

## **Bayesian Classifier for a Gaussian Distribution, Decision Surface Equation, with Application**

**Nawzad. M. Ahmad<sup>\*</sup>**

---

### **ABSTRACT**

Bayesian decision theory is a fundamental statistical approach to the problem of classification as for pattern recognition. It makes the assumption that the decision problem is posed in probabilistic term, and all of the relevant probability values are known. To minimize the error probability in classification problem, one must choose the state of nature that maximizes the posterior probability. Bayes formula allows us to calculate such probabilities given the prior probabilities, and the conditional densities for different classes or categories.

Statistical classification is one of the most useful analysis tool which can be used for analyzing several kinds of data in various sciences. Its interesting in separating more than one class or category from each other, when their behaviors are near each other. The Bayesian surface decision equation is one of the classification rule that produces a linear separable equation for these interacted classes, specially where they contain vectors of random variables that are identically independently distributed (iid),(or all vectors

---

<sup>\*</sup>Prof. Dr/Department of Statistics/College of Administration & Economy/ University of Sulaimni. [www.nawzadstat@yahoo.com](mailto:www.nawzadstat@yahoo.com)

have the same distribution with same parameters values). This was done as a special case by *Muller, P. & Insua, D.R, (1995)*, This study is a trial to generalize (*Bayesian Surface Decision Equation*) to produce a such (*Linear Separable Equation*) as a linear classifier for those classes contain random vectors distributed identically Gaussian) but with different parameters values (  $\mathbf{m}_i$  ,  $\Sigma_i$  ), moreover, in this study the researcher tried to search for equivalence between *Bayesian Surface Decision Equation*, and a *linear perceptron* for classification for this general case, with a numeric application, (encoded data vectors) that is illustrated latter.

( ) ( )

,(classification)

state of )

(posterior)

(nature

(conditional p.d.f)

.(Prior p.d.f)

(separation)

) .

(Bayse Surface Decision Equation

(vectors of random variables)

.(iid)

. (1995) Insua. D Muller .P

( ) \_\_\_\_\_

(id)

(Gaussian)

\_\_\_\_ , ( Insua Muller)

(Linear Seperable Equation) )

(Perceptron)

(Binary)

\_\_\_\_ , (Threshold)

(Perceptron) (BSDE)

(logical Forms )

,

**Keywords:** Perceptron, Multivariate Gaussian Distribution, , threshold( $\theta$ ), Bayesian Surface Decision Equation (BSDE), Classification, logical Form. Linear separable equation (LSE), hard limiter (HL), Minimum risk decision rule (MRDR).

## 1- Introduction

Classification is a statistical task of learning a target function that maps each vector( $\underline{X}$ ), or set to one of the predefined class labels( $k_i$ ), the target function is also known informally as a classification model which is useful for *descriptive modeling* which can serve as an explanatory tool to distinguish between objects of different classes, or for *predictive model* which is used to predict the class label of unknown or records never seen before. Classification model can be treated as a black box that automatically assigns a class label when presented with the attribute unknown set or vector of random variables.

Classification techniques (*Classifiers*), such as (Rule Bayes Classifier, Neural Net Works, Naïve Bayes Classifier, and Support Vector Machine) are most suited for predicting or describing

categorical data, then classification is a systematic approach for building classification models for an input data records.

Classification techniques employ a learning algorithm to identify a model that best fits the relationship between attribute set and class label for the input data, then the fitted classifier from the algorithm should both fit the input data well, and correctly predict the class labels of records that have never seen before.

Therefore a key objective of classification is to build model that accurately predict the class labels or patterns of previously unknown records.

Its clear that the Bayesian Decision Rule(BDR) has great role in statistical data analysis for various directions in our live, specially in stochastic processes, it is also a good tool if it is used in classification to identify and to compare several groups or statistical populations(classes), and also taking classifier knowledge about the source of observations that they come from, by identifying the behaviors of random variables and playback or replacing them to their origin sources(behaviors of classes, or statistical distributions must be known previously). This can be done when one needs to make classification among some classes come from several populations, to return to their origin. And also this criteria can be done by using what is so called (Perceptron), or generally (linear

perceptron). Then it must have a similarity or in convenience to a mathematical equivalence form between (BSDE), and the perceptron.

