3.1 INTRODUCTION

Classification is an important problem in Pattern Recognition. Classification of data is based on the nature of data collected according to spatial or temporal. In temporal data collection, time series forms an important class. Time series classification is one of the most predominant factor for efficient data analysis in image processing, signal processing, medical imaging, speech recognition, etc.,. Much work has been reported regarding classification of time series. Usually in classification problems a model is built describing a predetermined set of data classes or concepts. The model is constructed by analyzing data base tuples. Each tuple is assumed to be predefined class as determined by one of the attributes called the class label attribute. The data tuples analyzed to build the model collectively form training data set. Since the class labels of each training data set is provided or known, this classification problem is called supervised learning.

In time series classification, given a set of time series with class labels one can develop a model for each class. When a new time series is given, a label is assigned to it based on the model. For example, in medical data analysis, like classification of cardiac patients, the data collected on pulse rate over various time points is used. It is known that persons can be classified into three categories viz. normal persons, pro cardiac patients or cardiac patients. This class labeling is highly important for categorizing a person who has undergone medical checkup for efficient medical treatment to reduce risk for cardiac arrest. For this sort of data sets several supervised
learning algorithms are developed with various assumptions to suit a wide class of data sets.


In all these papers, the authors considered univariate time series data for supervised learning with single attribute. But in many time series data, there will be more than one attributes which play a significant role for class labeling. For example, in classification of cardiac patients, the pulse rate alone does not serve well unless it is monitored or observed along with other attributes like blood pressure. Similarly, in studies of classification of stock portfolios the B.S.E index as well as N.S.C. index both are correlated and form an important feature vector. Hence, to have an accurate classification it is needed to develop a supervised learning algorithm with bivariate time series classification. With this motivation a supervised learning algorithm based on bivariate autoregressive process of order 1 is developed and analyzed.

The autoregressive process is a generative model in the sense that the current value of time series is generated as a linear combination of previous values. In particular, the bivariate time series forms a linear combination of both variates under consideration. The model parameters are estimated through Expectation
Maximization (EM) algorithm with suitable initial estimates. A supervised learning algorithm with bivariate AR(1) model is developed by maximizing the class likelihood function. The accuracy of developed algorithm is studied through confusion matrices, true positive rate (TPR), false positive rate (FPR), precision, recall, F-measure and misclassification rate. A comparative study of this algorithm with earlier algorithms is also carried through real time data collected from a hospital. This algorithm is much useful for analyzing data sets in medical, meteorological, signal processing and other areas.

### 3.2 Finite Mixture of Bivariate AR(1) Model

In this section, a bivariate autoregressive process of order 1 and mixture of bivariate autoregressive process of order 1 are presented.

A bivariate process \( \left\{ \begin{pmatrix} X_t \\ Y_t \end{pmatrix}, t \in T \right\} \) is said to follow a bivariate autoregressive process of order 1 if it can be expressed as

\[
Z_t = \mu + \Phi (Z_{t-1} - \mu) + \epsilon_t,
\]

where, \( Z_t = \begin{pmatrix} X_t \\ Y_t \end{pmatrix} \), \( \mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \), \( \Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \) and \( \epsilon_t = \begin{pmatrix} \epsilon_{x_t} \\ \epsilon_{y_t} \end{pmatrix} \)

and \( \epsilon_t \) of \( Z_t, t=1,2,3,...,T \) follows a bivariate Gaussian distribution with mean vector as null vector and Variance–Covariance matrix as

\[
\Sigma = \begin{pmatrix} \sigma_{x_t}^2 & \rho \sigma_{x_t} \sigma_{y_t} \\ \rho \sigma_{x_t} \sigma_{y_t} & \sigma_{y_t}^2 \end{pmatrix}
\]

The probability density function of \( \epsilon_t \) is

\[
f(e_{x_t}, e_{y_t}) = \frac{1}{2\pi \sigma_{x_t} \sigma_{y_t} \sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left( \frac{e_{x_t}^2}{\sigma_{x_t}^2} - \frac{2\rho e_{x_t} e_{y_t}}{\sigma_{x_t} \sigma_{y_t}} + \frac{e_{y_t}^2}{\sigma_{y_t}^2} \right) \right\}
\]

\(-1 < \rho < 1, \sigma_{x_t} > 0 and \sigma_{y_t} > 0 \) (3.2.1)
This process reduces to univariate AR(1) process if $\phi_{12}, \phi_{21}$ and $\phi_{22}$ reduce to 0. This bivariate AR(1) process is an extension of AR(1) process usually known as Markov process of order 1.

This process is a stationary process in mean and variance. Its mean function is

$$E(Z_t) = \begin{bmatrix} \mu_x \\ 1 - \phi_{11} \\ \mu_y \\ 1 - \phi_{22} \end{bmatrix}$$

The conditional likelihood function of the time series can be expressed as

$$P(Z_t, \Phi) = \prod_{u=2}^{n} f\left(\left(e_{x,t,u}, e_{y,t,u}\right), \Phi\right)$$

(3.2.2)

since residual terms in the sample realisation of time series starts from $u = 2$ in bivariate AR(1) process,

i.e,  
$$e_{x,t,u} = X_{t,u} - \phi_{11}X_{t,u-1} - \phi_{12}Y_{t,u-1} \quad \text{and} \quad e_{y,t,u} = X_{t,u} - \phi_{21}X_{t,u-1} - \phi_{22}Y_{t,u-1}.$$

where, $\Phi = \{\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \sigma_{x}, \sigma_{y}, \rho\}$ is the set of model parameters and $n$ is the number of observations in each time series.

