Chapter 3

Classifying by Bayesian Method and Some Applications

Tai Vovan

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.70052

Abstract

This chapter sums up and proposes some results related to classification problem by Bayesian method. We present the classification principle, Bayes error, and establish its relationship with other measures. The determination for Bayes error in reality for one and multi-dimensions is also considered. Based on training set and the object that we need to classify, an algorithm to determine the prior probability that can make to reduce Bayes error is proposed. This algorithm has been performed by the MATLAB procedure that can be applied well with real data. The proposed algorithm is applied in three domains: biology, medicine, and economics through specific problems. With different characteristics of applied data sets, the proposed algorithm always gives the best results in comparison to the existing ones. Furthermore, the examples show the feasibility and potential application in reality of the researched problem.

Keywords: Bayesian method, classification, error, prior, application

1. Introduction

Classification problem is one of the main subdomains of discriminant analysis and closely related to many fields in statistics. Classification is to assign an element to the appropriate population in a set of known populations based on certain observed variables. It is an important development direction of multivariate statistics and has applications in many different fields [25, 27]. Recently, this problem is interested by many statisticians in both theories and applied areas [14–18, 22–25]. According to Tai [22], we have four main methods to solve the classification problem: Fisher method [6, 12], logistic regression method [8], support vector machine (SVM) method [3], and Bayesian method [17]. Because Bayesian method does not require normal condition for data and can classify for two and more populations it has many advantages [22–25]. Therefore, it has been used by many scientists in their researches.

Given k populations $\{w_i\}$, with probability density functions (pdfs) and the prior probabilities respectively $\{f_i\}$ and $\{q_i\}$, $i = 1, 2,\ldots, k$, where $q_i \in (0; 1)$, $\sum_{i=1}^{k} q_i = 1$. Pham–Gia et al. [17] used the maximum function of pdfs as a tool to study about Bayesian method and obtained important results. The classification principle and Bayes error were established based on the $g_{\max}(x) = \max\{q_1f_1(x), q_2f_2(x), \ldots, q_kf_k(x)\}$. The relationship between the upper and lower bounds of the Bayes error and the L^1—distance of the pdfs and the overlap coefficient of the pdfs—were established. The function $g_{\max}(x)$ played a very important role in the classification problem by Bayesian method and Pham–Gia et al. [17] continued to do research on it. Using the MATLAB software, Pham–Gia et al. [18] succeeded in identifying $g_{\max}(x)$ for some cases of the bivariate normal distribution. With similar development, Tai [22] has proposed the L^1—distance of the $\{q_if_i(x)\}$—and established its relationship with Bayes error. This distance is also used to calculate Bayes error as well as to classify new element. This research has been applied in classifying ability to repay debt of bank customers. However, we think that the survey of two Bayesian approach relevant research was not yet completed. There are some relations between Bayes error and other statistical measures.

Bayesian method has many advantages. However, to our knowledge, the field of applications of this method in practice is narrower than other methods. We can find many applications in banking and medicine using Fisher method, SVM method, logistic method [1, 3, 8, 12]. Recently, all statistics software can effectively and quickly process the classification of large data sets and multivariate statistics using either three of the methods mentioned above, whereas the Bayesian method does not have this advantage. The cause of this problem is the ambiguity in determining prior probability, in estimating pdfs, and the complexity in calculating Bayes error. Although all these issues have been discussed by many authors, the optimal methods have yet to be found [22, 25]. In this chapter, we consider to estimate the pdf and to calculate Bayes error to apply in reality. We will present the problem on how to determine the prior probability in this chapter. In case of noninformation, we normally choose prior probabilities by uniform distribution. If we have some types of past data or training set, the prior probabilities are estimated either by Laplace method: $q_i = (n_i + n/k)/(N + n)$ or by the frequencies of the sample: $q_i = n_i/N$, where n_i and N are the number of elements in the ith population and training set, respectively, n is the number of dimensions, and k is the number of groups. The above-mentioned approaches have been studied and applied by many authors [14, 15, 22, 25]. We will also propose an algorithm to determine prior probability based on the training set, classified objective, and fuzzy cluster analysis. The proposed algorithm is applied in some specific problems of biology, medicine, and economics and has advantages over existing approaches. All calculations are performed by MATLAB procedures.

The next section of this chapter is structured as follows. Section 2 presents the classification principle and Bayes error. Some results of the Bayes error are also established in this section. Section 3 resolves the related problems in real application of the Bayes method. There are estimation of pdfs and determination of Bayes error in case of one dimension and multidimension. This section also proposes an algorithm to determine prior probability. Section 4 applies the proposed algorithm in real problems and compares outcome results to those obtained using existing approaches. Section 5 concludes this chapter.

2. Classifying by Bayesian method

The classification problem by Bayesian method has been presented in many documents [15, 16, 27], where the classification principle and the Bayes error are established based on Bayes theorem. In this section, we present them via the maximum function of $q_i f_i(x)$, $i = 1, 2, \ldots, k$ that they have advantages over existing approaches in real application [17, 18, 21–25]. This section also establishes the upper and lower bounds of the Bayes error and the relationships of Bayes error with other measures in statistical pattern recognition.

2.1. Classification principle and Bayes error

Given k populations $w_1, w_2, \ldots, w_k$ with $q_i \in (0;1)$ and $f_i(x)$ are the prior probability and pdf of ith population, respectively, $i = 1, 2, \ldots, k$. According to Pham–Gia et al. [17], element x_0 will be assigned to w_i if

$$g_i(x_0) = g_{\max}(x_0),\ \ i = 1, 2, \ldots, k \tag{1}$$

where $g_i(x) = q_i f_i(x), g_{\max}(x) = \max\{q_1 f_1(x), q_2 f_2(x), \ldots, q_k f_k(x)\}$.

Bayes error is given by the formula:

$$Pe_{1,2,\ldots,k}^{(q)} = \sum_{i=1}^{k} \int\limits_{R^n \setminus R_i^n} q_i f_i dx = 1 - \sum_{i=1}^{k} \int\limits_{R_i^n} q_i f_i(x) dx, \tag{2}$$

where $R_i^n = \left\{x | q_i f_i(x) > q_j f_j(x), \forall i \neq j, i, j = 1, 2, \ldots, k\right\}, (q) = (q_1, q_2, \ldots, q_k)$.

From Eq. (2), we can prove the following result:

$$\begin{aligned} Pe_{1,2,\ldots,k}^{(q)} &= \sum_{j=1}^{k} \int\limits_{R^n \setminus R_j^n} q_j f_j(x) dx \\ &= \sum_{j=1}^{k} \left[\int\limits_{R^n} q_j f_j(x) dx - \int\limits_{R_j^n} \max_{1 \le l \le k}\{q_l f_l(x)\} dx \right] \\ &= \int\limits_{R^n} \sum_{j=1}^{k} q_j f_j(x) dx - \sum_{j=1}^{k} \int\limits_{R_j^n} \max_{1 \le l \le k}\{q_l f_l(x)\} dx \\ &= 1 - \int\limits_{R^n} \max_{1 \le l \le k}\{q_l f_l(x)\} dx \end{aligned}$$

or

$$Pe_{1,2,\ldots,k}^{(q)} = 1 - \int\limits_{R^n} g_{\max}(x) dx. \tag{3}$$

The correct probability is determined by $Ce^{(q)}_{1,2,\ldots,k} = 1 - Pe^{(q)}_{1,2,\ldots,k}$.

For $k = 2$, we have

$$Pe^{(q,1-q)}_{1,2} = \int_{R^n} \min\{qf_1(x), (1-q)f_2(x)\}dx = \lambda^{(q,1-q)}_{1,2} = \frac{1}{2}\left[1 - \|qf_1, (1-q)f_2\|_1\right], \tag{4}$$

where

$\lambda^{(q,1-q)}_{1,2}$ is the overlap area measure of $qf_1(x)$ and $(1-q)f_2(x)$ and $\|qf_1, (1-q)f_2\|_1 = \int_{R^n} |qf_1(x) - (1-q)f_2(x)|dx$.