### **1-1 The aim of the study**

The aim of this study is trying to establish the following two important points:

**1<sup>st</sup>**/ Making sure, about the equivalence between single layer perceptron, and the difference between two (BDR) for two classes [ $d(X_1)-d(X_2)$ ]. This difference is called (Bayesian Surface Decision Equation (BSDE), or a Bayesian decision boundary that is evaluated from two independent but not identical Gaussian classes (multivariate normal density).ie. each class contains a set of vector of random variables with Gaussian distribution , but with different mean vector and positive semi-definite variance-covariance matrix as shown below.

**2<sup>nd</sup>**/Applying single perceptron based on (BSDE) criterion on data under consideration in the study, to define a Bayesian decision

equation classifier and the decision boundary for a set of variable vectors in each class refer to two different Gaussian patterns defined as:

$$\mathbf{X}_{1p} \sim N(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_{x1}) , p = 1, 2, \dots, P \text{ for } 1^{\text{st}} \text{ class } (k_1)$$

$$\mathbf{X}_{2p} \sim N(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_{x2}) , p = 1, 2, \dots, P \text{ for } 2^{\text{nd}} \text{ class } (k_2)$$

Such that:  $P$  = maximum number of random variates in class  $(k_i)$ ,  
 $i = 1, 2$

$\boldsymbol{\mu}_p$  = mean vector , and  $\boldsymbol{\Sigma}$  = variance-covariance matrix

## 1-2 Hypothesis of the Study

The above illustrated two points in the aim of the study can be suggested to be the main hypotheses that must be studied.

## 2- Single Perceptron, and (BSDE) for Multivariate Normal Distribution

### 2-1 Single Perceptron

It is usually named by linear perceptron or simply perceptron which is defined as a (Linear Classifier, or linear Separable Equation LSE) having one input , and one output, these two layers and their nodes (Neurons) are connected with each other by a linear

function named by (*hard limiter*), through a random set of weights ( $w_i$ ), as illustrated by the figure below.

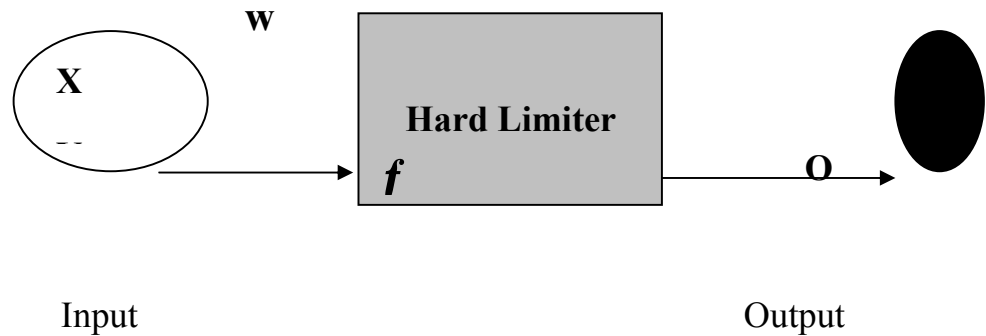


Figure (1) : perceptron

The diagram above can be expressed by the following system:

$$\text{Net}(y) = \sum x_i w_i + \theta \quad i = 1, 2, \dots, n$$

$$1 \quad \text{if} \quad \text{net}(y) \geq \theta \quad \dots \dots \dots (1)$$



**Hard limiter ( net(y)=**

$$0 \quad \text{if} \quad \text{net}(y) < \theta$$

It is important to explain that the *threshold* ( $\theta$ ) is similar to the *intercept* in any linear system, and the *net(y)* is output produced by the system, at the same time

(  $y_i$  )is the *desired output* that we search for. Now after transforming sets of data referring to two populations, to a *logical form* (*AND, OR, XOR*), and introduced it to the perceptron system as an *encoded data patterns* (binary data). Then, one can use it to separate the data input to their origin categories (classes) which are coming from, this can be done by finding a relationship called *linear separable function*. This process is indeed a *classification*, and the perceptron system is a *classifier*.