(George E.P.Box, Gwilym M. Jenkins, Gregory C.Reinsel (2003))

Substituting the value of $f\left(e_{x}, e_{y}\right)$ given in equation (3.2.1) in equation (3.2.2), the likelihood function of time series is obtained as

$$P(Z_t, \Phi) = \left(\frac{1}{2\pi \sigma_{e_x} \sigma_{e_y} \sqrt{1 - \rho^2}}\right)^{n-1} \exp\left\{-\frac{1}{2(1-\rho^2)} \sum_{u=2}^{n} \left(\frac{e_{x,t,u}^2}{\sigma_{e_x}^2} - \frac{2 \rho e_{x,t,u} e_{y,t,u}}{\sigma_{e_x} \sigma_{e_y}} + \frac{e_{y,t,u}^2}{\sigma_{e_y}^2}\right)\right\}$$

(3.2.3)
The logarithm of this likelihood is
\[
\ln P(Z_t, \Phi) = -(n-1) \ln \left( 2\pi \sigma_{\varepsilon_t} \sigma_{\eta_t} \sqrt{1-\rho^2} \right) - \frac{1}{2(1-\rho^2)} \sum_{s=2}^{n} \left( \frac{e_{\chi_{ts}}^2}{\sigma_{\varepsilon_t}^2} - \frac{2\rho e_{\chi_{ts}} e_{\varepsilon_{ts}}}{\sigma_{\varepsilon_t} \sigma_{\eta_t}} + \frac{e_{\varepsilon_{ts}}^2}{\sigma_{\eta_t}^2} \right)
\]

(3.2.4)

Let the whole time series data are generated by M different bivariate AR(1) models which corresponds to M class labels of interest with weights \(P(\omega_1), P(\omega_2), ..., P(\omega_M)\).

Let \(P(Z_t | \omega_k, \Phi_k)\) denote the conditional likelihood function of \(k^{th}\) class with \(\Phi_k\) as set of parameters for the model. Then the conditional likelihood function of mixture of finite bivariate AR process of order 1 can be expressed in the form

\[
P(Z_t | \Theta) = \sum_{k=1}^{M} P(Z_t | \omega_k, \Phi_k) P(\omega_k)
\]

(3.2.5)

where, \(\Theta = \{\Phi_1, \Phi_2, ..., \Phi_M, P(\omega_1), (\omega_2), ..., P(\omega_M)\}\) represents the set of model parameters for mixture model.

A bivariate time series \(Z_t\) is assigned to cluster \(\omega_k\) with posterior probability \(P(\omega_k | Z_t)\) such that \(\sum_{k=1}^{M} P(\omega_k) = 1\).

This model includes mixture of univariate autoregressive process of order 1 model when the model parameters \(\phi_{12}, \phi_{21}\) and \(\phi_{22}\) are zero for \(k=1, 2, ..., M\). If \(M=1\), then this process becomes bivariate AR(1) process.
3.3 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 estimates of the parameters $\sigma_{x_1}^2, \sigma_{y}^2, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}$ and $\rho$. For estimating the parameters through EM algorithm, one can use likelihood function of the sample.

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

$$P(D | \Theta) = \prod_{i=1}^N P(Z_i | \Theta)$$

$$= \prod_{i=1}^N \sum_{\omega_j, \Phi_k} P(Z_i | \omega_j, \Phi_k) P(\omega_k)$$

The Log likelihood function of realization $D$ is

$$\ln P(D | \Theta) = \sum_{i=1}^N \ln \left( \sum_{\omega_j, \Phi_k} P(Z_i | \omega_j, \Phi_k) P(\omega_k) \right)$$

In EM algorithm, the first step is to obtain expectation of log likelihood. 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

51
\[ Q(\Theta | \Theta(r)) = E(\ln P(D | \Theta)) \]
\[ = \sum_{i=1}^{N} \sum_{k=1}^{M} P(\omega_k | Z_i, \Theta(r)) \ln P(Z_i | \omega_k, \Phi_k) + \sum_{i=1}^{N} \sum_{k=1}^{M} P(\omega_k | Z_i, \Theta(r)) \ln P(\omega_k) \]

(3.3.1)

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

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

(3.3.2)

Using equation (3.2.4) and equation (3.3.2), equation (3.3.1) can be written as

\[ Q(\Theta | \Theta(r)) = \]
\[ \sum_{i=1}^{N} \sum_{k=1}^{M} P(\omega_k | Z_i, \Theta(r)) \left\{ -(n-1) \ln \left( 2\pi \sigma_{\epsilon_k} \sigma_{\mu_k} \sqrt{1-\rho_{\mu_k}^2} \right) - \frac{1}{2(1-\rho_{\mu_k}^2)} \sum_{u=2}^{n} \left( \frac{\epsilon_{u,\epsilon_k}^2}{\sigma_{\epsilon_k}^2} - \frac{2\rho_{\mu_k} \epsilon_{u,\mu_k} \epsilon_{u,\epsilon_k} + \epsilon_{u,\mu_k}^2}{\sigma_{\epsilon_k} \sigma_{\mu_k}} + \frac{\epsilon_{u,\mu_k}^2}{\sigma_{\mu_k}^2} \right) \right\} \]
\[ + \sum_{i=1}^{N} \sum_{k=1}^{M} P(\omega_k | Z_i, \Theta(r)) \ln \left\{ \frac{P(Z_i | \omega_k, \Phi_k) P(\omega_k)}{\sum_{q=1}^{M} P(Z_i | \omega_q, \Phi_q) P(\omega_q)} \right\} \]

(3.3.3)