2.2. Some results about Bayes error

Theorem 1. Let $f_i(x)$, $i = 1, 2, \ldots, k$, $k \geq 3$ be k pdfs defined on R^n, $n \geq 1$, $q_i \in (0;1)$. We have the relationships of Bayes error with other measures as follow:

i.
$$Pe^{(q)}_{1,2,\ldots,k} \leq 1 - \frac{1}{k-1}\left(1 - \prod_{j=1}^{k} q_j^{\alpha_j} D_T(f_1, f_2, \ldots, f_k)^{\alpha}\right), \tag{5}$$

ii.
$$Pe^{(q)}_{1,2,\ldots,k} \leq \sum_{i<j} q_i^{\beta} q_j^{1-\beta} D_T\left(f_i, f_j\right)^{(\beta,1-\beta)}, \tag{6}$$

iii.
$$\left\{(k-1) - \sum_i \sum_j \|g_i, g_j\|_1\right\}/k \leq Pe^{(q)}_{1,2,\ldots,k} \leq 1 - (1/2)\max_{i<j}\left\{\|g_i, g_j\|_1\right\} - \min_i\{q_i\}, \tag{7}$$

iv.
$$0 \leq Pe^{(q)}_{1,2,\ldots,k} \leq \max_i\{q_i\}, \tag{8}$$

where

$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k);\ \alpha_j, \beta \in (0,1), \sum_{j=1}^{k} \alpha_j = 1$, $i, j = 1, 2, \ldots, k$, and

$D_T(f_1, f_2, \ldots, f_k)^{\alpha} = \int_{R^n} \prod_{j=1}^{k} \left[f_j(x)\right]^{\alpha_j} dx$ is affinity of Toussaint [26].

http://dx.doi.org/10.5772/intechopen.70052

Proof:

i. For each $j = 1,2,\ldots,k$, we have

$$\left(\sum_{j=1}^{k} q_j f_j\right)^{\alpha_i} \geq (q_i f_i)^{\alpha_i}, i = 1, 2, \ldots, k.$$

Therefore,

$$\left(\sum_{j=1}^{k} q_j f_j\right)^{\alpha_1+\alpha_2+\ldots+\alpha_k} \geq \prod_{j=}^{k}\left(q_j f_j\right)^{\alpha_j} \Leftrightarrow \sum_{j=1}^{k} q_j f_j \geq \prod_{j=}^{k}\left(q_j f_j\right)^{\alpha_j}. \tag{9}$$

On the other hand,

$$\left(\min_{1\leq j\leq k}\left\{q_j f_j\right\}\right)^{\alpha_1} \leq (q_1 f_1)^{\alpha_1}, \ldots\ldots, \left(\min_{1\leq j\leq k}\left\{q_j f_j\right\}\right)^{\alpha_k} \leq (q_k f_k)^{\alpha_k},$$

So

$$\left(\min_{1\leq j\leq k}\left\{q_j f_j\right\}\right)^{\alpha_1+\cdots+\alpha_k} \leq \prod_{j=1}^{k}\left(q_j f_j\right)^{\alpha_j}.$$

or

$$\min_{1\leq j\leq k}\left\{q_j f_j\right\} \leq \prod_{j=1}^{k}\left(q_j f_j\right)^{\alpha_j}. \tag{10}$$

Combining Eqs. (9) and (10), we obtain

$$0 \leq \sum_{j=1}^{k} q_j f_j - \prod_{j=1}^{k}\left(q_j f_j\right)^{\alpha_j} \leq \sum_{j=1}^{k} q_j f_j - \min_{1\leq j\leq k}\left\{q_j f_j\right\}.$$

Because $\sum_{j=1}^{k} q_j f_j - \min_{1\leq j\leq k}\left\{q_j f_j\right\}$ includes $(k-1)$ terms, we have

$$\sum_{j=1}^{k} q_j f_j - \min_{1\leq j\leq k}\left\{q_j f_j\right\} \leq (k-1)\max_{1\leq j\leq k}\left\{q_j f_j\right\}.$$

Thus,

$$0 \leq \sum_{j=1}^{k} q_j f_j - \prod_{j=1}^{k}\left(q_j f_j\right)^{\alpha_j} \leq (k-1)\max_{1\leq j\leq k}\left\{q_j f_j\right\}.$$

Integrating the above relation, we obtain:

$$1 - \prod_{j=1}^{k} q_j^{\alpha_j} D_T\left(f_1, f_2, \ldots, f_k\right)^{\alpha} \le (k-1) \int_{R^n} g_{\max}(x)dx. \tag{11}$$

Using $\int_{R^n} g_{\max}(x) = 1 - Pe_{1,2,\ldots,k}^{(q)}$ for Eq. (11), we have Eq. (5).

ii. From Eq. (2), we have

$$\begin{aligned} Pe_{1,2,\ldots,k}^{(q)} &= \sum_{j=1}^{k} \int_{R^n \setminus R_j^n} q_j f_j(x)dx \\ &= \sum_{j=1}^{k} \sum_{j \ne i} \int_{R_j^n} \min\left\{q_i f_i(x),\ q_j f_j(x)\right\}dx \\ &= \sum_{i<j} \int_{R_i^n} \min\left\{q_i f_i(x), q_j f_j(x)\right\}dx. \end{aligned}$$

Since

$$\left[\min\left\{q_i f_i(x), q_j f_j(x)\right\}\right]^{\beta} \le \left(q_i f_i\right)^{\beta} \text{ and } \left[\min\left\{q_i f_i(x), q_j f_j(x)\right\}\right]^{1-\beta} \le \left(q_i f_i\right)^{1-\beta},$$

then

$$\min\left\{q_i f_i(x), q_j f_j(x)\right\} \le \left(q_i f_i\right)^{\beta} \left(q_j f_j\right)^{1-\beta}.$$

Integrating the above inequality, we obtain:

$$Pe_{1,2,\ldots,k}^{(q)} \le \sum_{i<j} \int_{R_i^n} \left[\left(q_i f_i(x)\right)^{\beta} \left(q_j f_j(x)\right)^{1-\beta}\right]dx \le \sum_{i<j} q_i^{\beta} q_j^{1-\beta} D_T\left(f_i, f_j\right)^{(\beta, 1-\beta)} dx.$$

iii. We have

$$\int_{R^n} \max\{g_1(x), g_2(x), \ldots, g_k(x)\}dx \ge \max_{i<j} \int_{R^n} \max\left\{g_i(x), g_j(x)\right\}dx$$

On the other hand,

$$\begin{aligned} \max_{i<j}\left\{\int_{R^n} \max\{g_i(x), g_j(x)\}dx\right\} &= \max_{i<j}\left\{\frac{1}{2}\|g_i, g_j\|_1 + \frac{1}{2}(q_i + q_j)\right\} \\ &\ge \max_{i<j}\left\{\frac{1}{2}\|g_i, g_j\|_1\right\} + \min_{i<j}\left\{\frac{1}{2}(q_i + q_j)\right\} \\ &\ge \max_{i<j}\left\{\frac{1}{2}\|g_i, g_j\|_1\right\} + \min_{i<j}\left\{(q_1, q_2, \ldots, q_k)\right\}. \end{aligned}$$

http://dx.doi.org/10.5772/intechopen.70052

Hence,

$$\int_{R^n} g_{\max}(x)dx \geq \frac{1}{2}\max_{i<j}\left\{\|g_i, g_j\|_1\right\} + \min_{i<j}\{(q_1, q_2, \dots, q_k)\}. \tag{12}$$

We also have

$$\sum_{i<}\sum_{j}|g_i - g_j| \geq \sum_{j=1}^{k}\left[\max\{g_1, g_2, \cdots g_k\} - g_j\right]$$
$$= k\left[\max\{g_1, g_2, \cdots g_k\}\right] - \sum_{j=1}^{k} g_j$$

Therefore,

$$\max\{g_1, g_2, \cdots g_k\} \leq \frac{1}{k}\sum_{i<}\sum_{j}|g_i - g_j| + \frac{1}{k}\sum_{j=1}^{k} g_j. \tag{13}$$

Since

$\int_{R^n} g_i(x)dx = q_i$ and $\sum_{i=1}^{k} q_i = 1$, the inequality Eq. (13) becomes:

$$\int_{R^n} g_{\max}(x)dx \leq \frac{1}{k}\sum_{i<j}\|g_i, g_j\|_1 + \frac{1}{k}. \tag{14}$$

Replacing $\int_{R^n} g_{\max}(x) = 1 - Pe_{1,2,\dots,k}^{(q)}$ to Eqs. (12) and (14), we have Eq. (7).

iv. We have

$q_i f_i(x) \leq \max\{q_1 f_1(x), q_2 f_2(x), \dots, q_k f_k(x)\} \leq \sum_{i=1}^{k} q_i f_i(x)$ for all $i = 1,\dots,k$.

