Ann. Inst. Statist. Math.
34 (1982), Part A, 161-165

AN UPPER BOUND ON THE PROBABILITY OF MISCLASSIFICATION IN TERMS OF MATUSITA'S MEASURE OF AFFINITY

BINAY K. BHATTACHARYA AND GODFRIED T. TOUSSAINT

(Received June 1, 1981; revised Nov. 9, 1981)

Summary

A distribution-free upper bound is derived on the Bayes probability of misclassification in terms of Matusita's measure of affinity among several distributions for the M-hypothesis discrimination problem. It is shown that the bound is as sharp as possible.

1. Introduction

Let us consider the discrimination problem of classifying an observation X as coming from one of M possible classes $\theta \in \{1, 2, \cdots, M\}$. Let $\eta_i = \Pr\{\theta = i\}$, $i=1, 2, \cdots, M$ denote the prior probabilities of the classes. Let $f_1(x), \cdots, f_M(x)$ denote the conditional probability density functions given the true class or hypothesis. We assume that the $f_i(x)$ and η_i, $i=1, \cdots, M$ are completely known. In such a situation it is well known that the decision rule which minimizes the probability of error is the Bayes decision rule which chooses the hypothesis with the largest posterior probability. We denote the resulting probability of error by

$$P_e = 1 - \int \max_i \{\eta_i f_i(x)\} dx . \tag{1}$$

Matusita [4] has defined the *affinity* of $f_1(x), \cdots, f_M(x)$ as

$$\rho_M = \int [f_1(x) f_2(x) \cdots f_M(x)]^{1/M} dx . \tag{2}$$

For the two-hypothesis problem the *affinity*, also known as the *Bhattacharyya coefficient* [1], is given by

Key words: Probability of misclassification, Matusita's measure of affinity, Bhattacharyya coefficient, information measures, discrimination rules, pattern classification, decision theory.

CR categories: 5.25, 5.30, 5.5, 3.36, 3.63.

(3) $$\rho_{12}=\int \sqrt{f_1(x)f_2(x)}\,dx\ .$$

In [6] Matusita applied ρ_M to discriminant analysis techniques. An axiomatic foundation for ρ_M in the multivariate discrete case was given by Kaufman and Mathai [2], and some properties of ρ_M were derived by Toussaint [7].

Matusita [5], [6] also derived a lower bound on P_e in terms of ρ_M given by

(4) $$P_e \geqq \frac{M-1}{M^{M-1}}\eta_1\eta_2\cdots\eta_M(\rho_M)^M\ .$$

Although he gave no upper bound on P_e in terms of ρ_M, he offered the following upper bound in terms of the pairwise affinities ρ_{ij}:

(5) $$P_e \leqq \sum_{i<j}^{M}\sqrt{\eta_i\eta_j}\,\rho_{ij}\ .$$

A corresponding lower bound on P_e in terms of ρ_{ij} was later exhibited by Kirmani [3] who showed that

(6) $$P_e \geqq \left(\frac{M-1}{M}\right)-\frac{1}{M}\sum_{i<j}^{M}\sqrt{(\eta_i+\eta_j)^2-4\eta_i\eta_j\rho_{ij}^2}\ .$$

Kirmani [3] suggested that (6) was sharper than (4) by proving that this was so when $M=2$.

The lower bound problem was finally settled by Toussaint [9] who showed that

(7) $$\rho_M \leqq K(M,\theta)(1-P_e)^{1/M}(P_e)^{(M-1)/M}$$

where

$$K(M,\theta)=(\eta_1\eta_2\cdots\eta_M)^{-1/M}(M-1)^{(1-M)/M}\ ,$$

and that this bound is as sharp as possible. If, for example, (7) is loosened by using the relation

$$(P_e)^{(M-1)/M} \leqq (P_e)^{1/M}$$

then one obtains

(8) $$P_e \geqq \frac{1}{2}-\frac{1}{2}\sqrt{1-4(M-1)^{M-1}\eta_1\eta_2\cdots\eta_M(\rho_M)^M}\ ,$$

where if M is set equal to two we obtain Kirmani's result (6).

An upper bound on P_e in terms of ρ_M was derived by Toussaint [8] and is given by:

$$P_e \leqq \left(\frac{M-2}{2}\right) + \frac{M}{2}(\eta_1\eta_2\cdots\eta_M)^{1/M}\rho_M \, . \tag{9}$$

For $M=2$ (9) reduces to Matusita's result (5). For the case of equally likely hypotheses (9) reduces to

$$P_e \leqq \left(\frac{M-2}{2}\right) + \frac{1}{2}\rho_M \, . \tag{10}$$

From (10) we can see that this bound is very loose and in fact, for $M \geqq 4$ it becomes useless.

In this paper we settle the upper bound problem by deriving a distribution-free upper bound on P_e in terms of ρ_M and proving that the bound is as sharp as possible.