## 2-2 Bayesian Decision Rule (BDR) from Gaussian distribution.

As a special case of Gaussian for n-class problem, Let us have a given random vector ( $\underline{X}$ ) with mean vector depends on whether it belongs to class ( $k_i$ ) with non-diagonal positive semi-definite covariance matrix, means that the samples drawn from the classes are correlated as follow:

For class  $k_i$  ;  $E_i(\underline{X}) = m_i$  , and  $E_i[(\underline{X} - m_i)(\underline{X} - m_i)^T] = \Sigma_i$   
 $i=1,2,\dots,n$  classes

Such that  $|\Sigma|$  is the determinant of a non-singular matrix, then  $\Sigma^{-1}$  exists.

Now we can express the conditional probability density function of the random vector ( $\underline{X}$ ) as follows:

$$p(\underline{X} / k_i) = \frac{1}{(2\pi)^{n/2} |\Sigma_i|^{1/2}} \exp \left[ -\frac{1}{2} (\underline{X} - \underline{m}_i)^T \Sigma_i^{-1} (\underline{X} - \underline{m}_i) \right] \quad \dots (2)$$

The covariance matrix  $\Sigma_i$  is symmetric and positive semi-definite matrix having always a positive determinant; the diagonal element  $\sigma_{kk}$  is the variance of the  $k^{\text{th}}$  element of the pattern vectors. The off-diagonal element is the covariance of ( $x_j$  and  $x_k$ ) when they are stochastically independent  $\sigma_{jk} = 0$ . The decision function for class ( $k_i$ ) may be chosen as:

$$d_i(\underline{X}) = p(\underline{X} / k_i) p(k_i)$$

Since the normal density has an exponential form, then working with logarithmic function of this decision would be more convenient, in other words we may use the following form:

$$d_i(\underline{X}) = \log[p(\underline{X} / k_i) p(k_i)] = \log p(\underline{X} / k_i) + \log p(k_i) \quad \dots (3)$$

Substituting equation (2) in equation (3) yields:

$$d_i(\underline{X}) = \log p(k_i) - (n/2) \log 2\pi - (1/2) \log |\Sigma_i| - (1/2) \left[ (\underline{X} - \underline{m}_i)^T \Sigma_i^{-1} (\underline{X} - \underline{m}_i) \right]. \quad (4)$$

And also for class  $k_j$  then we get the decision equation as follows:

$$d_j(\underline{X}) = \log p(k_j) - (n/2) \log 2\pi - (1/2) \log |\Sigma_j| - (1/2) \left[ (\underline{X} - \underline{m}_j)^T \Sigma_j^{-1} (\underline{X} - \underline{m}_j) \right] \dots (5)$$

Subtracting equation (5) from equation (4) yields **Bayesian surface boundary decision equation** given by:

$$d_i(\underline{X}) - d_j(\underline{X}) = \{ \log p(k_i) - \log p(k_j) \} - (1/2) \{ \log |\Sigma_i| - \log |\Sigma_j| \} + \left\{ \underline{X}^T \Sigma_i^{-1} \underline{m}_i - \underline{X}^T \Sigma_j^{-1} \underline{m}_j \right\}$$

$$- (1/2) \left\{ \underline{m}_i^T \Sigma_i^{-1} \underline{m}_i - \underline{m}_j^T \Sigma_j^{-1} \underline{m}_j \right\} = 0 \quad \dots \dots \dots (6)$$

If the var-cov matrix for classes is common then:

$\Sigma_i = \Sigma$  for all  $i = 1, 2, \dots, K$   $K$  is a maximum number of classes then

$$d_i(\underline{X}) - d_j(\underline{X}) = \log p(k_i) - \log p(k_j) + \left\{ (\underline{X}^T) \Sigma^{-1} (\underline{m}_i - \underline{m}_j) \right\}$$

$$-(1/2) \left\{ \underline{m}_i^T \underline{\Sigma}^{-1} \underline{m}_i - \underline{m}_j^T \underline{\Sigma}^{-1} \underline{m}_j \right\} = 0 \quad \dots\dots\dots(7)$$

