Correspondence

A Problem of Dimensionality: A Simple Example

G. V. TRUNK

Abstract—In pattern recognition problems it has been noted that beyond a certain point the inclusion of additional parameters (that have been estimated) leads to higher probabilities of error. A simple problem has been formulated where the probability of error approaches zero as the dimensionality increases and all the parameters are known; on the other hand, the probability of error approaches one-half as the dimensionality increases and parameters are estimated.

Index Terms—Dimensionality.

I. Introduction

In hypothesis-testing problems involving unknown parameters, a common way of proceeding is the generalized-likelihood approach [1] where the parameters are estimated, substituted into the likelihood ratio, and are then treated as if the estimates were the true parameters. Duda and Hart [2] comment that beyond a certain point the inclusion of additional features (parameters) leads to higher probabilities of error. They note that the basic source of the problem is the finite number of design samples and that the numerical examples are rather complicated. They also discuss several topics relating dimensionality and sample size, such as the capacity of a separating plane and the problem-average error rate [3]. The purpose of this correspondence is to give a simple example where the error rate initially decreases but then eventually increases as the dimensionality increases and parameters are estimated. In Section II the problem is formulated and solved assuming all parameters are known. In Section III the problem is solved estimating the unknown parameters, and it is shown that the probability of error approaches one-half.

II. Mean Values Known

Consider the following binary hypothesis testing problem: *a priori* probabilities are $p(\omega_l) = p(\omega_2) = \frac{1}{2}$, and the conditional probabilities are Gaussian

$$p(X|\omega_1) \sim G(\mu_1, I)$$

$$p(X|\omega_2) \sim G(\mu_2, I)$$

where $\mu_1 = \mu$, $\mu_2 = -\mu$, μ is an n-vector mean value whose ith component is $(1/i)^{1/2}$, and I is an identity covariance matrix. The Bayes detector which minimizes the probability of error is the correlator

$$\text{decide } \omega_1 \text{ if } \tilde{X}\mu > 0$$

where $\tilde{X}$ is the transpose of X. The probability of error is given by

$$P_e = \int_{r/2}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \tag{1}$$

where

$$r^2 = \|\mu_1 - \mu_2\|^2 = 4 \sum_{i=1}^{n} (1/i). \tag{2}$$

Since $1/i$ is a divergent series, the probability of error can be made to approach zero by increasing the dimensionality of the problem. That is, $P_e \to 0$ as $n \to \infty$.

III. Mean Value Unknown

Let us now assume that μ is unknown and that m independent labeled samples $X_1, \cdots, X_m$ are available. The best estimate of μ is the sample mean

$$\hat{\mu} = \frac{1}{m} \sum_{i=1}^{m} X_i \tag{3}$$

where X_i has been replaced by $-X_i$ if X_i came from ω_2. $\hat{\mu}$ is an unbiased estimate whose covariance matrix equals I/m. Since the *a priori* probabilities are equal, the probability of error is given by

$$P_e = P_r(\tilde{X}\hat{\mu} > 0|\omega_2). \tag{4}$$

Unfortunately, the density of $z = \tilde{X}\hat{\mu}$ is difficult to calculate for finite dimensionality n. However, it is rather straightforward to show that the mean and variance of z are given by

$$E(z) = \sum_{i=1}^{n} (1/i) \tag{5}$$

$$\text{var}(z) = \left(1 + \frac{1}{m}\right) \sum_{i=1}^{n} (1/i) + n/m. \tag{6}$$

The Lindeberg conditions [4] are satisfied, and consequently, the normalized variate $[z - E(z)]/\text{var}(z)^{1/2}$ has a Gaussian distribution as n approaches infinity. The probability of error is then given by

$$P_e = \int_{\gamma_n}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \tag{7}$$

where

$$\gamma_n = E(z)/[\text{var}(z)]^{1/2}.$$

It will now be shown that

$$\lim_{n \to \infty} \gamma_n = 0 \tag{8}$$

and thus the probability of error approaches one-half as the

Manuscript received May 10, 1978; revised July 24, 1978.

