Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 2120–2132
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Toward a tight upper bound for the error probability of the binary Gaussian classification problem

Moataz M.H. El Ayadi*, Mohamed S. Kamel, Fakhri Karray

Pattern Analysis and Machine Intelligence Lab, Electrical and Computer Engineering, University of Waterloo, Ont., Canada N2L 3G1

Received 22 March 2007; received in revised form 23 October 2007; accepted 26 October 2007

Abstract

It is well known that the error probability, of the binary Gaussian classification problem with different class covariance matrices, cannot be generally evaluated exactly because of the lack of closed-form expression. This fact pointed out the need to find a tight upper bound for the error probability. This issue has been for more than 50 years ago and is still of interest. All derived upper-bounds are not free of flaws. They might be loose, computationally inefficient particularly in highly dimensional situations, or excessively time consuming if high degree of accuracy is desired. In this paper, a new technique is developed to estimate a tight upper bound for the error probability of the well-known binary Gaussian classification problem with different covariance matrices. The basic idea of the proposed technique is to replace the optimal Bayes decision boundary with suboptimal boundaries which provide an easy-to-calculate upper bound for the error probability. In particular, three types of decision boundaries are investigated: planes, elliptic cylinders, and cones. The new decision boundaries are selected in such a way as to provide the tightest possible upper bound. The proposed technique is found to provide an upper bound, tighter than many of the often used bounds such as the Chernoff bound and the Bayesian-distance bound. In addition, the computation time of the proposed bound is much less than that required by the Monte-Carlo simulation technique. When applied to real world classification problems, obtained from the UCI repository [H. Chernoff, A measure for asymptotic efficiency of a hypothesis based on a sum of observations, Ann. Math. Statist. 23 (1952) 493–507.], the proposed bound was found to provide a tight bound for the analytical error probability of the quadratic discriminant analysis (QDA) classifier and a good approximation to its empirical error probability.

Keywords: Binary classification; Bayesian decision rule; Decision boundary; Error probability; Monte-Carlo simulations; Multivariate normal distribution; Quadratic surfaces

1. Introduction

It is well known that there is no closed-form expression for the error probability of the binary Gaussian classification problem when the covariance matrices of the two class are different. Though the attempts of finding a tight upper bound for this error probability dates back to more than 50 years ago [1], it has been found that this problem is still of interest in many applications in different research areas. For example, in pattern recognition applications, the quadratic discriminant analysis (QDA) classifier has been found to provide powerful classification performances in some pattern recognition applications [2] despite its simplicity in comparison with other classification techniques. Since the QDA is based on the assumption that the class conditional densities are Gaussian, there have been some attempts to give a tight upper bound for the error probability of the QDA classifier [3]. In addition, estimates of this error probability has been used recently as a feature selection criterion [4]. In digital communication, there have been many attempts to estimate an upper bound for the error probability of different communication systems such as non-coherent coded modulation [5] and code division multiple access [6,7] when the communication channel suffers from an additive white Gaussian noise (AWGN), slow fading, or rapid fading. It can be shown that estimating the detection error probability in all these cases is equivalent to the considered Gaussian classification problem. In addition, in some radio astronomy and sonar applications, both the signal and the noise are best modelled as Gaussian

* Corresponding author. Tel.: +1 519 888 4576.
E-mail address: moataz@pami.uwaterloo.ca (M.H. El Ayadi).

0031-3203/$30.00 Crown Copyright © 2007 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2007.10.028

random processes since signals are perturbed by propagation through turbulent media [8]. Hence, the problem of evaluating the detection-error probability is also equivalent the considered Gaussian classification problem.

Because of the popularity of this problem, there have been many attempts to upper bound this error probability. The oldest and most known bounds for this error probability are the Chernoff bound [1] and the Bhattacharyya bound [9]. Though these bounds are easy to calculate, they are found to be significantly loose in many problems. Tighter error bounds have also been proposed: the equivocation bound [10]; the Bayesian distance bound [11]; the sinusoidal bound [3]; and the exponential bound [12]. However, again, there is no closed-form expressions for these bounds and numerical integrations may be necessary. Thus, evaluation of these bounds may well be inefficient for problems with high dimensionality because they require a great deal of computations.