In the M- step of EM algorithm, one has to find estimates of parameters by maximizing \( Q(\Theta | \Theta(r)) \). First maximize \( Q(\Theta | \Theta(r)) \) with respect to each \( P(\omega_k) \) by using Lagrangian multiplier method subject to the condition that

\[ \sum_{k=1}^{M} P(\omega_k) = 1 \]

(3.3.4)

Let \( \eta = Q(\Theta | \Theta(r)) - \lambda \left( \sum_{k=1}^{M} P(\omega_k) - 1 \right) \)

52
Then \( \frac{\partial \eta}{\partial P(\omega_k)} = 0 \), implies

\[
\frac{\partial Q(\Theta|\Theta(r))}{\partial P(\omega_k)} = \lambda
\]  

(3.3.5)

Differentiating \( Q(\Theta|\Theta(r)) \) given in equation (3.3.1) with respect to \( P(\omega_k) \), one can get

\[
\frac{\partial Q(\Theta|\Theta(r))}{\partial P(\omega_k)} = \frac{\partial}{\partial P(\omega_k)} \left[ \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \ln P(Z_i | \omega_k, \Phi_k) + \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \ln P(\omega_k) \right]
\]

\[
= \frac{1}{P(\omega_k)} \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r))
\]  

(3.3.6)

From equations (3.3.5) and (3.3.6), one can get

\[
\lambda = \frac{1}{P(\omega_k)} \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r))
\]  

(3.3.7)

This implies \( \lambda P(\omega_k) = \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \)

Hence, \( \lambda \sum_{k=1}^{M} P(\omega_k) = \sum_{i=1}^{N} \sum_{k=1}^{M} P(\omega_k | Z_i, \Theta(r)) \)  

(3.3.8)

Using equation (3.3.2), equation (3.3.8) and the condition given in equation (3.3.4)

\[
\lambda = N
\]  

(3.3.9)

Hence, from equations (3.3.7) and (3.3.9),

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

(3.3.10)

To obtain refined estimates of the parameters \( \sigma^2_{e_{a_x}}, \sigma^2_{e_{a_y}}, \phi_{11k}, \phi_{12k}, \phi_{21k}, \phi_{22k} \) and \( \rho_k \), differentiate \( Q(\Theta|\Theta(r)) \) partially with respect to the parameters and equate them to zero.
Differentiating $Q(\Theta|\Theta(r))$ with respect to $\sigma_{ek}^2$ and equating to zero implies

$$\frac{\partial Q(\Theta|\Theta(r))}{\partial \sigma_{ek}^2} = 0$$  \hspace{1cm} (3.3.11)$$

Substituting equation (3.3.3) in equation (3.3.11), one can get

\[
\frac{\partial}{\partial \sigma_{ek}^2} \left[ \sum_{i=1}^{N} P(\omega_k|Z_t, \Theta(r)) \right] \left[ (n-1) \ln \left( 2\pi \sigma_{ek} \right) \sigma_{ek} \sqrt{1-\rho_k^2} + \frac{1}{2(1-\rho_k^2)} \sum_{u=2}^{n} \left( \frac{e_{X,u}^2}{\sigma_{ex}^2} - \frac{2\rho_k e_{X,u} e_{Y,u} + e_{Y,u}^2}{\sigma_{ex} \sigma_{ek}} \right) \right] \\
- \sum_{i=1}^{N} P(\omega_k|Z_t, \Theta(r)) \ln P(\omega_k) \frac{P(Z_t|\omega_k, \Phi_k) P(\omega_k)}{\sum_{q=1}^{M} P(Z_t|\omega_q, \Phi_q) P(\omega_q)} = 0
\]

Which implies

$$\frac{\partial}{\partial \sigma_{ek}^2} \left[ \sum_{i=1}^{N} P(\omega_k|Z_t, \Theta(r)) \right] \left[ (n-1) - \frac{1}{(1-\rho_k^2)} \sum_{u=2}^{n} \left( \frac{e_{X,u}^2}{\sigma_{ex}^2} - \frac{\rho_k e_{X,u} e_{Y,u} + e_{Y,u}^2}{\sigma_{ex} \sigma_{ek}} \right) \right] = 0$$

This implies

$$\sum_{i=1}^{N} P(\omega_k|Z_t, \Theta(r)) \left[ (n-1) - \frac{1}{(1-\rho_k^2)} \sum_{u=2}^{n} \left( \frac{e_{X,u}^2}{\sigma_{ex}^2} - \frac{\rho_k e_{X,u} e_{Y,u} + e_{Y,u}^2}{\sigma_{ex} \sigma_{ek}} \right) \right] \left[ X_{t,u}^2 - \left( \frac{\phi_{1k}^2}{\sigma_{ex}^2} - \frac{\rho_k \phi_{12k} \phi_{22k}}{\sigma_{ex} \sigma_{ek}} \right) \right] \left[ X_{t,u-1}^2 - \left( \frac{2\phi_{12k}^2}{\sigma_{ex}^2} - \frac{\rho_k \phi_{22k}^2}{\sigma_{ex} \sigma_{ek}} \right) \right] \\
+ \left( \frac{2\phi_{1k} \phi_{12k} + \phi_{21k} \phi_{22k}}{\sigma_{ex}^2} \right) X_{t,u-1} = \left( \frac{2\phi_{12k} + \phi_{22k}}{\sigma_{ex}^2} \right) X_{t,u-1} Y_{t,u-1} - \left( \frac{\phi_{1k} \phi_{12k} + \phi_{21k} \phi_{22k}}{\sigma_{ex}^2} \right) \left[ X_{t,u} - \left( \frac{\phi_{1k} \phi_{12k} + \phi_{21k} \phi_{22k}}{\sigma_{ex}^2} \right) \right] \left[ Y_{t,u} - \left( \frac{\phi_{1k} \phi_{12k} + \phi_{21k} \phi_{22k}}{\sigma_{ex}^2} \right) \right]$$

$$- \frac{\rho_k}{\sigma_{ex} \sigma_{ek}} \left[ X_{t,u} Y_{t,u} - \phi_{1k} X_{t,u-1} Y_{t,u} - \phi_{21k} Y_{t,u-1} Y_{t,u} \right] = 0$$  \hspace{1cm} (3.3.12)
Differentiating $Q(\Theta | \Theta (r))$ with respect to $\sigma^2_{\theta_k}$ and equating to zero implies

$$\frac{\partial Q(\Theta | \Theta (r))}{\partial \sigma^2_{\theta_k}} = 0$$

(3.3.13)