2. Upper bound on P_e in terms of ρ_M

THEOREM.

$$P_e \leqq \left(\frac{M-2}{M-1}\right) + \frac{1}{(M-1)}(\eta_1\eta_2\cdots\eta_M)^{1/M}\rho_M \, . \tag{11}$$

PROOF. Let $g_i(x)=f_i(x)\eta_i$. Then it is obvious that:

$$(M-1)\max_i \{g_i(x)\} \geqq \sum_{i=1}^{M} g_i(x) - \min_i \{g_i(x)\} \, . \tag{12}$$

Rearranging (12) yields

$$\frac{1}{M-1}\sum_{i=1}^{M} g_i(x) - \max_i \{g_i(x)\} \leqq \frac{1}{M-1}\min_i \{g_i(x)\} \, . \tag{13}$$

Now since the $f_i(x)$ and η_i are non-negative it is always true that for any x

$$\min_i \{g_i(x)\} \leqq \left[\prod_{i=1}^{M} g_i(x)\right]^{1/M} . \tag{14}$$

Substituting (14) into (13) we obtain

$$\frac{1}{M-1}\sum_{i=1}^{M} g_i(x) - \max_i \{g_i(x)\} \leqq \frac{1}{M-1}\left[\prod_{i=1}^{M} g_i(x)\right]^{1/M} . \tag{15}$$

The left term of (15) can be broken up into

$$\sum_{i=1}^{M} g_i(x) - \left(\frac{M-2}{M-1}\right)\sum_{i=1}^{M} g_i(x)$$

which, after integrating both sides of the inequality and using the fact that $\int f_i(x)dx=1$ yields

$$1-\int \max_i \{f_i(x)\eta_i\}dx \leqq \left(\frac{M-2}{M-1}\right)+\frac{1}{M-1}\int \left[\prod_{i=1}^{M} f_i(x)\eta_i\right]^{1/M} dx \ . \tag{16}$$

Applying the definitions of P_e and ρ_M in (1) and (2) to (16) completes the proof.

We now show that inequality (11) is as sharp as possible by exhibiting distributions for which the equality in (11) is achieved for any value of M. We need only consider the one-dimensional case. Define

$$f_i(x)=\begin{cases} 0 & \text{for } i-1\leqq x\leqq i-1+\delta \\ 0 & \text{for } x\leqq 0 \\ 0 & \text{for } x\geqq M \\ 1/(M-\delta) & \text{elsewhere} \end{cases} \tag{17}$$

for $i=1, 2, \cdots, M$ and where δ is a positive constant such that $0<\delta<1$, and let $\eta_1=\eta_2=\cdots=\eta_M=1/M$.

Substituting the equal priors and the densities defined in (17) into equation (2) and integrating yields

$$\rho_M=\frac{M(1-\delta)}{M-\delta} \ . \tag{18}$$

Alternately we can write

$$\delta=(M-M\rho_M)/(M-\rho_M) \ . \tag{19}$$

Substituting equal priors and (17) into (1) and integrating we obtain

$$P_e=\frac{M-\delta-1}{M-\delta} \ . \tag{20}$$

Substituting (19) for δ in (20) and performing some algebra yields

$$P_e=\left(\frac{M-2}{M-1}\right)+\frac{1}{M(M-1)}\cdot\rho_M$$

thus establishing that the equality in (11) can be achieved.

McGILL UNIVERSITY

REFERENCES

[1] Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection, *IEEE Trans. Commun. Tech.*, COM-15, 52-60.

[2] Kaufman, H. and Mathai, A. M. (1973). An axiomatic foundation for a multivariate measure of affinity among a number of distributions, *J. Multivariate Anal.*, **3**, 236-242.

[3] Kirmani, S. N. U. A. (1976). A lower bound on Bayes risk in classification problems, *Ann. Inst. Statist. Math.*, **28**, A, 385-387.

[4] Matusita, K. (1967). On the notion of affinity of several distributions and some of its applications, *Ann. Inst. Statist. Math.*, **19**, 181-192.

[5] Matusita, K. (1971). Some properties of affinity and applications, *Ann. Inst. Statist. Math.*, **23**, 137-155.

[6] Matusita, K. (1973). Discrimination and the affinity of distributions, *Discriminant Analysis and Applications* (ed. T. Cacoullos), Academic Press, New York, 213-223.

[7] Toussaint, G. T. (1974). Some properties of Matusita's measure of affinity of several distributions, *Ann. Inst. Statist. Math.*, **26**, 389-394.

[8] Toussaint, G. T. (1977). An upper bound on the probability of misclassification in terms of the affinity, *Proc. IEEE*, **65**, 275-276.

[9] Toussaint, G. T. (1978). Probability of error, expected divergence, and the affinity of several distributions, *IEEE Trans. Systems, Man, and Cybernetics*, SMC-**8**, 482-485.