Integrating the above relation, we obtain:

$$q_i \leq \int_{R^n} g_{\max}(x)dx \leq 1.$$

Above inequality is true for all $i = 1,\dots,k$, so

$$\max\{q_i\} \leq \int_{R^n} g_{\max}(x)dx \leq 1.$$

Replacing $\int_{R^n} g_{\max}(x) = 1 - Pe_{1,2,\dots,k}^{(q)}$ in above relation, we have Eq. (8).

From the result of Eqs. (5) and (6), with $\alpha_1 = \alpha_2 = \dots = \alpha_k = 1/k,$, we have the relationship between Bayes error and affinity of Matusita [11]. Especially, when $k = 2$, we have the relationship between $Pe_{1,2}^{(q,1-q)}$ and Hellinger's distance.

In addition, we also have the relation between Bayes error and overlap coefficients as well as L^1–distance of $\{g_1(x), g_2(x), \ldots, g_k(x)\}$ (see Ref. [22]). For special case: $q_1 = q_2 = \ldots = q_k = 1/k$, we had established expressions about relations between Bayes error and L^1–distance of $\{f_1(x), f_2(x), \ldots, f_k(x)\}$, $Pe^{(1/k)}_{1,2,\ldots,k}$ and $Pe^{(1/(k+1))}_{1,2,\ldots,k+1}$ (see Ref. [17]).

3. Related problems in applying of Bayesian method

To apply Bayesian method in reality, we have to resolve three main problems: (i) Determine prior probability, (ii) compute Bayes error, and (iii) estimate pdfs. In this section, we propose an algorithm to solve for (i) based on fuzzy cluster analysis and classified objective that can reduces Bayes error in comparing with traditional approaches. For (ii), Bayes error is established by closed expression for general case and determine it by an algorithm to find maximum function of $g_i(x)$, $i = 1, 2, \ldots, k$ for one dimension case. The quasi-Monte Carlo method is proposed to compute Bayes error in this section. For (iii), we review the problem to estimate pdfs by kernel function method where the bandwidth parameter and kernel function are specified.

3.1. Prior probability

In the n-dimensions space, given N populations $N^{(0)} = \left\{W_1^{(0)}, W_2^{(0)}, \ldots, W_N^{(0)}\right\}$ with data set $Z = [z_{ij}]_{nxN}$. Let matrix $U = [\mu_{ik}]_{c\times n}$, where μ_{ik} is probability of the kth element belonging to w_i. We have $\mu_{ik} \in [0, 1]$ and satisfies the following conditions:

$$\sum_{i=1}^{c} \mu_{ik} = 1,\ 0 < \sum_{k=1}^{N} \mu_{ik} < N,\ 1 \le i \le c,\ 1 \le k \le N.$$

We call

$$M_{zc} = \left\{ U = [\mu_{ik}]_{cxN} | \mu_{ik} \in [0, 1], \forall i, k; \sum_{i=1}^{c} \mu_{ik} = 1, \forall k; 0 < \sum_{k=1}^{N} \mu_{ik}, \forall i \right\} \tag{15}$$

be fuzzy partitioning space of k populations,

$D^2_{ikA} = \|z_k - \mathrm{v}_i\|^2_{\mathrm{A}} = (z_k - \mathrm{v}_i)^{\mathrm{T}} \mathrm{A} (z_k - \mathrm{v}_i)$ is the matrix whose element d^2_{ik} is the square of distance from the object z_k to the ith representative population. This representative is computed by the following formula:

$$\mathrm{v}_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m z_k}{\sum_{k=1}^{N} (\mu_{ik})^m},\ 1 \le i \le c, \tag{16}$$

where $m \in [1,\infty)$ is the fuzziness parameter.

Given the data set Z including c known populations $w_1, w_2, \ldots, w_c$. Assume x_0 is an object that we need to classify. To identify the prior probabilities when classifying x_0, we propose the following prior probability by fuzzy clustering (PPC) algorithm:

Algorithm 1. Determining prior probability by fuzzy clustering (PPC)

Input: The data set $Z = [z_{ij}]_{n\times N}$ of c populations $\{w_1, w_2, \ldots, w_c\}$, x_0, ε, m and the initial partition matrix $U = U^{(0)} = [\mu_{ij}]_{c\times N+1}$, where $\mu_{ij} = 1$ if the jth object belongs to the w_i and $\mu_{ij} = 0$ for the opposite, $i = \overline{1, c}; j = \overline{1, N}$, $\mu_{ij} = 1/c$ for $j = N + 1$.

Output: The prior probability $\mu_{i(N+1)}, i = 1, 2, \ldots c$.

Repeat:

Find the representative object of w_i: $v_i = \dfrac{\sum_{k=1}^{N} (\mu_{ik})^m z_k}{\sum_{k=1}^{N} (\mu_{ik})^m}, \ 1 \le i \le c$

Compute the matrix $[D_{ik}]_{c\times N+1}$ (the pairwise distance between objects and representative objects).

Update the new partition matrix $U^{(\text{new})}$ by the following principle:

If $D_{ik} > 0$ for all $i = 1, 2, \ldots, c; k = 1, 2, \ldots, N + 1$ then

$$\mu_{ik}^{(\text{new})} = \frac{1}{\sum_{j=1}^{c} (D_{ik}/D_{jk})^{2/(m-1)}}, i \neq j = 1, 2, \ldots, c$$

Else, $\mu_{ik}^{(\text{new})} = 0$

End;

Compute $S = \|U^{(\text{new})} - U\| = \max_{ik}\left(\left|\mu_{ik}^{(\text{new})} - \mu_{ik}\right|\right)$

$U = U^{(\text{new})}$

Until $S < \varepsilon$;

The prior probability $\mu_{i(N+1)}, i = 1, 2, \ldots c$ (the final column of the matrix U);

In the above algorithm, we have:

i. ε is a really small number and is chosen arbitrarily. The smaller ε is, the more iterations time is taken. In the examples of this chapter, we choose $\varepsilon = 0.001$.

ii. The distance matrix D_{ik} depends on the norm-inducing matrix A. When $A = I$, D_{ik} is the matrix of Euclidean distances. Besides, there are several choices of A, such as diagonal matrix or the inverse of the covariance matrix. In this chapter, we chose the Euclidean distances in the numerical examples and applications.

iii. m is the fuzziness parameter, when $m = 1$, the fuzzy clustering becomes the nonfuzzy clustering. When $m \to \infty$, the partition becomes completely fuzzy $\mu_{ik} = 1/c$. The determining of this parameter, which affects the analysis result, is difficult. Even though Yu et al. [28] proposed two rules to determine the supermom of m for clustering problems, the searching of the specific m was done by meshing method (see [2, 4, 5, 9] for more details). By this process, the best m among several of given values will be chosen. In this chapter, m is also identified by meshing method for the classification problem. The best integer m between 2 and 10 will be used.

