam, it is understood that 2 variables Temperature and Humidity levels are most important for taking control measures of weather forecast of a region. To understand the dynamics of the weather in the regions, it is required to segment the
total area into various regions based on the vector $\begin{bmatrix} \text{Temperature} \\ \text{Humidity} \end{bmatrix}$. The number of regions in an area is not known as a priori and requires unsupervised learning methods to identify a sample time series with the region and analysis. Hence a study is carried out by collecting the bivariate variable $\begin{bmatrix} \text{Temperature} \\ \text{Humidity} \end{bmatrix}$ over a period with a sample collected over 100 locations covering the whole area.

Using the K-means algorithm, the number of regions is determined. For implementing K-means algorithm, the initial number of regions is required. Using training data the bivariate time series are plotted in scatter responses through a 3-dimensional graph shown in figure 6.1.

**Figure 6.1**

Scatter plot for Temperature and Humidity against time
From figure 6.1, it is observed that the time series are stationary after obtaining deviation from mean values and it is also observed that there are 3 surfaces which represent 3 regions. Taking the initial value of number of clusters as 3, the K-means algorithm is implemented on the training dataset of 100 time series found that the number of regions in that area also represent 3 levels. The whole training data is classified into 3 segments representing 3 regions.

For developing BAR(2) model for each region, one can obtain the updated equations of the model parameters for EM algorithm. Taking \( p=2 \) in equations (5.3.11) to (5.3.20) one can obtain the updated equations of the model parameters of 

\[
\begin{align*}
\sigma^2_{\varepsilon_k}, \sigma^2_{\varepsilon_k}, \phi^{(1)}_{11k}, \phi^{(1)}_{12k}, \phi^{(1)}_{21k}, \phi^{(1)}_{22k}, \phi^{(2)}_{11k}, \phi^{(2)}_{12k}, \phi^{(2)}_{21k}, \phi^{(2)}_{22k} \text{ and } \rho_k \text{ respectively as}
\end{align*}
\]

\[
\sum_{i=1}^{N} P(\omega_k|Z_t, \Theta(r)) \left\{ (n-2) - \frac{1}{1-\rho_k^2} \sum_{s=3}^{3} \frac{X^2_{t,u}}{\sigma^2_{\varepsilon_k}} + \frac{\left(\phi^{(1)}_{11k}\phi^{(1)}_{21k}\phi^{(1)}_{22k}\right)}{\sigma^2_{\varepsilon_k} \sigma_{\varepsilon_k} \sigma_{\varepsilon_k}} \right\} X^2_{t,u-1}\n\]

\[
+ \frac{\left(\phi^{(2)}_{11k}\right)^2}{\sigma^2_{\varepsilon_k}} - \frac{\rho_k \phi^{(2)}_{12k} \phi^{(2)}_{21k} \phi^{(2)}_{22k}}{\sigma^2_{\varepsilon_k} \sigma_{\varepsilon_k} \sigma_{\varepsilon_k}} X^2_{t,u-2} + \frac{\left(\phi^{(1)}_{11k}\right)^2}{\sigma^2_{\varepsilon_k}} - \frac{\rho_k \phi^{(1)}_{12k} \phi^{(1)}_{21k} \phi^{(1)}_{22k}}{\sigma^2_{\varepsilon_k} \sigma_{\varepsilon_k} \sigma_{\varepsilon_k}} Y^2_{t,u-1} + \frac{\left(\phi^{(2)}_{11k}\right)^2}{\sigma^2_{\varepsilon_k}} - \frac{\rho_k \phi^{(2)}_{12k} \phi^{(2)}_{21k} \phi^{(2)}_{22k}}{\sigma^2_{\varepsilon_k} \sigma_{\varepsilon_k} \sigma_{\varepsilon_k}} Y^2_{t,u-2}\n\]

\[
- \frac{2\phi^{(1)}_{11k}}{\sigma^2_{\varepsilon_k}} - \frac{\rho_k \phi^{(1)}_{12k} \phi^{(1)}_{21k} \phi^{(1)}_{22k}}{\sigma^2_{\varepsilon_k} \sigma_{\varepsilon_k} \sigma_{\varepsilon_k}} X_{t,u-1} X_{t,u} - \frac{2\phi^{(2)}_{11k}}{\sigma^2_{\varepsilon_k}} - \frac{\rho_k \phi^{(2)}_{12k} \phi^{(2)}_{21k} \phi^{(2)}_{22k}}{\sigma^2_{\varepsilon_k} \sigma_{\varepsilon_k} \sigma_{\varepsilon_k}} X_{t,u-2} X_{t,u}\n\]