Substituting equation (3.3.3) in equation (3.3.13), one can get

$$\frac{\partial}{\partial \sigma^2_{\theta_k}} \left[ \sum_{i=1}^{n} P(\omega_k | Z_i, \Theta (r)) \left\{ (n-1) \ln \left( 2\pi \sigma_{\theta_k} \right) \frac{1}{\sqrt{\sigma_{\theta_k}}} + \frac{1}{2(1-\rho_k^2)} \sum_{u=2}^{n} \left( \frac{\sigma_{\theta_k}^2}{\sigma_{\theta_k}^2} \rho_{\theta_k \theta_u, \theta_u} \epsilon_{\theta_u} \right) \right\} \right] = 0$$

Which implies

$$\frac{\partial}{\partial \sigma^2_{\theta_k}} \left[ \sum_{i=1}^{n} P(\omega_k | Z_i, \Theta (r)) \left\{ (n-1) - \frac{1}{2(1-\rho_k^2)} \sum_{u=2}^{n} \left( \rho_{\theta_k \theta_u, \theta_u} \sigma_{\theta_u}^2 - \frac{\epsilon_{\theta_u}^2}{\sigma_{\theta_u}^2} \right) \right\} \right] = 0$$

This implies

$$\sum_{i=1}^{n} P(\omega_k | Z_i, \Theta (r)) \left\{ (n-1) - \frac{1}{2(1-\rho_k^2)} \sum_{u=2}^{n} \left( \rho_{\theta_k \theta_u, \theta_u} \sigma_{\theta_u}^2 - \frac{\epsilon_{\theta_u}^2}{\sigma_{\theta_u}^2} \right) \right\} X_{t,u-1}^2$$

$$+ \left( \frac{\phi_{21k}^2}{\sigma_{\theta_k}^2} - \rho_{\theta_k \theta_{21k}} \phi_{21k} \right) Y_{t,u-1}^2$$

$$- \left( \frac{2 \phi_{21k} \phi_{11k}}{\sigma_{\theta_k}^2} - \frac{\rho_{\theta_k \theta_{11k}} \phi_{11k}}{\sigma_{\theta_k}^2} \right) X_{t,u-1} Y_{t,u}$$

$$+ \left( \frac{2 \phi_{11k} \phi_{21k}}{\sigma_{\theta_k}^2} - \rho_{\theta_k \theta_{21k}} \phi_{21k} \right) X_{t,u-1} Y_{t,u-1}$$

$$- \frac{\rho_{\theta_k}}{\sigma_{\theta_k} \sigma_{\theta_u}} \left( X_{t,u} Y_{t,u} - \phi_{21k} X_{t,u} X_{t,u-1} - \phi_{21k} X_{t,u} Y_{t,u-1} \right) = 0$$

(3.3.14)
Similarly, \( \partial Q(\Theta|\Theta(r)) / \partial \phi_{11k} = 0 \) implies

\[
\frac{\partial}{\partial \phi_{11k}} \left[ \sum_{i=1}^{n} P(\omega_i | Z_i, \Theta(r)) \left\{ (n-1) \ln \left( 2 \pi \sigma_{x_i} \sigma_{e_i} \sqrt{1 - \rho_k^2} \right) + \frac{1}{2(1-\rho_k^2)} \sum_{u=2}^{n} \left( \frac{e_{x,u}^2}{\sigma_{x_i}^2} - \frac{2 \rho_k e_{x,u} e_{e_i}}{\sigma_{x_i} \sigma_{e_i}} + \frac{e_{e_i}^2}{\sigma_{e_i}^2} \right) \right\} \right] = 0
\]

Which implies

\[
\frac{\partial}{\partial \phi_{11k}} \left( \sum_{i=1}^{n} P(\omega_i | Z_i, \Theta(r)) \sum_{u=2}^{n} \left\{ \frac{e_{x,u}}{\sigma_{x_i}} - \frac{\rho_k e_{e_i}}{\sigma_{e_i}} \right\} X_{i,u-1} \right) = 0
\]

This implies

\[
\sum_{i=1}^{n} P(\omega_i | Z_i, \Theta(r)) \left[ \frac{1}{\sigma_{x_i}} \sum_{u=2}^{n} X_{i,u} - \frac{\rho_k}{\sigma_{e_i}} \sum_{u=2}^{n} X_{i,u-1} Y_{i,u} \right]
\]

\[
- \left( \frac{\phi_{11k}}{\sigma_{x_i}} - \frac{\rho_k \phi_{21k}}{\sigma_{e_i}} \right) \sum_{u=2}^{n} X_{i,u-1} - \left( \frac{\phi_{12k}}{\sigma_{x_i}} - \frac{\rho_k \phi_{22k}}{\sigma_{e_i}} \right) \sum_{u=2}^{n} X_{i,u-1} Y_{i,u-1} \right] = 0 \quad (3.3.15)
\]