Now put  $i = 1$ , and  $j = 2$  then the decision boundary (separable equation) between two classes in equation (6) becomes:

$$d_1(\underline{X}) - d_2(\underline{X}) =$$

$$\{\log p(k_1) - \log p(k_2)\} - (1/2) \{\log |\underline{\Sigma}_1| - \log |\underline{\Sigma}_2|\} + \left\{ \underline{X}^T \underline{\Sigma}_1^{-1} \underline{m}_1 - \underline{X}^T \underline{\Sigma}_2^{-1} \underline{m}_2 \right\}$$

$$-(1/2) \left\{ \underline{m}_1^T \underline{\Sigma}_1^{-1} \underline{m}_1 - \underline{m}_2^T \underline{\Sigma}_2^{-1} \underline{m}_2 \right\} = 0 \quad (8)$$

Also, since the two classes have the same chance for appearance then:

$$p(k_i) = p(k) = 1/2 \quad \text{Then this quantity is ignored from the}$$

decision boundary or Bayesian surface decision between two classes and so the equation (8) is reduced to:

$$d_1(\underline{X}) - d_2(\underline{X}) =$$

$$-(1/2)\{\log|\Sigma_1|-\log|\Sigma_2|\}+X^T\{\Sigma_1^{-1}\underline{m}_1-\Sigma_2^{-1}\underline{m}_2\}-(1/2)\{\underline{m}_1^T\Sigma_1^{-1}\underline{m}_1-\underline{m}_2^T\Sigma_2^{-1}\underline{m}_2\}=0 \quad (9)$$

Then:  $d_1(\underline{X}) - d_2(\underline{X}) =$

$$\underline{X}^T\{\Sigma_1^{-1}\underline{m}_1-\Sigma_2^{-1}\underline{m}_2\}+(1/2)\{\underline{m}_2^T\Sigma_2^{-1}\underline{m}_2-\underline{m}_1^T\Sigma_1^{-1}\underline{m}_1\}-1/2\{\log|\Sigma_1|-\log|\Sigma_2|\}=0. \quad (10)$$

Now recalling equation (1) for the perceptron's mathematical formula:

$$\text{net}(y) = \sum x_i w_i + \theta$$

$$\text{Or in matrix notation } \text{net}(y) = \underline{X}^T \underline{W} + \theta \quad \dots\dots\dots(11)$$

Then if we suppose that the decision boundary above in equation (10) is the output for the perceptron or the network :

$$\text{Net}(y) = d_1(\underline{X}) - d_2(\underline{X}) \quad \dots\dots\dots(12)$$

Net(y) is the output of the perceptron =

$$\underline{X}^T\{\Sigma_1^{-1}\underline{m}_1-\Sigma_2^{-1}\underline{m}_2\}+(1/2)\{\underline{m}_2^T\Sigma_2^{-1}\underline{m}_2-\underline{m}_1^T\Sigma_1^{-1}\underline{m}_1\}-1/2\{\log|\Sigma_1|-\log|\Sigma_2|\}=0 \quad \dots\dots(13)$$

Comparing equation (11) with surface decision boundary in equation (10) we conclude that:

$$W = \left\{ \Sigma_1^{-1} \underline{m}_1 - \Sigma_2^{-1} \underline{m}_2 \right\}, \text{ and:}$$

$$\theta = (1/2) \left\{ \underline{m}_2^T \Sigma_2^{-1} \underline{m}_2 - \underline{m}_1^T \Sigma_1^{-1} \underline{m}_1 \right\} - (1/2) \left\{ \log |\Sigma_1| - \log |\Sigma_2| \right\} \quad .(14)$$

The equation (13) states that the **Byes classifier** (BSDE) for two Gaussian classes is really **perceptron** comparing with formula in equation (1).