\[
+ \frac{2\phi^{(1)}_{11k} \phi^{(1)}_{21k}}{\sigma^2_{\varepsilon_k} \sigma_{\varepsilon_k} \sigma_{\varepsilon_k}} X_{t,u-1} X_{t,u-2} + \frac{2\phi^{(2)}_{11k} \phi^{(2)}_{21k}}{\sigma^2_{\varepsilon_k} \sigma_{\varepsilon_k} \sigma_{\varepsilon_k}} X_{t,u-1} X_{t,u}\n\]

\[
- \frac{2\phi^{(1)}_{12k}}{\sigma^2_{\varepsilon_k}} - \frac{\rho_k \phi^{(1)}_{11k} \phi^{(1)}_{21k} \phi^{(1)}_{22k}}{\sigma^2_{\varepsilon_k} \sigma_{\varepsilon_k} \sigma_{\varepsilon_k}} X_{t,u} Y_{t,u-1} - \frac{2\phi^{(2)}_{12k}}{\sigma^2_{\varepsilon_k}} - \frac{\rho_k \phi^{(2)}_{11k} \phi^{(2)}_{21k} \phi^{(2)}_{22k}}{\sigma^2_{\varepsilon_k} \sigma_{\varepsilon_k} \sigma_{\varepsilon_k}} X_{t,u} Y_{t,u-2}\n\]
\[\begin{align*}
+ \left( \frac{2\phi^{(1)}_{12k} \phi^{(2)}_{12k}}{\sigma^2_{\varepsilon_k}} - \rho_k \left( \phi^{(1)}_{12k} + \phi^{(2)}_{12k} \right) \right) X_{t,u-1} - Y_{t,u-1} + \left( \frac{2\phi^{(1)}_{12k} \phi^{(2)}_{12k}}{\sigma^2_{\varepsilon_k}} - \rho_k \left( \phi^{(1)}_{12k} + \phi^{(2)}_{12k} \right) \right) X_{t,u-2} - Y_{t,u-2} \\
+ \left( \frac{2\phi^{(1)}_{12k} \phi^{(2)}_{12k}}{\sigma^2_{\varepsilon_k}} - \rho_k \left( \phi^{(1)}_{12k} + \phi^{(2)}_{12k} \right) \right) X_{t,u-2} - Y_{t,u-2} + \left( \frac{2\phi^{(1)}_{12k} \phi^{(2)}_{12k}}{\sigma^2_{\varepsilon_k}} - \rho_k \left( \phi^{(1)}_{12k} + \phi^{(2)}_{12k} \right) \right) X_{t,u-2} - Y_{t,u-2} \\
- \frac{\rho_k}{\sigma_{\varepsilon_k}} \left( X_{t,u} - \phi^{(1)}_{12k} X_{t,u-1} - \phi^{(2)}_{12k} X_{t,u-2} - \phi^{(1)}_{12k} Y_{t,u-1} - \phi^{(2)}_{12k} Y_{t,u-2} \right) Y_{t,u} \right] = 0 \quad (6.6.1)
\end{align*}\]
\[\begin{align*}
&\left(2\eta_{21k}^{(2)}\varphi_{22k}^{(1)} + \frac{\rho_k}{\sigma_{s_k}} (\varphi_{11k}^{(2)} + \varphi_{12k}^{(2)} + \varphi_{21k}^{(2)} + \varphi_{22k}^{(2)})\right) X_{t,u-2} Y_{t,u-1} + \left(2\eta_{21k}^{(2)}\varphi_{22k}^{(1)} + \frac{\rho_k}{\sigma_{s_k}} (\varphi_{11k}^{(2)} + \varphi_{12k}^{(2)} + \varphi_{21k}^{(2)} + \varphi_{22k}^{(2)})\right) X_{t,u-2} Y_{t,u-1} \\
&\left(2\eta_{22k}^{(2)}\varphi_{22k}^{(2)} + \frac{\rho_k}{\sigma_{s_k}} (\varphi_{12k}^{(2)} + \varphi_{22k}^{(2)})\right) Y_{t,u-1} Y_{t,u-2} \\
&\frac{\rho_k}{\sigma_{s_k}} \left(Y_{t,u} - \varphi_{21k}^{(1)} X_{t,u-1} - \varphi_{21k}^{(2)} X_{t,u-2} - \varphi_{21k}^{(1)} Y_{t,u-1} - \varphi_{21k}^{(2)} Y_{t,u-2}\right) X_{t,u} = 0 \quad (6.6.2)
\end{align*}\]