Similarly, \( \partial Q(\Theta|\Theta(r)) / \partial \phi_{12k} = 0 \) implies

\[
\frac{\partial}{\partial \phi_{12k}} \left[ \sum_{i=1}^{n} P(\omega_i | Z_i, \Theta(r)) \left\{ (n-1) \ln \left( 2 \pi \sigma_{x_i} \sigma_{e_i} \sqrt{1 - \rho_k^2} \right) + \frac{1}{2(1-\rho_k^2)} \sum_{u=2}^{n} \left( \frac{e_{x,u}^2}{\sigma_{x_i}^2} - \frac{2 \rho_k e_{x,u} e_{e_i}}{\sigma_{x_i} \sigma_{e_i}} + \frac{e_{e_i}^2}{\sigma_{e_i}^2} \right) \right\} \right] = 0
\]

\[
- \sum_{i=1}^{n} P(\omega_i | Z_i, \Theta(r)) \ln P(\omega_i) \left\{ \frac{P(Z_i | \omega_i, \Phi_i) P(\omega_i)}{\sum_{i=1}^{n} P(Z_i | \omega_i, \Phi_i) P(\omega_i)} \right\} = 0
\]
Which implies
\[ \frac{\partial}{\partial \phi_{12k}} \left( \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \left[ \sum_{u=2}^{n} \left( \frac{e_{x_{i,u}}}{\sigma_{x_k}} - \frac{\rho_k e_{y_{i,u}}}{\sigma_{e_k}} \right) Y_{i,u-1} \right] \right) = 0 \]

This implies
\[ \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \left[ \frac{1}{\sigma_{x_k}} \sum_{u=2}^{n} X_{i,u-1} Y_{i,u-1} - \frac{\rho_k}{\sigma_{e_k}} \sum_{u=2}^{n} Y_{i,u-1} Y_{i,u} \right] - \left( \frac{\phi_{11k}}{\sigma_{x_k}} - \frac{\rho_k \phi_{21k}}{\sigma_{e_k}} \right) \sum_{u=2}^{n} X_{i,u-1} Y_{i,u-1} - \left( \frac{\phi_{22k}}{\sigma_{e_k}} - \frac{\rho_k \phi_{22k}}{\sigma_{e_k}} \right) \sum_{u=2}^{n} Y_{i,u-1} Y_{i,u-1} = 0 \] (3.3.16)

Similarly, \( \frac{\partial Q(\Theta | \Theta(r))}{\partial \phi_{21k}} = 0 \) implies
\[ \frac{\partial}{\partial \phi_{21k}} \left( \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \left[ (n-1) \ln \left( 2\pi \sigma_{x_k} \sigma_{e_k} \sqrt{1-\rho_k^2} \right) + \frac{1}{2(1-\rho_k^2)} \sum_{u=2}^{n} \left( \frac{e_{x_{i,u}}^2}{\sigma_{x_k}^2} + \frac{2 \rho_k e_{x_{i,u}} e_{y_{i,u}}}{\sigma_{x_k} \sigma_{e_k}} + \frac{e_{y_{i,u}}^2}{\sigma_{e_k}^2} \right) \right] \right) - \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \ln P(\omega_k) \left[ \frac{P(Z_i | \omega_k, \Phi_k) P(\omega_k)}{\sum_{q=1}^{M} P(Z_i | \omega_q, \Phi_q) P(\omega_q)} \right] = 0 \]

Which implies
\[ \frac{\partial}{\partial \phi_{21k}} \left( \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \left[ \sum_{u=2}^{n} \left( \frac{\rho_k e_{x_{i,u}}}{\sigma_{x_k}} - \frac{e_{y_{i,u}}}{\sigma_{e_k}} \right) X_{i,u-1} \right] \right) = 0 \]

This implies
\[ \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \left[ \frac{\rho_k}{\sigma_{x_k}} \sum_{u=2}^{n} X_{i,u-1} X_{i,u} - \frac{1}{\sigma_{e_k}} \sum_{u=2}^{n} X_{i,u-1} Y_{i,u} \right] - \left( \frac{\rho_k \phi_{11k}}{\sigma_{x_k}} - \frac{\phi_{21k}}{\sigma_{e_k}} \right) \sum_{u=2}^{n} X_{i,u-1} X_{i,u-1} - \left( \frac{\rho_k \phi_{12k}}{\sigma_{x_k}} - \frac{\phi_{22k}}{\sigma_{e_k}} \right) \sum_{u=2}^{n} X_{i,u-1} Y_{i,u-1} = 0 \] (3.3.17)
Similarly, \( \frac{\partial Q(\Theta | \Theta(r))}{\partial \phi_{22k}} = 0 \) implies

\[
\frac{\partial}{\partial \phi_{22k}} \left[ \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \right] \left\{ (n-1) \ln \left( 2\pi \sigma_{e_k} \sigma_{e_n} \sqrt{1-\rho_k^2} \right) + \frac{1}{2(1-\rho_k^2)} \sum_{u=2}^{n} \left( \frac{e_{x,u}^2}{\sigma_{e_x}^2} - \frac{2\rho_k e_{x,u} e_{y,u} + e_{y,u}^2}{\sigma_{e_y}^2} \right) \right\}

- \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \ln P(\omega_k) \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\} = 0
\]

Which implies

\[
\frac{\partial}{\partial \phi_{22k}} \left( \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \right) \left\{ \sum_{u=2}^{n} \left( \frac{\rho_k e_{x,u} - e_{y,u}}{\sigma_{e_x}^2} \right) Y_{u-1} \right\} = 0
\]

This implies

\[
\sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \left[ \frac{\rho_k}{\sigma_{e_x}} \sum_{u=2}^{n} X_{t,u} Y_{t,u-1} - \frac{1}{\sigma_{e_n}} \sum_{u=2}^{n} Y_{t,u-1} Y_{t,u} \right.