### 2-3 Minimum Error Rate Classification

Generally in Bayesian decision theory for two classes/categories classification let  $\{k_1, \text{ and } k_2\}$  be two state of natures,  $\{d_1, d_2\}$  two possible actions, and  $\gamma(d_i / k_j)$  for  $(i, j = 1, 2)$  be the **loss** incurred for taking **action** ( $d_i$ ) when state of nature is ( $k_j$ ), then for a random vector ( $\underline{X}$ ) which so called **feature vector** the conditional risks can be written as :

$$\begin{aligned} R(d_1 / \underline{X}) &= \gamma_{11} P(k_1 / \underline{X}) + \gamma_{12} P(k_2 / \underline{X}) \\ R(d_2 / \underline{X}) &= \gamma_{21} P(k_1 / \underline{X}) + \gamma_{22} P(k_2 / \underline{X}) \quad \dots (15) \end{aligned}$$

The **minimum risk decision rule** to decide ( $k_i, i = 1, 2$ ) becomes :

$$\left[ \begin{array}{ll} k_1 & \text{if } (\gamma_{21} - \gamma_{11})P(k_1 / \underline{X}) > (\gamma_{12} - \gamma_{22})P(k_2 / \underline{X}) \dots \\ k_2 & \text{otherwise} \end{array} \right] \quad (16)$$

This corresponds to decide  $(k_1)$  if

$$\frac{P(\underline{X} / k_1)}{P(\underline{X} / k_2)} > \frac{(\gamma_{12} - \gamma_{22}) P(k_1)}{(\gamma_{21} - \gamma_{11}) P(k_2)} \dots \dots \dots (17)$$

The left hand side of inequality (17) is a likelihood ratio ( $L_{12}$ ), and the right hand side is a threshold ( $\theta$ ) that is independent of

observations in the vector  $(\underline{X})$ . Then:  $L_{12} = \frac{P(\underline{X} / k_1)}{P(\underline{X} / k_2)}$

$$\text{,and } \frac{(\gamma_{12} - \gamma_{22})P(k_1)}{(\gamma_{21} - \gamma_{11})P(k_2)} = \theta \dots (18)$$

Now for a given vector( $\underline{X}$ ) which is required to classify, and calculating the threshold value ( $\theta$ ) from equation (14), also calculating posteriors from equation (2) to find ( $L_{12}$ ) from relation (18), and at last comparing them one with each other. After achieving this comparison, it will be possible to take a ***minimum risk decision rule*** given by relation (17).

### 3- Application:

The application in this study is for making classification for two levels of income in Sulaimani (center) between two areas [(Azmar, class  $k_1$ ), and Shorsh, class  $k_2$ )] due to four variables as follows:

X1: Average of income, X2: Average of persons/Family, X3: Average of expenditure percentage from income), and X4: average of workers/family.

#### 3-1 Data Sampling

Cluster sampling technique has been used for collecting data, by dividing each area into (four) sectors. Each sector was divided into four lines. Five families (clusters) were selected randomly from each line. And the average for each variable that was described above was taken for these five families. So each sample containing (16 observations, each of them describes the four variables above, they came from the average of five families.



In order to achieve logical form for data, it was encoded to binary (0, 1) with respect to the arithmetic mean for each variable (if the value of the variable  $>$  mean, then it was encoded by 1, and otherwise was encoded by 0).

See the two following tables:

Table (1): First Encoded Sample(matrix) for Azmar ( class  $k_1$ ) Area in Sulaimani.

X1: Average income/Family	0	0	0	1	1	0	1	0	0	1	1	1	1	1	0	0
X2: Average of persons/family	1	0	0	1	0	0	0	0	1	0	1	1	0	1	1	0
X3:Average of expenditure /family By a percentage rate from income	1	1	1	1	1	0	1	0	0	0	0	0	0	1	1	1
X4:Average of workers /family	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0

Table (2): Second Encoded Sample for Shores area in Sulaimani.  
( $k_2$ ).

X1: Average income/Family	0	0	0	1	0	1	0	0	1	1	1	1	0	1	1	1
X2: Average of persons/family	1	1	0	0	1	1	1	1	0	0	1	0	0	0	0	1
X3:Average of expenditure/family	1	0	0	1	1	0	0	1	1	1	1	0	0	1	1	1

By a percentage rate from income																
X4:Average of workers /family	1	1	1	0	1	1	1	1	0	0	1	1	0	1	1	1