At the end of the PPC algorithm, we obtain the prior probabilities of x_0 based on the last column of the partition matrix U ($\mu_{i(N+1)}, i = 1, 2, \ldots c$). The PPC algorithm helps us determine the prior probabilities via the closeness degree between the classified object and the populations. Each object will receive its suitable prior probabilities.

In this chapter, Bayesian method with prior probabilities calculated by the uniform distribution approach, the ratio of samples approach, the Laplace approach, and the proposed PPC algorithm approach are respectively called BayesU, BayesR, BayesL, and BayesC.

Example 1. Given the studied marks (scale 10 grading system) of 20 students. Among them, nine students have marks that are lower than 5 (w_1: fail the exam) and 11 students have marks that are higher than 5 (w_2: pass the exam). The data are given in **Table 1**.

Assume that we need to classify the ninth object, $x_0 = 4.3$, to one in two populations. Using the PPC algorithm, we have the following final partition matrix:

$$\begin{pmatrix} 0.957 & 0.973 & 0.981 & 0.993 & 1 & 0.997 & 0.997 & 0.830 & 0.321 & 0.290 & 0.158 & 0.1 & 0.1 & 0.01 & 0.009 & 0.037 & 0.045 & 0.054 & 0.062 & 0.724 \\ 0.043 & 0.027 & 0.019 & 0.007 & 0 & 0.003 & 0.003 & 0.170 & 0.679 & 0.710 & 0.842 & 0.9 & 0.9 & 0.99 & 0.991 & 0.963 & 0.955 & 0.946 & 0.938 & 0.276 \end{pmatrix}$$

This matrix shows the prior probabilities when assigning the ninth object to w_1 and w_2 are 0.724 and 0.276, respectively. Meanwhile, the prior probabilities determined by BayesU, BayesR, and BayesL are (0.5; 0.5), (0.421; 0.579), and (0.429; 0.571), respectively.

From the data in **Table 1**, we might estimate the pdfs $f_1(x)$ and $f_2(x)$ and compute the values $q_1 f_1(x)$ and $q_2 f_2(x)$, where q_1 and q_2 are the calculated prior probabilities. The results of classifying x_0 by four approaches: BayesU, BayesR, BayesL, and BayesC are given in **Table 2**.

Because the actual population of x_0 is w_1, only BayesC gives the true result. The Bayes error of BayesC is also the smallest. Thus, in this example, the proposed method improves the drawback of the traditional method in determining the prior probabilities.

Objects	Marks	Groups	Objects	Marks	Groups
1	0.6	w_1	11	5.6	w_2
2	1.0	w_1	12	6.1	w_2
3	1.2	w_1	13	6.4	w_2
4	1.6	w_1	14	6.4	w_2
5	2.2	w_1	15	7.3	w_2
6	2.4	w_1	16	8.4	w_2
7	2.4	w_1	17	9.2	w_2
8	3.9	w_1	18	9.4	w_2
9	4.3	w_1	19	9.6	w_2
10	5.5	w_2	20	9.8	w_2

Table 1. The studied marks of 20 students and the actual classifications.

Methods	Priors	gmax(x0)	Populations	Bayes errors
BayesU	(0.5; 0.5)	0.0353	2	0.0538
BayesR	(0.421; 0.579)	0.0409	2	0.0558
BayesL	(0.429; 0.571)	0.0403	2	0.0557
BayesC	(0.724; 0.276)	0.0485	1	0.0241

Table 2. The results when classifying the ninth object.

3.2. Determining Bayes error

Theorem 2. Let $f_i(x)$, $i=1, 2, \ldots, k$, $k \geq 3$ be k pdfs defined on R^n, $n \geq 1$ and let $q_i \in (0;1)$,

$$\begin{cases} R_1^n = \left\{x \in R^n : q_1 f_1(x) > q_j f_j(x), 2 \leq j \leq k\right\}, \\ R_k^n = \left\{x \in R^n : q_k f_k(x) > q_j f_j(x), 1 \leq j \leq k\right\}, \\ R_l^n = \left\{x \in R^n : q_i f_i(x) > q_l f_l(x), 1 \leq i \leq k, 2 \leq l \leq k-1, i \neq l\right\}. \end{cases} \tag{17}$$

The Bayes error is determined by

$$Pe_{1,2,\ldots,k}^{(q)} = 1 - \int\limits_{R_1^n} q_1 f_1(x)dx - \sum_{l=2}^{k-1} \int\limits_{R_l^n} q_l f_l(x)dx - \int\limits_{R_k^n} q_k f_k(x)dx. \tag{18}$$

Proof:

To obtain Eq. (18), we need to prove two following results:

$$R_i^n \cap R_j^n = \phi, \ (1 \leq i \neq j \leq k)$$

and $\bigcup\limits_{i=1}^{k} R_i^n = R_1^n \cup \bigcup\limits_{i=2}^{k-1} R_i^n \cup R_k^n = R^n, \quad f_{\max}(x) = f_i(x), \quad \forall x \in R_i^n.$

Let $\overline{A} = R^n \setminus A$, we have

$$\overline{R}_{ij} = \{x \in R^n : q_i f_i(x) \leq q_j f_j(x)\}, R_{ij} = \{x \in R^n : q_i f_i(x) > q_j f_j(x)\}, (1 \leq i, j \leq k).$$

From Eq. (17), we obtain

$$R_1^n = \bigcap_{j=2}^{k} R_{1j}, R_l^n = \bigcap_{i \neq k} \overline{R}_{il}, (2 \leq l < k).$$

Therefore,

$$R_1^n \cap R_l^n = (\bigcap_{j=2}^{k} R_{ij}) \cap (\bigcap_{i \neq k} \overline{R}_{il}) \subset R_{il} \cap \overline{R}_{1l} = \phi \Rightarrow R_1^n \cap R_l^n = \phi, \ (2 \leq l < k).$$

On the other hand, from antithesis style of D'Morgan, we have

$$\overline{R_1^n \cup R_l^n} = (\bigcup_{j=2}^{n} \overline{R}_{ij}) \cup (\bigcup_{i \neq k} R_{il}) \subset \overline{R}_{il} \cap R_{1l} = \phi \Rightarrow R_1^n \cup R_l^n = R^n, (2 \leq l < k).$$

Similarly,

$$R_k^n \cap R_l^n = \phi, (2 \leq l < k), R_1^n \cap R_k^n = \phi,$$

so

$$\bigcup_{i=1}^{k} R_i^n = R^n, \ \cup (\bigcup_{l=2}^{k-1} R_l^n) \cup R_k^n = R_1^n \cup (\bigcup_{l=2}^{k-1} R_l^n) \cup R_k^n$$
$$= (\bigcup_{l=2}^{k-1} R_1^n \cup R_l^n) \cup (\bigcup_{l=2}^{k-1} R_k^n \cup R_l^n) = R^n \cup R^n = R^n \Rightarrow \bigcup_{i=1}^{k} R_i^n = R^n.$$

In addition, from Eq. (17), we can directly find out

$$g_{\max}(x) = g_i(x), \forall x \in R_i^n, (1 \leq i \leq k).$$

For $k = 2$, $q_1 = q_2 = 1/2$, we consider the two following special cases:

i. If $f_1(x)$ and $f_2(x)$ are two one-dimension normal pdfs ($N(\mu_i, \sigma_i)$, $i = 1, 2$), without loss of generality, we suppose that $\mu_1 < \mu_2$ (for $\mu_1 \neq \mu_2$), $\sigma_1 < \sigma_2$ (for $\sigma_1 \neq \sigma_2$), then

$$Pe_{1,2}^{(1/2,1/2)} = \begin{cases} \dfrac{1}{2}\left[\int\limits_{-\infty}^{x_1} f_2(x)dx + \int\limits_{x_1}^{+\infty} f_1(x)dx\right], & \text{if} \sigma_1 = \sigma_2, \\ \dfrac{1}{2}\left[\int\limits_{-\infty}^{x_2} f_1(x)dx + \int\limits_{x_2}^{x_3} f_2(x)dx + \int\limits_{x_3}^{+\infty} f_1(x)dx\right], & \text{if} \sigma_1 < \sigma_2, \end{cases}$$

where

$$x_1 = \frac{\mu_1 + \mu_2}{2}, \ x_2 = \frac{(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) - \sigma_1\sigma_2\sqrt{(\mu_1 - \mu_2)^2 + K}}{\sigma_2^2 - \sigma_1^2},$$
$$x_3 = \frac{(\mu_1\sigma_2^2 - \mu_2\sigma_1^2) + \sigma_1\sigma_2\sqrt{(\mu_1 - \mu_2)^2 + K}}{\sigma_2^2 - \sigma_1^2}, \ K = 2(\sigma_2^2 - \sigma_1^2)\ln\left(\frac{\sigma_2}{\sigma_1}\right) \geq 0.$$