\[
\sum_{u=1}^{\bar{u}} P(\omega_k | \Omega) \sum_{u=2} \left[\frac{\rho_k}{\sigma_{s_k}} Y_{t,u} \right] X_{t,u-1} - \left[\frac{\rho_k}{\sigma_{s_k}} Y_{t,u} \right] Y_{t,u-1} - \left[\frac{\rho_k}{\sigma_{s_k}} Y_{t,u-2} \right] X_{t,u-1} = 0 \quad (6.6.3)
\]

\[
\sum_{u=1}^{\bar{u}} P(\omega_k | \Omega) \sum_{u=2} \left[\frac{\rho_k}{\sigma_{s_k}} Y_{t,u} \right] Y_{t,u-1} - \left[\frac{\rho_k}{\sigma_{s_k}} Y_{t,u-2} \right] X_{t,u-1} = 0 \quad (6.6.4)
\]

\[
\sum_{u=1}^{\bar{u}} P(\omega_k | \Omega) \sum_{u=2} \left[\frac{\rho_k}{\sigma_{s_k}} X_{t,u} \right] X_{t,u-1} - \left[\frac{\rho_k}{\sigma_{s_k}} X_{t,u-2} \right] X_{t,u-1} = 0 \quad (6.6.5)
\]
\[ \sum_{r=1}^{N} P(\omega_k | Z_r, \Theta(r)) \sum_{u=3}^{n} \left[ \left( \frac{\rho_{\delta_{12k}}}{\sigma_{\delta_{xk}}} - \frac{\phi_{12k}}{\sigma_{\delta_{xk}}} \right) Y_{t,u-1} + \left( \frac{\rho_{\delta_{21k}}}{\sigma_{\delta_{xk}}} - \frac{\phi_{21k}}{\sigma_{\delta_{xk}}} \right) Y_{t,u-2} \right] = 0 \] (6.6.6)

\[ \sum_{r=1}^{N} P(\omega_k | Z_r, \Theta(r)) \sum_{u=3}^{n} \left[ \left( \frac{\phi_{12k}}{\sigma_{\delta_{xk}}} - \frac{\rho_{\delta_{12k}}}{\sigma_{\delta_{xk}}} \right) Y_{t,u-1} + \left( \frac{\phi_{21k}}{\sigma_{\delta_{xk}}} - \frac{\rho_{\delta_{21k}}}{\sigma_{\delta_{xk}}} \right) Y_{t,u-2} \right] = 0 \] (6.6.7)

\[ \sum_{r=1}^{N} P(\omega_k | Z_r, \Theta(r)) \sum_{u=3}^{n} \left[ \left( \frac{\phi_{12k}}{\sigma_{\delta_{xk}}} - \frac{\rho_{\delta_{12k}}}{\sigma_{\delta_{xk}}} \right) Y_{t,u-1} + \left( \frac{\phi_{21k}}{\sigma_{\delta_{xk}}} - \frac{\rho_{\delta_{21k}}}{\sigma_{\delta_{xk}}} \right) Y_{t,u-2} \right] = 0 \] (6.6.8)

\[ \sum_{r=1}^{N} P(\omega_k | Z_r, \Theta(r)) \sum_{u=3}^{n} \left[ \left( \frac{\rho_{\delta_{12k}}}{\sigma_{\delta_{xk}}} - \frac{\phi_{12k}}{\sigma_{\delta_{xk}}} \right) Y_{t,u-1} + \left( \frac{\rho_{\delta_{21k}}}{\sigma_{\delta_{xk}}} - \frac{\phi_{21k}}{\sigma_{\delta_{xk}}} \right) Y_{t,u-2} \right] = 0 \] (6.6.9)
\[
\sum_{i=1}^{N} P(\omega_i | Z_i, \Theta(r)) \sum_{w=3}^{n} \left[ \frac{\rho_x}{\sigma_{\epsilon x}} X_{i,w} - \frac{Y_{i,w}}{\sigma_{\epsilon y}} \right] Y_{i,w-1} \left[ \frac{\rho_x \phi^{(1)}_{12k}}{\sigma_{\epsilon x}} - \frac{\phi^{(1)}_{12k}}{\sigma_{\epsilon y}} \right] Y_{i,w-2} \left[ \frac{\rho_x \phi^{(2)}_{22k}}{\sigma_{\epsilon x}} - \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right] Y_{i,w-1} = 0 \quad (6.6.10)
\]