In order to apply (BSDE) as a linear perceptron, for each two above classes ( $k_1$ ), and ( $k_2$ ), the following quantities must be computed:

$$k_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \underline{m}_1 = (1/16) * \begin{bmatrix} 8 \\ 7 \\ 9 \\ 5 \end{bmatrix}$$

$$k_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}, \underline{m}_2 = (1/16) * \begin{bmatrix} 8 \\ 10 \\ 12 \end{bmatrix}$$

$$\Sigma_1 = (1/16)(k_1 k_1^T) - \underline{m}_1 \underline{m}_1^T =$$

$$= \begin{bmatrix} 0.2500 & 0.0313 & -0.0313 & 0.0313 \\ 0.0313 & 0.2461 & 0.0039 & 0.1133 \\ -0.0313 & 0.0039 & 0.2461 & -0.0508 \\ 0.0313 & 0.1133 & -0.0508 & 0.2148 \end{bmatrix}$$

$$\Sigma_1^{-1} = \begin{bmatrix} 4.1448 & -0.4069 & 0.4759 & -0.2759 \\ -0.4069 & 5.5241 & -0.7655 & -3.0345 \\ 0.4759 & -0.7655 & 4.4207 & 1.3793 \\ -0.2759 & -3.0345 & 1.3793 & 6.6207 \end{bmatrix} \quad \text{Det}$$

$$\Sigma_1 = 0.0022 \quad -0.2759 \quad -3.0345 \quad 1.3793 \quad 6.6207$$

Also for class  $k_2$  :

$$\Sigma_2 = (1 / 16)(k_2 k_2^T) - \underline{m_2 m_2^T} =$$

$$\Sigma_2^{-1} = \begin{bmatrix} 5.6073 & 2.3686 & -2.1269 & -0.5317 \\ 2.3686 & 7.0695 & -1.4502 & -4.3625 \\ -2.1269 & -1.4502 & 5.2205 & 1.3051 \\ -0.5317 & -4.3625 & 1.3051 & 8.3263 \end{bmatrix} \quad \text{Det}$$

$$\Sigma_2 = 0.0013 \quad -0.5317 \quad -4.3625 \quad 1.3051 \quad 8.3263$$

After calculating (W), and ( $\theta$ ) from two equation (14),and substituting them in equation (13, perceptron formula). Now For any given vector ( $\underline{X}$ ), it is easy to compute (BSDE).

$$\mathbf{W} = \begin{bmatrix} 2.0759 \\ 0.8345 \\ 2.8207 \\ 1.3793 \end{bmatrix} - \begin{bmatrix} 2.6103 \\ 0.6888 \\ 2.3202 \\ 4.5801 \end{bmatrix} = \begin{bmatrix} -1.5344 \\ 0.1457 \\ 0.5005 \\ -3.2008 \end{bmatrix}$$

And the threshold value  $\theta = - (1/2) [(- 6.1136) - (- 6.6745)] = - 0.28045$

Now using these two values, the vector (W), and ( $\theta$ ) that was calculated at last, and we suppose that we have a vector  $\underline{X}^T = [x_1, x_2, x_3, x_4]$ , for those variables described previously in data tables (1, and 2) without any knowledge for which class they were returned Class (a), or class (b), then by substituting these two values in equation (13) we get:

$$\text{Net} = (x_1 \ x_2 \ x_3 \ x_4) * \begin{bmatrix} -1.53440 \\ 0.145700 \\ 0.500500 \\ -3.32008 \end{bmatrix} - 0.28045 = 0$$

After knowing a true value of the random vector and transforming it to binary data (encoded), then if  $(\text{net} \geq 0)$  then the vector ( $\underline{X}$ ) is coming from class ( $k_1$ ), and otherwise is from class ( $k_2$ ).

Note (1): Each Vector ( $\underline{X}$ ) represents an averages ( $x_1, x_2, x_3, x_4$ ) from each two classes above.

Note(2) :  $m_1, m_2$  represent the mean vectors of class( $k_1$ ), and class( $k_2$ ), also  $\Sigma_1$ , and  $\Sigma_2$  are their var-covariance matrices, respectively.

### 3-2 Evaluation of classification

The major goal in (BDR), or generally in decision theory for classification, is to minimize the cost of classification, then to evaluate the result of classification that the chosen vector ( $\underline{X}$ ) is referring to one of two classes ( $k_1$  , or  $k_2$  ), one can calculate ( Likelihood ratio ) to test the power of classification, now in Bayesian classifier for a Gaussian environment, when it is compared with linear perceptron, let  $L_{12}$  , see inequality( 16 ) be a likelihood ratio that the vector ( $\underline{X}$ ) is coming from class ( $k_1$ ) but not from ( $k_2$ ).