- \left( \frac{\rho_k \phi_{12k} - \phi_{21k}}{\sigma_{e_x}} \sum_{u=2}^{n} X_{t,u-1} Y_{t,u-1} \right) \left( \frac{\rho_k \phi_{12k} - \phi_{22k}}{\sigma_{e_n}} \sum_{u=2}^{n} Y_{t,u-1}^2 \right) = 0 \quad (3.3.18)
\]

Similarly, \( \frac{\partial Q(\Theta | \Theta(r))}{\partial \rho_k} = 0 \) implies

\[
\frac{\partial}{\partial \rho_k} \left[ \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \right] \left\{ (n-1) \ln \left( 2\pi \sigma_{e_k} \sigma_{e_n} \sqrt{1-\rho_k^2} \right) + \frac{1}{2(1-\rho_k^2)} \sum_{u=2}^{n} \left( \frac{e_{x,u}^2}{\sigma_{e_x}^2} - \frac{2\rho_k e_{x,u} e_{y,u} + e_{y,u}^2}{\sigma_{e_y}^2} \right) \right\}

- \sum_{i=1}^{N} P(\omega_k | Z_i, \Theta(r)) \ln P(\omega_k) \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\} = 0
\]
Which implies
\[
\sum_{i=1}^{n} P(\omega_k | Z_t, \Theta(r)) \left[ (n-1) \rho_k - \frac{\rho_k}{1 - \rho^2_k} \sum_{w=2}^{n} \left( \frac{e_{x,t,w}^2}{\sigma_{x,t}^2} + \frac{e_{y,t,w}^2}{\sigma_{y,t}^2} \right) + \left( 1 + \rho^2_k \right) \sum_{w=2}^{n} \frac{e_{x,t,w} e_{y,t,w}}{\sigma_{x,t} \sigma_{y,t}} \right] = 0
\]

This implies
\[
\sum_{i=1}^{n} P(\omega_k | Z_t, \Theta(r)) \left[ (n-1) \rho_k - \frac{\rho_k}{1 - \rho^2_k} \sum_{w=2}^{n} \left( \frac{X_{t,u}^2}{\sigma_{x,t}^2} + \frac{Y_{t,u}^2}{\sigma_{y,t}^2} \right) + \left( \phi_{11k}^2 + \phi_{21k}^2 \right) X_{t,u-1}^2 + \left( \phi_{12k}^2 + \phi_{22k}^2 \right) Y_{t,u-1}^2 \right. \\
-2 \left( \phi_{11k} X_{t,u} X_{t,u-1} \sigma_{x,t}^2 + \phi_{12k} X_{t,u} Y_{t,u-1} \sigma_{y,t}^2 \right) X_{t,u-1} Y_{t,u-1} + \left( \phi_{21k} Y_{t,u} X_{t,u-1} \sigma_{y,t}^2 + \phi_{22k} Y_{t,u} Y_{t,u-1} \sigma_{y,t}^2 \right) Y_{t,u-1}^2 \\
+ \left( \phi_{11k} + \phi_{12k} \right) X_{t,u-1} Y_{t,u-1} - \phi_{21k} X_{t,u} X_{t,u-1} - \phi_{22k} X_{t,u} Y_{t,u-1} - \phi_{11k} X_{t,u-1} Y_{t,u} + \phi_{11k} \phi_{21k} X_{t,u-1} \right] = 0
\]

\[ (3.3.19) \]

Solving the equations (3.3.10), (3.3.12), (3.3.14), (3.3.15), (3.3.16), (3.3.17), (3.3.18) and (3.3.19) simultaneously, the refined estimates of model parameters \( P(\omega_k), \sigma_{x,t}^2, \sigma_{y,t}^2, \phi_{11k}, \phi_{12k}, \phi_{21k}, \phi_{22k} \) and \( \rho_k \) can be obtained.

### 3.4 INITIAlISATION OF PARAMETERS

To utilize EM algorithm, one has to initialize the parameters using training data set. For the parameters \( \sigma_{x,t}^2, \sigma_{y,t}^2 \) and \( \rho_k \), consider sample variances of the estimated residual of \( X_t, Y_t \) and sample correlation coefficient as initial estimates respectively.

59
Thus the initial estimates of $\sigma_{x_k}^2$ and $\sigma_{y_k}^2$ are

$$
\hat{\sigma}_{x_k}^2 = \frac{1}{(n-1)G_k} \sum_{g=1}^{G_k} \sum_{u=2}^{n} e_{x_ku}^2 \quad \text{and} \quad \hat{\sigma}_{y_k}^2 = \frac{1}{(n-1)G_k} \sum_{g=1}^{G_k} \sum_{u=2}^{n} e_{y_ku}^2, \quad k=1,2,\ldots,M
$$

(3.4.1)

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 $Y_t$ and is given by

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

(3.4.2)

where, $G_k$ is the number of time series in the $k^{th}$ cluster.