For $\mu_1 = \mu_2 = \mu$, the above result becomes:

$$Pe_{1,2}^{(1/2,1/2)} = \begin{cases} 1, & \text{if} \sigma_1 = \sigma_2, \\ \dfrac{1}{2}\left[\int\limits_{-\infty}^{x_4} f_1(x)dx + \int\limits_{x_4}^{x_5} f_2(x)dx + \int\limits_{x_5}^{+\infty} f_1(x)dx\right] & \text{if } \sigma_1 < \sigma_2, \end{cases}$$

where $x_4 = \mu - \sigma_1\sigma_2\sqrt{E}$ and $x_5 = \mu + \sigma_1\sigma_2\sqrt{E}$ with $E = \frac{2}{\sigma_2^2-\sigma_1^2}\ln\left(\frac{\sigma_2}{\sigma_1}\right) \geq 0$.

ii. If $f_1(x)$ and $f_2(x)$ are two n-dimension normal pdfs $(N(\mu_i, \Sigma_i), n\geq 2, i = 1, 2)$ then

$$Pe_{1,2}^{(1/2,1/2)} = \frac{1}{2}\left[\int_{R_1} f_2(x)dx + \int_{R_2} f_1(x)dx\right],$$

where

$$R_1^n = \{x : d(x)\leq 0\}, R_2^n = \{x : d(x) > 0\},$$
$$d(x) = \left[\mu_1^T(\Sigma_1)^{-1} - \mu_2^T(\Sigma_2)^{-1}\right]x - \frac{1}{2}x^T\left[(\Sigma_1)^{-1} - (\Sigma_2)^{-1}\right]x - m,$$
$$m = \frac{1}{2}\left[\ln\frac{|\Sigma_1|}{|\Sigma_2|} + \mu_1^T(\Sigma_1)^{-1}\mu_1 - \mu_2^T(\Sigma_2)^{-1}\mu_2\right].$$

In case of $n = 2$, $d(x)$ can be straight lines or parabola or ellipses or hyperbola.

3.3. Maximum function in the classification problem

To classify a new element by the principle (1) and to determine Bayes error by the formula (3), we must find $g_{\max}(x)$. Some authors, such as Pham–Gia et al. [15, 17] and Tai [21, 22], have surveyed relationships between $g_{\max}(x)$ with some related quantities of classification problem. The specific expression for $g_{\max}(x)$ in some special case has been found [18]. However, the general expression for all of cases is a complex problem that has not been still found yet.

Given k pdfs $f_i(x)$ and q_i, $i = 1, 2, \ldots, k$ with $q_1 + q_2 + \ldots + q_k = 1$ and let $g_i(x) = q_if_i(x)$, $g_{\max}(x) = \max\{g_i(x)\}$. Now, we take interest in determining $g_{\max}(x)$.

(a) For one dimension

In this case, we can find $g_{\max}(x)$ by the following algorithm:

Algorithm 2. Find the $g_{\max}(x)$ function
Input: $g_i(x) = q_if_i(x)$, where $f_i(x)$ and q_i are the probability density function and the prior probability of w_i, $i = 1, 2, \ldots, k$, respectively.
Output: The $g_{\max}(x)$ function.

Find all roots of the equations $g_i(x) - g_j(x) = 0, i = \overline{1, k-1}, j = \overline{i+1, k}$.
Let B be the set of all roots.
For $x_{lm} \in B$ (the roof of equation $g_l(x) - g_m(x) = 0$) do
    For $p \in \{1,2,\ldots,k\}\setminus\{l,m\}$ do
        If $g_l(x_{lm}) < g_p(x_{lm})$ then $B = B\setminus\{x_{lm}\}$
        End
    End
End
Arrange the elements of B in order from smallest to largest:
$$B = \{x_1, x_2, \ldots, x_h\}, x_1 < x_2 < \ldots < x_h$$

(*Determine the function* $g_{\max}(x)$ *in interval* $(-\infty, x_1]$)

```
For i = 1 to k do
      If g_i(x_1 − ε_1) = max{g_1(x_1 − ε_1), g_2(x_1 − ε_1), …, g_k(x_1 − ε_1)} then
            g_max(x) = g_i(x), for all x ∈ (−∞,x_1]
      End
End
```

(*Determine the function* $g_{\max}(x)$ *in interval* $(x_j, x_{j+1}]$, $j = \overline{1, h-1}$)

```
For i =1 to k do
   For j =1 to h-1 do
      If g_i(x_j + ε_2) = max{g_1(x_j + ε_2), g_2(x_j + ε_2), …, g_k(x_k + ε_2)} then
            g_max(x) = g_i(x), for all x ∈ (x_j, x_{j+1}]
      End
   End
End
```

(*Determine the function* $g_{\max}(x)$ *in interval* $(h, +\infty)$)

```
For i = 1 to k do
   If g_i(x_h + ε_3) = max{g_1(x_h + ε_3), g_2(x_h + ε_3), …, g_k(x_h + ε_3)} then
            g_max(x) = g_i(x), for all x ∈ (h, +∞)
   End
End
```

In the above algorithm, ε_1, ε_2, ε_3 are the positive constants such that:

$$x_1 + \varepsilon_1 < x_2,\ x_h - \varepsilon_3 > x_{h-1},\ x_i - \varepsilon_2 < x_{i-1} \text{ and } x_i + \varepsilon_2 < x_{i+1}.$$

From this algorithm, we have written a MATLAB code to find the $g_{\max}(x)$. When $g_{\max}(x)$ is determined, we will easily calculate Bayes error by using formula (3), as well as classify a new element by principle (1).

Example 2. Given seven populations having univariate normal pdfs $\{f_1, f_2, \ldots, f_7\}$ with specific parameters as follows (**Figure 1**):

$$\mu_1 = 0.3, \mu_2 = 4.0, \mu_3 = 9.1, \mu_4 = 1.9, \mu_5 = 5.3, \mu_6 = 8, \mu_7 = 4.8,$$
$$\sigma_1 = 1.0, \sigma_2 = 1.3, \sigma_3 = 1.4, \sigma_4 = 1.6, \sigma_5 = 2, \sigma_6 = 1.9, \sigma_7 = 2.3.$$

Using codes written with $q_i = 1/7, g_i(x) = q_i f_i(x),\ i = 1, 2, .., 7,$ we have the results:

$$g_{\max}(x) = \begin{cases} g_1 & \text{if} \quad -1.28 < x \le 0.99, \\ g_2 & \text{if} \quad 2.58 < x \le 4.89, \\ g_3 & \text{if} \quad 8.30 < x \le 12.52, \\ g_4 & \text{if} \quad \{-7.86 < x \le -1.28\} \cup \{0.99 < x \le 2.58\}, \\ g_5 & \text{if} \quad 4.89 < x \le 6.65, \\ g_6 & \text{if} \quad \{6.65 < x \le 8.30\} \cup \{12.52 < x \le 23.33\}, \\ g_7 & \text{if} \quad \{x \le -7.86\} \cup \{x > 23.33\}. \end{cases}$$