Then from relation (17):

$$L_{12} = \frac{P(\underline{X} / k_1)}{P(\underline{X} / k_2)}$$

And remember that ( $\theta$ ) was the threshold in expression (14) that has been calculated previously ( $\theta = -0.28045$ ). Now: the three following results can be observed:

- 1/ Assign the chosen vector ( $\underline{X}$ ) to class ( $k_1$ ) if  $L_{12}(\underline{X}) > \theta$ .
- 2/ Assign the chosen vector ( $\underline{X}$ ) to class ( $k_2$ ) if  $L_{12}(\underline{X}) < \theta$ .
- 3/ But if  $L_{12}(\underline{X}) = \theta$ , then an arbitrary decision could be made.

The assurance above is a test procedure of the fitness for classification model that may be decided by the investigator, as an additional comparison for that was done by using equation (14), and it gives the same result.

#### 4- Results and conclusions

Through this study the following important results and conclusions are achieved:

- 1/ A generalized Bayesian surface decision equation (equation 12) was established as a modified classifier not only for (identically independent) vectors as it was suggested by **Muller, P. & Insua, D.R, (1995)**, (equation 7), but also for those which have identical distributions, specially in Gaussian case.
- 2/ The generalization established in equation (12) is also equivalent to the linear perceptron illustrated in equation (13). And it can be used as a linear separable function classifier to determine the source class of a given vector ( $\underline{X}$ ) taken among two separable classes.

3/ The threshold value in equation (14) is equivalent to the quantity in inequality (17) as shown in the relation (18), this means, one can conclude that the minimum risk decision rule might be achieved either by comparing (BSDE) with the likelihood ratio given in relation (17), or with the value of threshold ( $\theta$ ).

4/ Since the non-zero threshold value had been used in classifier perceptron(LSE) then, in this study it was observed that the real data must be applied in (BSDE), but when it was generalized or modified, as it had been used, see equation(12, and 13), then the encoded(binary) data (0,1) is better or necessary must be applied in any (BSDE) problem, as it was shown in the application).

5/ Fitting (BSDE) in this study is very important and necessary for the economic planners in national income distribution, to determine a separable equation for income classes, specially those classes having nearly behaviors which make separation between them without (BSDE) is more difficult, to decide, and refer to a given set of records for their origin classes.

6/ It is not necessary for classes to have equal chances for appearance to make (BSDE) capable as shown above that:  $p(k_i) = p(k) = 1/2$ , but they may take different chances, with property  $P(k_1) + P(k_2) = 1$  must be observed always.

7/ It is concluded that using the modified (BSDE) with linear perceptron in equation(14) for separating and classifying classes is more applicable than using modified (BSDE) alone, specially for those classes which seem to be equally likely in their behaviors.

## References

- 1/ C. Jason, "Baysian Decision Theory." Jhon-Wiely & Sons, London,(2009).
- 2/G. Arnauld, V. Corinne, and V. Jean-Paul."Multivariate mathematical morphology and Bayesian classifier: Application to color and medical emages". SPIE-IS &T. Vol.6812, 681203. (2008).
- 3/ G. L.Webb, J. Boughton, and Z.Wang "Not so naive bayes: Aggregating one-dependence estimators". J. Machine Learning 58(1):5-24, Jan-(2005).
- 4/ Moran. S, He. Y, and L. Kecheng." Choosing the best Baysian classifier: An imperical study."IAENG International Journal of Computer science. 34:4, IJCSNS\_36\_4\_09. 19-nov-(2009).
- 5/ P. Bhargavi, and S. Jyothi." Appling Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils" IJCSNS, VOL.9 No.8, August-2009.
- 6/ S. Moran, H.Yulan, and L. Kecheng," Choosing the best Bayesian Classifier: An Imperial Study" IJCS\_36\_4\_09, Nov. (2009).
- 7/ T. N.Shankar." Neural networks" University Science Press. Dalhi.India (2009).