For initializing parameters $\phi_{11k}, \phi_{12k}, \phi_{21k}$ and $\phi_{22k}$, ordinary least squares estimates of the parameters of the bivariate autoregressive process of order 1 are obtained for each time series in each class and averaged over the number of time series in that class. The sample realization of bivariate autoregressive process of order 1 can be written in the following format

$$
\begin{bmatrix}
X_{t,2} & Y_{t,2} \\
X_{t,3} & Y_{t,3} \\
\vdots & \vdots \\
X_{t,n} & Y_{t,n}
\end{bmatrix}
= \begin{bmatrix}
X_{t,1} & Y_{t,1} \\
X_{t,2} & Y_{t,2} \\
\vdots & \vdots \\
X_{t,n-1} & Y_{t,n-1}
\end{bmatrix}
\begin{bmatrix}
\phi_{11t} & \phi_{12t} \\
\phi_{21t} & \phi_{22t}
\end{bmatrix}
+ \begin{bmatrix}
e_{x_{t,2}} & e_{y_{t,2}} \\
e_{x_{t,3}} & e_{y_{t,3}} \\
\vdots & \vdots \\
e_{x_{t,n}} & e_{y_{t,n-1}}
\end{bmatrix}
$$

which can be represented as

$$
Y_t = X_t \Phi_t + \xi_t
$$
where,

\[
\begin{bmatrix}
X_{t,2} & Y_{t,2} \\
X_{t,3} & Y_{t,3} \\
\vdots & \vdots \\
X_{t,n} & Y_{t,n}
\end{bmatrix}
= \begin{bmatrix}
X_{t,1} & Y_{t,1} \\
X_{t,2} & Y_{t,2} \\
\vdots & \vdots \\
X_{t,n-1} & Y_{t,n-1}
\end{bmatrix}
\text{ and }
\begin{bmatrix}
e_{t,1} \\
e_{t,2}
\end{bmatrix}
= \begin{bmatrix}
\phi_{11t} & \phi_{12t} \\
\phi_{21t} & \phi_{22t}
\end{bmatrix}
\]

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

\[
\hat{\Phi}_t = (X_t^T X_t)^{-1} X_t^T Y_t \quad \text{for } t = 1, 2, \ldots, N
\] (3.4.3)

**The Expectation Maximization Algorithm**

Step 1: Find initial parameters using the equations (3.4.1) to (3.4.3).

Step 2: Obtain the revised estimates of parameters \( P(\omega_k), \sigma_{\omega_k}^2, \sigma_{\phi_k}^2, \phi_{11k}, \phi_{12k}, \phi_{21k}, \phi_{22k} \)

and \( \rho_k \) using the equations (3.3.10), (3.3.12), (3.3.14), (3.3.15), (3.3.16), (3.3.17), (3.3.18) and (3.3.19).

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 the final estimates of parameters \( P(\omega_k), \sigma_{\omega_k}^2, \sigma_{\phi_k}^2, \phi_{11k}, \phi_{12k}, \phi_{21k}, \phi_{22k} \)

and \( \rho_k \).

**3.5 SUPERVISED LEARNING ALGORITHM**

In this section, the supervised learning algorithm is presented. The steps in the algorithm are as follows:

Step 1: Identify number of class labels.

Step 2: Obtain initial estimates of the model parameters using equations (3.4.1), (3.4.2) and (3.4.3).
Step 3: Obtain the refined estimates of model parameters using Expectation
Maximization algorithm until we get final estimates.

Step 4: For the new data point compute conditional likelihood with the model
parameters of \( k^{th} \) class derived from step 3 and assign it to the class for which
sample conditional likelihood is maximum. i.e. the classification is
\[
\hat{C} = \arg \max_k P(Z_t | C_k)
\]

where, \( \hat{C} \) is the maximum likelihood class and \( Z_t \) is the new time series data.

3.6 EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

In this section, the utility of developed supervised learning algorithm for
classifying the patients with cardiac problem is considered. After discussions with
medical personal, it has been observed that 3 categories of persons namely, Normal,
Pro-Cardiac and Cardiac patients, usually visit hospital. These 3 categories can be
taken as class labels and the number of groups is known as 3.

The data of 200 patients who visited the hospital with respect to two attributes
namely, blood pressure (BP) and pulse rate (PR) is collected for 96 time points. This
bivariate time series for each individual is characterized with bivariate autoregressive
process of order 1. From this data set, 100 time series are considered as the training
data set and 100 time series for the test data. A sample scatter plot of BP and for PR
against time is plotted and shown in figure 3.1.
From figure 3.1, it is observed that the time series are stationary after obtaining deviation from mean values. For determining the lag of AR process for each variate of every class, the autocorrelations and partial autocorrelations for both variables for a sample of time series are computed and the correlograms are presented in figure 3.2 to figure 3.13.
From figure 3.2 to figure 3.13, the partial autocorrelogram cut off at lag1 and autocorrelogram tails off. Therefore the assumption that the sample time series of each class follows a bivariate autoregressive process of order 1 is suitable.

Using the initialization of parameters discussed in section 3.4, the initial estimates of parameters $P(\omega_k), \sigma_{\epsilon_k}^2, \sigma_{\epsilon_2}^2, \rho$, $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_7, \phi_8$ are obtained and presented in Table 3.1.
Using the initial estimates and EM algorithm, refined estimates of parameters for each class are obtained and presented in Table 3.2.

With these final estimates, the models characterizing the three groups of persons are estimated as
Here, $X_t$ is blood pressure (BP) of the patients at time $t$ and $Y_t$ is pulse rate (PR) of the patient at time $t$.

Therefore the model that characterizes the whole data set is a three component mixture of bivariate autoregressive process of order 1 (BAR(1)) with component weights as $P(\omega_1) = 0.299$, $P(\omega_2) = 0.502$ and $P(\omega_3) = 0.199$ respectively.