\[
\sum_{i=1}^{N} P(\omega_i | Z_i, \Theta(r)) \left[ (n-2) - \frac{1}{1-\rho_x^2} \sum_{w=3}^{n} \left[ \frac{X_{i,w}^2}{\sigma_{\epsilon x}^2} + \frac{Y_{i,w}^2}{\sigma_{\epsilon y}^2} + \left( \frac{\phi^{(1)}_{12k}}{\sigma_{\epsilon x}} \right)^2 + \left( \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right)^2 \right] X_{i,w-1} + \left( \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right)^2 \right] X_{i,w-2}
\]

\[
-2 \left[ \left( \frac{\phi^{(1)}_{12k}}{\sigma_{\epsilon x}} \right)^2 + \left( \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right)^2 \right] X_{i,w-1} + \left[ \left( \frac{\phi^{(1)}_{12k}}{\sigma_{\epsilon x}} \right)^2 + \left( \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right)^2 \right] X_{i,w-2}
\]

\[
+ \left[ \left( \frac{\phi^{(1)}_{12k}}{\sigma_{\epsilon x}} \right)^2 + \left( \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right)^2 \right] \left( \frac{\phi^{(1)}_{21k}}{\sigma_{\epsilon x}} \right)^2 + \left( \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right)^2 \right] Y_{i,w}
\]

\[
+ \left( 1 + \frac{\rho_x^2}{1-\rho_x^2} \sum_{w=3}^{n} \left[ (Y_{i,w} - \phi^{(1)}_{12k} X_{i,w-1} + \phi^{(2)}_{22k} X_{i,w-2}) - \phi^{(1)}_{22k} X_{i,w-1} - \phi^{(2)}_{22k} X_{i,w-2} \right) X_{i,w}
\]

\[
- \left( \frac{\phi^{(1)}_{12k}}{\sigma_{\epsilon x}} \right)^2 + \left( \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right)^2 \right] X_{i,w-1}
\]

\[
+ \left( \frac{\phi^{(1)}_{12k}}{\sigma_{\epsilon x}} \right)^2 + \left( \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right)^2 \right] X_{i,w-2}
\]

\[
+ \left( \frac{\phi^{(1)}_{12k}}{\sigma_{\epsilon x}} \right)^2 + \left( \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right)^2 \right] Y_{i,w-1}
\]

\[
+ \left( \frac{\phi^{(1)}_{12k}}{\sigma_{\epsilon x}} \right)^2 + \left( \frac{\phi^{(2)}_{22k}}{\sigma_{\epsilon y}} \right)^2 \right] Y_{i,w-2}
\]

\[
= 0 \quad (6.6.11)
\]
Using the initialization of parameters discussed in section 6.4, the initial estimates of parameters $P(\omega_k)$, $\sigma^2_{y_k}$, $\sigma^2_{x_k}$, $\phi^{(j)}_{11k}$, $\phi^{(j)}_{12k}$, $\phi^{(j)}_{21k}$, $\phi^{(j)}_{22k}$ and $\rho_k$ are obtained for three regions. The computed initial estimates of the model parameters are presented in Table 6.1.