(b) For multidimension

In multidimension cases, it should be very complicated to obtain closed expression for $g_{\max}(x)$. The difficulty comes from the various forms of the intersection space curves between the pdfs surfaces. This problem has been interested by many authors in Refs. [17, 18, 21–25]. Pham–Gia

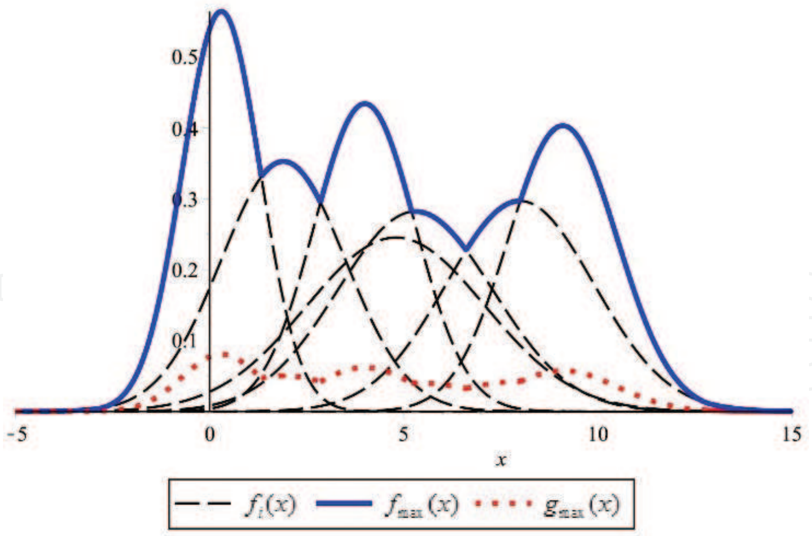

Figure 1. The graph of seven one-dimension normal pdfs, $f_{\max}(x)$ and $g_{\max}(x)$.

et al. [18] have attempted to find the function $g_{\max}(x)$; however, it has been only established for some cases of bivariate normal distribution.

Example 3. Given the four bivariate normal pdfs $N(\mu_i, \Sigma_i)$ with the following specific parameters [16]:

$$\mu_1 = \begin{bmatrix} 40 \\ 20 \end{bmatrix}, \mu_2 = \begin{bmatrix} 48 \\ 24 \end{bmatrix}, \mu_3 = \begin{bmatrix} 43 \\ 32 \end{bmatrix}, \mu_4 = \begin{bmatrix} 38 \\ 28 \end{bmatrix},$$

$$\Sigma_1 = \begin{pmatrix} 35 & 18 \\ 18 & 20 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 28 & -20 \\ -20 & 25 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 15 & 25 \\ 25 & 65 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 5 & -10 \\ -10 & 7 \end{pmatrix}$$

With q_1 = 0.25, q_2 = 0.2, q_3 = 0.4, and q_4 = 0.15, we have the graphs of $g_i(x) = q_i f_i(x)$ and their intersection curves as shown in **Figure 2**.

Here, we do not find the expression of $g_{\max}(x)$. We compute Bayes error instead by taking integration of $g_{\max}(x)$ by quasi-Monte Carlo method [17]. An algorithm for doing calculations has been constructed, and a corresponding MATLAB procedure is used in Section 4.

3.4. Estimate the probability density function

There are many parameter and nonparameter methods to estimate pdfs. In the examples and applications of Section 4, we use the kernel function method, the popular one in practice nowadays. It has the following formula:

$$\widehat{f}(x) = \frac{1}{Nh_1h_2\ldots h_n} \sum_{i=1}^{N} \prod_{j=1}^{n} f_j\left(\frac{x_j - x_{ij}}{h_j}\right), \quad (19)$$

where x_j, j = 1,2,…,n are variables, x_j, i = 1,2,…,N are the ith data of the jth variable, h_j is the bandwidth parameter for the jth variable, $f_j(.)$ is the kernel function of the jth variable which is usually normal, Epanechnikov, biweight, and triweight. According to this method, the choice

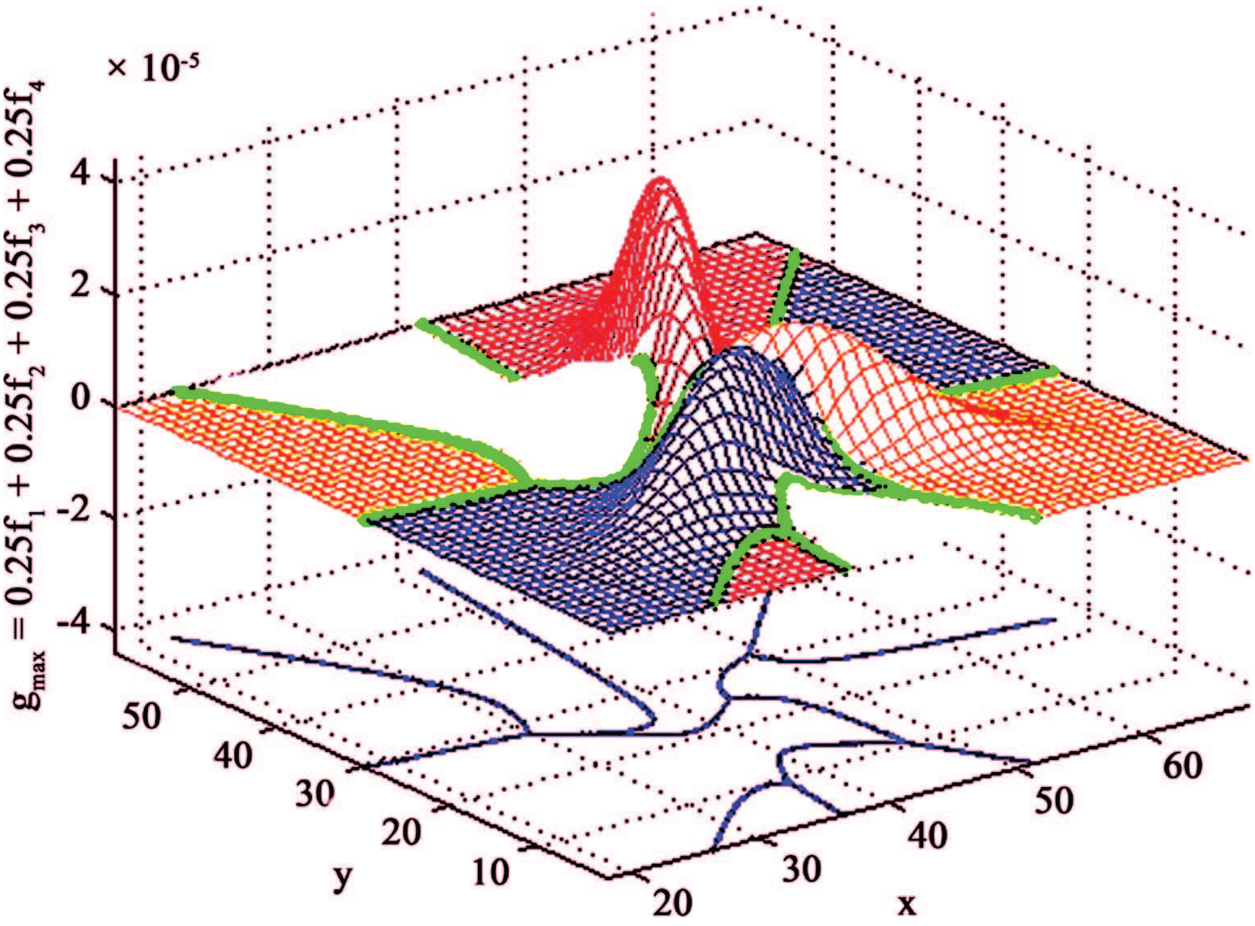

Figure 2. The graph of three bivariate normal pdfs and their $g_{\max}(x)$.

of smoothing parameter and the type of kernel function play an important role and affect the result. Although Silverman [20], Martinez and Martinez [10], and some other authors [7, 13, 27] had discussions about this problem, the optimal choice still has not been found yet. In this chapter, the smoothing parameter is from the idea of Scott [19] and the kernel function is the Gaussian one. We have also written the code by MATLAB software to estimate the pdfs in n-dimensions space using this method.