A third approach to approximate the error probability is through the generation of a relatively large number of random instances that follow the distribution of each class. The generated samples are then classified according to the Bayesian decision rule. The error probability is simply estimated as the ratio of the number of misclassified samples to the total number of samples. This technique in the literature is known as the Monte-Carlo simulations [13]. Amazingly, the complexity of this technique does not grow with increasing the dimensionality of the data. However, the accuracy of the obtained estimate is exponentially proportional to the number of the generated points. In particular, in order to increase the accuracy of the obtained estimate to one digit of precision, the number of generated samples should be increased by two orders of magnitude. Therefore, the estimation of the error probability using this technique is time consuming due to the large number of samples required for a sufficient degree of accuracy. In order to overcome with this deficiency, some improvement has been proposed to the *conventional* Monte-Carlo simulation technique such as importance sampling [13]. Though these methods have been applied successfully in some problems, it is difficult to apply them in some complex problems such as Viterbi decoding [14].

In this paper, a new method is proposed for the estimation of a tight upper bound for the error probability of the binary Gaussian classification problem. This method has the advantages of providing good approximation to the error probability with a relatively small computation time. The basic idea is to replace the Bayesian boundary with sub-optimal decision boundaries. In particular, three types of decision boundaries are considered: planes, elliptic cylinder, and cones. The main motivation behind this replacement is the relative easiness of calculating the error probability when these surfaces are used as classification boundaries. At the same time, these boundaries are suboptimal in the sense that classification using them must result in inferior classification performance than that provided by the optimal Bayesian decision rule. Hence, these boundaries provide an upper bound for the true error probability. However, in order not to obtain a loose upper bound, their parameters should be optimized in such a way to obtain the least possible upper bound.

The rest of this paper is organized as follows. In Section 2, the binary Gaussian classification problem is briefly reviewed with the introduction of known error bounds. The possible shapes of the optimal Bayesian decision boundary are discussed in Section 3. The proposed error upper bound is described in Section 4. Comparative performance evaluation of proposed bounding technique with other selected techniques is provided in Section 5. Finally, concluding remarks are mentioned in Section 6.

2. The binary Gaussian classification problem: A brief review

The binary Gaussian classification problem is generally formulated as follows. Given a vector, $\mathbf{x} \in \mathbb{R}^d$, which belongs to one of either two possible classes: $\mathscr{C}_1$ or $\mathscr{C}_2$. It is required to determine which one of the following two hypotheses is more likely to occur:

$\mathscr{H}_1$: $\mathbf{x}$ comes from the first class, $\mathscr{C}_1$.

$\mathscr{H}_2$: $\mathbf{x}$ comes from the second class, $\mathscr{C}_2$.

It is usually assumed that data vectors belonging to each class $\mathscr{C}_i$, $i = 1, 2$ follow a certain distribution $\mathrm{p}(\mathbf{x}|\mathscr{C}_i)$, $i = 1, 2$. Usually, these distributions are incompletely known and the unknown parameters are estimated based on a set of training data. In this paper, it will be assumed that the amount of the training data is enough to estimate the unknown parameters with a sufficient degree of accuracy. Moreover, it will be assumed that the distributions take the form a multivariate normal distribution, i.e.,

$$\mathrm{p}(\mathbf{x}|\mathscr{C}_i) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right), \tag{1}$$

where $\boldsymbol{\mu}_i$ and $\mathbf{\Sigma}_i$ are the mean vector and the covariance matrix of $\mathscr{C}_i$. According to the Bayesian decision theory, the optimal decision rule is given as

$$P(\mathscr{C}_1)\mathrm{p}(\mathbf{x}|\mathscr{C}_1) \underset{\mathscr{H}_1}{\overset{\mathscr{H}_2}{\lessgtr}} P(\mathscr{C}_2)\mathrm{p}(\mathbf{x}|\mathscr{C}_2). \tag{2}$$

Substituting Eq. (1) into Eq. (2) and performing some manipulations, the optimal Bayes decision rule for the Gaussian case is given by

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} - 2\mathbf{b}^{\mathrm{T}}\mathbf{x} + c \underset{\mathscr{H}_1}{\overset{\mathscr{H}_2}{\gtrless}} 0, \tag{3}$$

where

$$\mathbf{A} = \mathbf{\Sigma}_1^{-1} - \mathbf{\Sigma}_2^{-1},$$

$$\mathbf{b} = \mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \mathbf{\Sigma}_2^{-1}\boldsymbol{\mu}_2,$$

$$c = \boldsymbol{\mu}_1^{\mathrm{T}}\mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^{\mathrm{T}}\mathbf{\Sigma}_2^{-1}\boldsymbol{\mu}_2 + \log\frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_2|} - 2\log\frac{P(\mathscr{C}_1)}{P(\mathscr{C}_2)},$$

A classification error occurs if a data vector $\mathbf{x}$ belongs to one class but falls in the decision region of the other class.