Table 6.1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cluster 1 (Region I)</th>
<th>Cluster 2 (Region II)</th>
<th>Cluster 3 (Region III)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\omega_k)$</td>
<td>0.3333</td>
<td>0.3333</td>
<td>0.3333</td>
</tr>
<tr>
<td>$\sigma^2_{x_k}$</td>
<td>13.5424</td>
<td>10.8860</td>
<td>13.6331</td>
</tr>
<tr>
<td>$\sigma^2_{y_k}$</td>
<td>15.8794</td>
<td>20.9672</td>
<td>9.3483</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.1726</td>
<td>0.1826</td>
<td>0.4143</td>
</tr>
<tr>
<td>$\phi^{(1)}_{11}$</td>
<td>-0.5576</td>
<td>-0.1566</td>
<td>0.2256</td>
</tr>
<tr>
<td>$\phi^{(1)}_{12}$</td>
<td>-0.0320</td>
<td>-0.0075</td>
<td>-0.544</td>
</tr>
<tr>
<td>$\phi^{(1)}_{21}$</td>
<td>-0.1198</td>
<td>0.0089</td>
<td>0.1110</td>
</tr>
<tr>
<td>$\phi^{(1)}_{22}$</td>
<td>-0.2174</td>
<td>0.0329</td>
<td>-0.0054</td>
</tr>
<tr>
<td>$\phi^{(2)}_{11}$</td>
<td>0.3700</td>
<td>0.3327</td>
<td>0.0266</td>
</tr>
<tr>
<td>$\phi^{(2)}_{12}$</td>
<td>-0.1738</td>
<td>0.2782</td>
<td>-0.5018</td>
</tr>
<tr>
<td>$\phi^{(2)}_{21}$</td>
<td>0.3572</td>
<td>0.0522</td>
<td>0.0382</td>
</tr>
<tr>
<td>$\phi^{(2)}_{22}$</td>
<td>0.2559</td>
<td>0.075</td>
<td>0.3254</td>
</tr>
</tbody>
</table>

Using the initial estimates of the parameters and EM algorithm, the refined estimates of parameters for each class are obtained and presented in Table 6.2.
Table 6.2

Final Estimates of the Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cluster 1 (Region I)</th>
<th>Cluster 1 (Region II)</th>
<th>Cluster 1 (Region III)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\omega_k)$</td>
<td>0.342</td>
<td>0.412</td>
<td>0.246</td>
</tr>
<tr>
<td>$\sigma_{e_x}^2$</td>
<td>12.243</td>
<td>9.584</td>
<td>14.6996</td>
</tr>
<tr>
<td>$\sigma_{e_y}^2$</td>
<td>8.9341</td>
<td>22.467</td>
<td>10.040</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.156</td>
<td>-0.05</td>
<td>0.024</td>
</tr>
<tr>
<td>$\phi_{11}^{(1)}$</td>
<td>-0.3891</td>
<td>0.036</td>
<td>-0.4053</td>
</tr>
<tr>
<td>$\phi_{12}^{(1)}$</td>
<td>0.0493</td>
<td>0.016</td>
<td>0.0531</td>
</tr>
<tr>
<td>$\phi_{21}^{(1)}$</td>
<td>-0.0294</td>
<td>0.142</td>
<td>-0.0295</td>
</tr>
<tr>
<td>$\phi_{22}^{(1)}$</td>
<td>-0.0925</td>
<td>0.148</td>
<td>-0.0925</td>
</tr>
<tr>
<td>$\phi_{11}^{(2)}$</td>
<td>-0.2965</td>
<td>-0.0009</td>
<td>-0.321</td>
</tr>
<tr>
<td>$\phi_{12}^{(2)}$</td>
<td>0.4321</td>
<td>0.009</td>
<td>0.4284</td>
</tr>
<tr>
<td>$\phi_{21}^{(2)}$</td>
<td>-0.1815</td>
<td>-0.140</td>
<td>-0.1815</td>
</tr>
<tr>
<td>$\phi_{22}^{(2)}$</td>
<td>0.1395</td>
<td>0.019</td>
<td>0.1395</td>
</tr>
</tbody>
</table>

With these final estimates, the models characterizing the three groups of time series is estimated as

**Region I**

\[
\begin{pmatrix}
X_t \\
Y_t
\end{pmatrix} = \begin{pmatrix}
-0.3891 & 0.0493 & -0.2965 & 0.4321 \\
-0.0294 & -0.0925 & -0.1815 & 0.1395
\end{pmatrix} \begin{pmatrix}
X_{t-1} \\
Y_{t-1} \\
X_{t-2} \\
Y_{t-2}
\end{pmatrix} + \begin{pmatrix}
e_{x_t} \\
e_{y_t}
\end{pmatrix}
\]

**Region II**

\[
\begin{pmatrix}
X_t \\
Y_t
\end{pmatrix} = \begin{pmatrix}
0.036 & 0.016 & -0.0009 & 0.009 \\
0.142 & 0.148 & -0.0140 & 0.019
\end{pmatrix} \begin{pmatrix}
X_{t-1} \\
Y_{t-1} \\
X_{t-2} \\
Y_{t-2}
\end{pmatrix} + \begin{pmatrix}
e_{x_t} \\
e_{y_t}
\end{pmatrix}
\]

144
Here, $X_t$ is Temperature in a region at time $t$ and $Y_t$ is Humidity level at time $t$.

Therefore the model that characterizes the whole data set is a three component mixture of bivariate autoregressive process of order 2 (BAR(2)) with component weights as $P(\omega_1) = 0.342$, $P(\omega_2) = 0.412$ and $P(\omega_3) = 0.246$ respectively for regions I, II and III.