We have written the complete code for the proposed algorithm by MATLAB software. It is applied effectively for the examples of Section 4.

4. Some applications

In this section, we will consider three applications in three domains: biology, medicine, and economics to illustrate for present theories and to test established algorithms. They also show that the proposed algorithm presents more advantages than the existing ones.

Application 1. We consider classification for well-known Iris flower data, which have been presented in many documents like in Ref. [17]. These data are often used to compare the new method and existing ones in classifying. The three varieties of Iris, namely, Setosa (Se), Versicolor (Ve), and Virginica (Vi), have data in four attributes: $X1$ = sepal length, $X2$ = sepal width, $X3$ = petal length, and $X4$ = petal width.

In this application, the cases of one, two, three and four variables are respectively considered to classify for three groups (Se), (Ve), and (Vi) by Bayesian method with different prior probabilities. The purpose of this performance is to compare the results of BayesC with BayesU, BayesR, and BayesL. Because the numbers of the three groups are equal, and the results of BayesU, BayesR, and BayesL are the same. The correct probability of methods is summarized in **Table 3**.

Table 3 shows that in almost all cases, the results of proposed algorithm are better than those using other algorithms, and in the case using three variables X1, X2, and X3, it gives the best results.

Application 2. This application considers thyroid gland disease (TGD). Thyroid gland is an important and the largest gland in our body. It is responsible for the metabolism and work process of all cells. Some of the common diseases of gland thyroid are hypothyroidism, hyperthyroidism, thyroid nodules, and thyroid cancer. They are dangerous diseases. Recently, the rate of thyroid gland disease has been increasing in some poor countries. Data includes 3772 person with 3541 for ill group (I) and 231 ones for nonill group (NI). Detail for this data is given in http://www.cs.sfu.ca/wangk/ucidata/dataset/thyroid–disease, in which the surveyed variables are Age ($X1$), Query on thyroxin ($X2$), Anti-thyroid medication ($X3$), Sick ($X4$), Pregnant ($X5$), Thyroid surgery ($X6$), Thyroid Stimulating Hormone ($X7$),

Variables	B BayesU = BayesL = BayesR	BayesC
$X1$	0.667	0.679
$X2$	0.668	0.579
$X3$	0.903	0.916
$X4$	0.815	0.827
$X1$, $X2$	0.715	0.807
$X1$, $X3$	0.893	0.895
$X1$, $X4$	0.807	0.850
$X2$, $X3$	0.891	0.898
$X2$, $X4$	0.809	0.815
$X3$, $X4$	0.843	0.866
$X1$, $X2$, $X3$	0.892	0.919
$X1$, $X2$, $X4$	0.764	0.810
$X1$, $X3$, $X4$	0.762	0.814
$X2$, $X3$, $X4$	0.736	0.822
$X1$, $X2$, $X3$, $X4$	0.725	0.745

Table 3. The correct probability (%) in classifying Iris flower.

Triiodothyronine ($X8$), Total thyroxin ($X9$), T4U measured ($X10$), and Referral source ($X11$). In this application, this chapter will use random 70% of the data size (2479 elements belong to group I and 162 elements belong to group NI) as the training set to determine significant variables, to estimate pdfs, and to find suitable model. About 30% of the remaining data will be used as test set (1062 elements belong to group I and 69 elements belong to group NI). The result of Bayesian method is also compared to others.

To assess the effect of independent variables in TGD, we build the logistic regression model log ($p/1-p$) with variables Xi, $i = 1, 2, \ldots, 11$ (p is the probability of TGD). The analytical results are summarized in **Table 4**.

In **Table 4**, the three variables $X1$, $X8$, and $X11$ in bold face have statistical significance in classifying the two groups (I) and (NI) at 5% level, so we use them to classify TGD.

Applying the PPC algorithm for cases of one variable, two variables, and three variables with all prior probabilities, we obtain the results given in **Table 5**.

Table 5 shows that the correct probability is high, in which BayesC always gives the best result in all three cases of variables. BayesC gives the almost exact result with three variables. We also compare BayesC with existing methods (Fisher, SWM, and logistic) for all the above three cases. All cases show that BayesC is more advantageous than others in reducing Bayes error.

Variable	Sig.	Variable	Sig.
$X1$	**0.000**	$X7$	0.304
$X2$	0.279	**$X8$**	**0.000**
$X3$	0.998	$X9$	0.995
$X4$	0.057	$X10$	0.999
$X5$	0.997	**$X11$**	**0.000**
$X6$	0.997	Const	0.992

Table 4. Value Sigs of logistic regression model.

Cases	Variables	BayesU	BayesR	BayesL	BayesC
One variable	$X1$	91.13	97.47	97.46	97.97
	$X8$	90.72	98.51	98.50	98.65
	$X11$	90.53	97.48	97.47	98.19
Two variables	$X1$, $X8$	98.73	98.77	98.77	99.78
	$X1$, $X11$	98.11	98.65	97.65	99.44
	$X8$, $X11$	98.71	98.77	98.77	99.82
Three variables	$X1$, $X8$, $X11$	98.35	98.89	98.89	99.96

Table 5. The correct probability (%) in classifying TGD by Bayesian method from training set.

Using the best results for each case of methods from **Table 6**, classifying for test set (1131 elements), we have the results given in **Table 7**.

From **Table 7**, we see that with the test set, BayesC also gives the best result.

Application 3. This application considers the problem of repaying bank debt (RBD) by customers. In bank credit operations, determining the repayment ability of customers is really important. If the lending is too easy, the bank may have bad debt problems. In contrast, the bank will miss a good business. Therefore, in the current years, the classification of credit application on assessing the ability to repay bank debt has been specially studied and has been a difficult problem in Vietnam. In this section, we appraise this ability of companies in Can Tho city (CTC), Vietnam by using the proposed approach. We collect a data on 214 enterprises operating in key sectors as agriculture, industry, and commerce, including 143 cases of good debt (G) and 71 cases of bad debt (B). Data are provided by responsible organizations of CTC. Each company is evaluated by 13 independent variables in the expert opinion. The specific variables are given in **Table 8**.

Because of sensitive problem, author has to conceal real data and use training data set. The steps to perform in this application are similar as in Application 2. Training set has 100 elements belonging to group G and 50 elements belonging to group B, and the test set has 43 elements belonging to group G and 21 elements belonging to group B. With training set, the logistic regression model shows only three variables $X1$, $X4$, and $X7$ have statistical significance at 5% level, so we use these three variables to perform BayesU, BayesR, BayesL, and BayesC. Their results are given in **Table 9**.

From **Table 9**, we see that BayesC gives the highest probability in all the cases. We also use logistic method, Fisher, and SVM with training set to find the best results. We have the correct probability given in **Table 10**.

Methods	One variable	Two variables	Three variables
Logistic	93.90	93.90	93.90
Fisher	72.30	73.60	71.70
SVM	93.87	93.87	93.87
BayesC	98.65	99.82	99.96

Table 6. The correct probability (%) for optimal models of methods in classifying TGD.

Methods	Correct numbers	False numbers	Correct probability
Logistic	835	296	73.8
Fisher	835	296	73.8
SVM	1062	69	90.9
BayesC	1062	69	93.9

Table 7. Compare the correct probability (%) in classifying TGD from test set.

Using the best model for each case of methods from **Table 10** to classify the test set (67 elements), we obtain the results given in **Table 11**.

Once again from **Table 11**, we see that with test data, BayesC also gives the best result.