For evaluating the developed algorithm discussed in section 3.5, the test data consisting of 100 patients with 30 patients under normal group, 50 patients as Pro Cardiac group and 20 patients as Cardiac group is considered. The developed supervised learning algorithm with bivariate AR(1) model identified 29 persons as normal group, 49 persons as pro-cardiac group and 18 persons as cardiac group. To compare the efficiency of the developed algorithm with existing univariate supervised learning algorithms with AR(1) model for both attributes BP and PR the confusion matrices are computed (Sumit Goswami et al. (2009)). The confusion matrices
summarizing the instances predicted correctly or incorrectly by the classifiers namely, the classifier with mixture of BAR(1) and classifier with mixture of AR(1) for each attribute are presented in Table 3.3, Table 3.4 and Table 3.5.

Table 3.3
Confusion Matrix for Classifier with Mixture of BAR(1)

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Pro Cardiac</th>
<th>Cardiac</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>29</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Pro Cardiac</td>
<td>1</td>
<td>49</td>
<td>0</td>
</tr>
<tr>
<td>Cardiac</td>
<td>0</td>
<td>2</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3.4
Confusion Matrix for Classifier with Mixture of AR(1) for BP

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Pro Cardiac</th>
<th>Cardiac</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>28</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Pro Cardiac</td>
<td>3</td>
<td>46</td>
<td>1</td>
</tr>
<tr>
<td>Cardiac</td>
<td>1</td>
<td>3</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 3.5
Confusion Matrix for Classifier with Mixture of AR(1) for PR

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Pro Cardiac</th>
<th>Cardiac</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>24</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Pro Cardiac</td>
<td>3</td>
<td>42</td>
<td>5</td>
</tr>
<tr>
<td>Cardiac</td>
<td>3</td>
<td>2</td>
<td>15</td>
</tr>
</tbody>
</table>
Using these confusion matrices, the precision and recall for each class by each classifier are computed. The combined measure for precision and recall is F-measure which is the harmonic mean between precision and recall are also computed for each group and presented in Table 3.6.

### Table 3.6

**Comparative Performance of Supervised Learning Algorithm**

<table>
<thead>
<tr>
<th>Class</th>
<th>Model</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Bivariate</td>
<td>0.9666</td>
<td>0.0142</td>
<td>0.9666</td>
<td>0.9666</td>
<td>0.9666</td>
</tr>
<tr>
<td></td>
<td>Univariate (BP)</td>
<td>0.9333</td>
<td>0.0571</td>
<td>0.8750</td>
<td>0.9333</td>
<td>0.9032</td>
</tr>
<tr>
<td></td>
<td>Univariate (PR)</td>
<td>0.8000</td>
<td>0.0857</td>
<td>0.8000</td>
<td>0.8000</td>
<td>0.8000</td>
</tr>
<tr>
<td>Pro Cardiac</td>
<td>Bivariate</td>
<td>0.9800</td>
<td>0.0400</td>
<td>0.9800</td>
<td>0.9800</td>
<td>0.9800</td>
</tr>
<tr>
<td></td>
<td>Univariate (BP)</td>
<td>0.9200</td>
<td>0.0600</td>
<td>0.9387</td>
<td>0.9200</td>
<td>0.9292</td>
</tr>
<tr>
<td></td>
<td>Univariate (PR)</td>
<td>0.8400</td>
<td>0.1000</td>
<td>0.8936</td>
<td>0.8400</td>
<td>0.8659</td>
</tr>
<tr>
<td>Cardiac</td>
<td>Bivariate</td>
<td>0.9000</td>
<td>0.0125</td>
<td>0.9473</td>
<td>0.9000</td>
<td>0.9230</td>
</tr>
<tr>
<td></td>
<td>Univariate (BP)</td>
<td>0.8000</td>
<td>0.0375</td>
<td>0.8421</td>
<td>0.8000</td>
<td>0.8205</td>
</tr>
<tr>
<td></td>
<td>Univariate (PR)</td>
<td>0.7500</td>
<td>0.1000</td>
<td>0.6521</td>
<td>0.7500</td>
<td>0.6976</td>
</tr>
</tbody>
</table>

From Table 3.6, it is observed that the F-measure in all the three groups for bivariate supervised learning algorithm are close to 1 ensures that the classifier classifies persons more accurately to their corresponding class. It is also observed that both precision and recall are reasonably high for the developed bivariate supervised learning algorithm compared to the existing univariate supervised learning algorithm.

The classifier accuracy and misclassification rate for each class using supervised learning algorithm with bivariate AR(1) model of order 1 and supervised learning algorithm with univariate AR(1) model of order 1 for attributes blood pressure and pulse rate are also computed and presented in Table 3.7.
Table 3.7

Misclassification Rates of the Classifiers

<table>
<thead>
<tr>
<th>Group</th>
<th>Bivariate AR(1)</th>
<th>Univariate AR(1)</th>
<th>Univariate AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blood Pressure</td>
<td>Pulse Rate</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>96%</td>
<td>3%</td>
<td>75%</td>
</tr>
<tr>
<td>II</td>
<td>98%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>90%</td>
<td>19%</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>93%</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>92%</td>
<td>84%</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>80%</td>
<td>75%</td>
<td></td>
</tr>
</tbody>
</table>

From Table 3.7, it is observed that the accuracy of developed classifier for bivariate is 96% with respect to Class I, 98% with respect to Class II and 90% with respect to Class III and overall misclassification rate is 4%, whereas for univariate blood pressure the accuracy is 93% with respect to Class I, 92% with respect to Class II and 80% with respect to Class III with overall misclassification rate as 10%. The accuracy for univariate pulse rate is 80% with respect to Class I, 84% with respect to Class II and 75% with respect to Class III with overall misclassification rate as 19%. Therefore the developed supervised learning algorithm outperforms than the available supervised learning algorithms with univariate autoregressive process of order 1. This algorithm is also useful for classifying the bivariate time series data arriving in many applications from scientific, sociological, financial and other domains.