For evaluating the performance of the proposed algorithm, true positive rate (TPR), false positive rate (FPR), false discovery rate (FDR) and F- measure are used. For the proposed supervised learning algorithm with bivariate autoregressive process of order 2, the performance measures for each class are computed and presented in Table 6.3.

<table>
<thead>
<tr>
<th></th>
<th>True Positive Rate (TPR)</th>
<th>False Positive Rate (FPR)</th>
<th>False Discovery Rate (FDR)</th>
<th>F Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>0.9667</td>
<td>0.0143</td>
<td>0.0333</td>
<td>0.9667</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>0.9800</td>
<td>0.0400</td>
<td>0.0392</td>
<td>0.9703</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>0.9500</td>
<td>0.0250</td>
<td>0.0952</td>
<td>0.9268</td>
</tr>
</tbody>
</table>

To compare the efficiency of the developed bivariate AR(2) classifier with the earlier AR(2) classifiers, the true positive rate (TPR), false positive rate (FPR), false
discover rate (FDR) and F-Measure are computed and presented in Table 6.4 and Table 6.5.

**Table 6.4**

Performance Measures of the Mixture of Univariate AR(2) Classifier for Temperature

<table>
<thead>
<tr>
<th>Clusters</th>
<th>True Positive Rate (TPR)</th>
<th>False Positive Rate (FPR)</th>
<th>False Discovery rate (FDR)</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>0.9333</td>
<td>0.0571</td>
<td>0.1250</td>
<td>0.9032</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>0.9400</td>
<td>0.0400</td>
<td>0.0408</td>
<td>0.9495</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>0.8947</td>
<td>0.0370</td>
<td>0.1500</td>
<td>0.8718</td>
</tr>
</tbody>
</table>

**Table 6.5**

Performance Measures of the Mixture of Univariate AR(2) Classifier for Humidity

<table>
<thead>
<tr>
<th>Clusters</th>
<th>True Positive Rate (TPR)</th>
<th>False Positive Rate (FPR)</th>
<th>False Discovery rate (FDR)</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>0.9355</td>
<td>0.0435</td>
<td>0.0968</td>
<td>0.9355</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>0.9592</td>
<td>0.0196</td>
<td>0.0208</td>
<td>0.9691</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>0.9000</td>
<td>0.0250</td>
<td>0.1000</td>
<td>0.9000</td>
</tr>
</tbody>
</table>

From Table 6.3, Table 6.4 and Table 6.5, it is observed that the F value for all categories using the proposed classifier are more compared to that of the classifier with univariate AR(2) models for both the variables. This indicates that the proposed classifier with bivariate AR(2) model is much better in classifying time series data than the other two classifiers.
For evaluating the developed algorithm discussed in section 6.5, the test data consisting of 100 time series is considered. The developed unsupervised learning algorithm with bivariate AR(2) model identified 29 as region I, 49 as region II and 19 region III. To compare the efficiency of developed algorithm with existing univariate unsupervised learning algorithms with AR(2) model for both the variables temperature and humidity, the same test data have been considered and the misclassification rates are computed. Table 6.6 presents the misclassification rates of BAR(2) classifier, AR(2) classifiers for Temperature and Humidity separately.

### Table 6.6

**Performance Evaluation of Misclassification Rate**

<table>
<thead>
<tr>
<th>Classifier with</th>
<th>Misclassification Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAR(2)</td>
<td>3%</td>
</tr>
<tr>
<td>AR(2) of temperature</td>
<td>8%</td>
</tr>
<tr>
<td>AR(2) of humidity</td>
<td>6%</td>
</tr>
</tbody>
</table>

From Table 6.6, it is observed that the misclassification rate for bivariate AR(2) classifier is less compared to misclassification rate of AR(2) classifiers of the individual variables Temperature and Humidity. Therefore the developed unsupervised learning algorithm with bivariate autoregressive process of order 2 outperforms the existing unsupervised learning algorithm with autoregressive processes of order 2. This algorithm is useful in other domains like medical, biological, financial applications etc.