*X*i	Independent variables	Detail
*X*1	Financial leverage	Total debt/total equity
*X*2	Reinvestment	Total debt/total equity
*X*3	Roe	Net profit/equity
*X*4	Interest	(Net income + depreciation)/total assets
*X*5	Floating capital	(Current assets − current liabilities)/total assets
*X*6	Liquidity	(Cash + Short-term investments)/current liabilities
*X*7	Profits	Net profit/total assets
*X*8	Ability	Net sales/Total assets
*X*9	Size	Logarithm of total assets
*X*10	Experience	Years in business activity
*X*11	Agriculture	Agricultural and forestry sector
*X*12	Industry	Industry and construction
*X*13	Commerce	Trade and services

Table 8. The surveyed independent variables.

Cases variables		BayesU	BayesR	BayesL	BayesC
One variable	*X*1	86.21	86.14	84.13	87.13
	*X*4	81.12	82.91	86.16	88.19
	*X*7	83.21	84.63	83.14	84.52
Two variables	*X*1, *X*4	87.25	88.72	87.19	89.06
	*X*1, *X*7	88.16	88.34	83.26	89.56
	*X*4, *X*7	89.25	89.04	89.02	91.34
Three variables	*X*1, *X*5, *X*7	91.15	91.53	90.17	93.18

Table 9. The correct probability (%) in classifying RBD by Bayesian method from training set.

Methods	One variable	Two variables	Three variables
Logistic	84.04	88.29	88.69
Fisher	84.73	80.73	79.32
SWM	82.34	82.03	83.07
BayesC	88.19	91.34	93.18

Table 10. The correct probability (%) for optimal models of methods in classifying RBD.

Methods	Correct numbers	False numbers	Correct probability
Logistic	53	11	82.81
Fisher	52	12	81.25
SVM	53	11	82.81
BayesC	57	7	89.06

Table 11. Compare the correct probability (%) in classifying RBD from test set.

5. Conclusion

This chapter presents the classification algorithm by Bayesian method in both theory and application aspect. We establish the relations of Bayes error with other measures and consider the problem to compute it in real application for one and multidimensions. An algorithm to determine the prior probabilities which may decrease Bayes error is proposed. The researched problems are applied in three different domains: biology, medicine, and economics. They show that the proposed approach has more advantages than existing ones. In addition, a complete procedure on MATLAB software is completed and is effectively used in some real applications. These examples show that our works present potential applications for research on real problems.

Author details

Tai Vovan

Address all correspondence to: vvtai@ctu.edu.vn

College of Natural Sciences, Can Tho University, Can Tho City, Vietnam

References

[1] Altman DG. Statistics in medical journals: Development in 1980s. Statistical Medicine. 1991;**10**:546-551. DOI: 10.1002/sim.4780101206

[2] Bora DJ, Gupta AK. Impact of exponent parameter value for the partition matrix on the performance of fuzzy C means Algorithm. International Journal of Scientific Research in Computer Science Applications and Management Studies. 2014;**3**:1-6. DOI: arXiv:1406.4007

[3] Cristiani S, Shawe TJ. An introduction to support vector machines and other kernel-based learning method. 2nd ed. London: Cambridge University; 2000. p. 204. DOI: 10.1108/k.2001.30.1.103.6

[4] Cannon RL, Dave JV, Bezdek JC. Efficient implementation of the fuzzy c-means clustering algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1986;**2**:248-255. DOI: 10.1109/TPAMI.1986.4767778

[5] Fadili MJ, et al. On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series. Medical Image Analysis. 2001;**5**(1):55-67. DOI: 10.1016/S1361-8415(00)00035-9

[6] Fisher RA. The statistical utilization of multiple measurements. Annals of Eugenic. 1936;**7**: 376-386. DOI: 10.1111/j.1469-1809.1938.tb02189

[7] Ghosh AK. Classification using kernel density estimates. Technometrics. 2006;**48**:120-132. DOI: 10.1198/004017005000000391

[8] Jan YK, Cheng CW, Shih YH. Application of logistic regression analysis of home mortgage loan prepayment and default. ICIC Express Letters. 2010;**2**:325-331. DOI: 325-331. 10.12783/ijss.2015.03.014

[9] Hall LO, et al. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. IEEE Transactions on Neural Networks. 1992;**3**(5):672-682. DOI: 10.1109/72.159057

[10] Martinez WL, Martinez AR. Computational statistics handbook with Matlab. 1st ed. Boca Raton: CRC Press; 2007. DOI: 1198/tech.2002.s89

[11] Matusita K. On the notion of affinity of several distributions and some of its applications. Annals of the institute of Statistical Mathematics. 1967;**19**:181-192. DOI: 10.1007/BF02481018

[12] Marta E. Application of Fisher's method to materials that only release water at high temperatures. Portugaliae Etecfochlmlca Acta. 2001;**15**:301-311. DOI: 10.1016/S0167-7152(02)00310-3

[13] McLachlan GJ, Basford KE. Mixture Models: Inference and Applications to Clustering. 1st ed. New York: Marcel Dekker; 1988. DOI: 10.2307/2348072

[14] Miller G, Inkret WC, Little TT. Bayesian prior probability distributions for internal dosimetry. Radiation Protection Dosimetry. 2001;**94**:347-352. DOI: 10.1093/oxfordjournals.rpd.a006509

[15] Pham-Gia T, Turkkan T. Bounds for the Bayes error in classification: A Bayesian approach using discriminant analysis. Statistical Methods and Applications. 2006;**16**:7-26. DOI: 10.1007/s10260-006-0012-x

[16] Pham–Gia T, Turkkan N, Bekker A. Bayesian analysis in the L1–norm of the mixing proportion using discriminant analysis. Metrika. 2008;**64**:1-22. DOI: 10.1007/s00184-006-0027-1

[17] Pham–Gia T, Turkkan N, Tai VV. Statistical discrimination analysis using the maximum function. Communications in Statistics-Simulation and Computation. 2008;**37**:320-336. DOI: 10.1080/03610910701790475

[18] Pham–Gia T, Nhat ND, Phong, NV. Statistical classification using the maximum function. Open Journal of Statistics. 2015;**15**:665-679. DOI: 10.4236/ojs.2015.57068

[19] Scott DW. Multivariate density estimation: Theory, practice, and visualization. 1st ed. New York: Wiley; 1992. DOI: 10.1002/9780470316849

[20] Silverman BW. Density Estimation for Statistics and Data Analysis. 1st ed. Boca Raton: CRC Press; 1986. DOI: 10.1007/978-1-4899-3324-9

[21] Tai VV, Pham–Gia T. Clustering probability distributions. Journal of Applied Statistics. 2010;**37**:1891-1910. DOI: 10.1080/02664760903186049

[22] Tai VV. L^1–distance and classification problem by Bayesian method. Journal of Applied Statistics. 2017; **44**(3):385-401. DOI: 10.1080/02664763.2016.1174194

[23] Tai VV, Thao NT, Ha CN. The prior probability in classifying two populations by Bayesian method. Applied Mathematics Engineering and Reliability. 2016;**6**:35-40. DOI: 10.1201/b21348-7

[24] Thao NT, Tai VV. Fuzzy clustering of probability density functions. Journal of Applied Statistics. 2017;**44**(4):583-601. DOI: 0.1080/02664763.2016.117750

[25] Thao NT, Tai VV. A new approach for determining the prior probabilities in the classification problem by Bayesian method. Advances in data analysis and classification. Forthcoming. DOI: 10.1007/s11634-016-0253

[26] Toussaint GT. Some inequalities between distance: Measures for feature evaluation. IEEE Transactions on Computers. 1972;**21**:405-410. DOI: 10.1109/TC1972.5008990

[27] Webb AR. Statistical Pattern Recognition. 1st ed. New York: Wiley; 2003. DOI: 10.1109/34.824819

[28] Yu J, Cheng Q, Huang H. Analysis of the weighting exponent in the FCM. IEEE Transactions on Systems, Man, and Cybernetics, Part B. 2004;**34**(1):634-639