The author is with the Naval Research Laboratory, Washington, DC 20375.

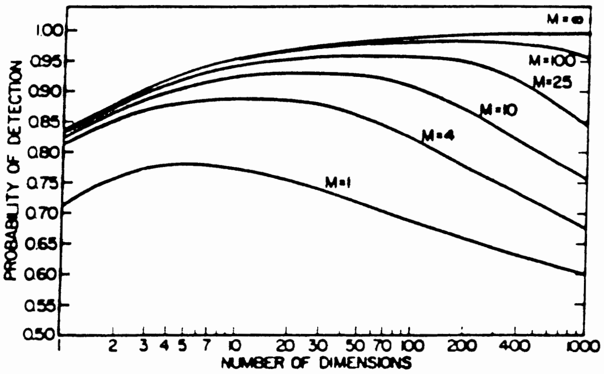

Fig. 1. Probability of detection versus dimensionality for a various number of design samples M.

dimensionality of the problem becomes very high. By ignoring the sum term in the variance, one obtains

$$\left(\frac{m}{n}\right)^{1/2} \sum_{i=1}^{n} (1/i) > \gamma_n. \tag{9}$$

Since

$$\lim_{n\to\infty} \left(\sum_{i=1}^{n} (1/i) - \log n\right) = \text{Euler's constant}, \tag{10}$$

dividing by $n^{1/2}$ yields the fact that

$$\lim_{n\to\infty} \left(\frac{m}{n}\right)^{1/2} \sum_{i=1}^{n} (1/i) = 0. \tag{11}$$

Thus, since (11) bounds γ_n, (8) has been proven and the probability of error approaches one-half for any finite set of m labeled samples.

To investigate the convergence rate of the probability of error [given by (4)] to one-half, the problem was simulated. For each dimension between 1 and 1000 and for each value of m (m = 1, 4, 10, 25, or 100), 500 repetitions were run. The results of the simulation are plotted in Fig. 1. For the problem selected, the dimensionality interval of near optimal performance is rather large and the convergence to one-half is rather slow. The interval of near optimal performance would be smaller and the convergence to one-half more rapid if the ith component of μ is $(1/i)^K$ where $K > \frac{1}{2}$. Of course, P_e would then not approach zero as $n \to \infty$ since $\|\mu_1 - \mu_2\|^2$ would not approach infinity.

Summary

A simple problem has been formulated where the probability of error approaches zero as the dimensionality increases and all the parameters are known. On the other hand, the probability of error approaches one-half as the dimensionality increases and parameters are estimated from a finite number of design samples.

References

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York, NY: Wiley, 1968.

[2] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York, NY: Wiley, 1973.

[3] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 55-63, Jan. 1968.

[4] M. G. Kendall, *The Advanced Theory of Statistics, Vol. 1*. London: Charles Griffin, 1943.

Comment on "A Declustering Criterion for Feature Extraction in Pattern Recognition"

G. EDEN

Abstract—The purpose of this correspondence is to point out that the scaling procedure (as well as any nonsingular transformation) described in Section VI of the above paper[1] is unnecessary.

Index Terms—Clustering, feature extraction, Rayleigh quotients.

Consider the form of the generalized Rayleigh quotient $J = (d^tBd)/(d^tAd)$ where A and B are kernel matrices based on means and covariances of random vectors in a space X.[1] It follows that B is at least positive semidefinite and A is positive definite. Let T: $X \to X$ be a nonsingular transformation. Let w_i optimize J; i.e., $A^{-1}Bw_i = \lambda_i w_i$ and z_i optimize J on the transformed space; i.e., $(TAT^t)^{-1}(TBT^t)z_i = \rho_i z_i$. Since T is nonsingular it follows that $(T^t)^{-1}A^{-1}BT^t z_i = \rho_i z_i$ which implies that $\{\rho_i\} = \{\lambda_i\}$ and $z_i = (T^t)^{-1}w_i$. The projections of the transformed vectors on the eigenvectors z_i are $(Tx)^t z_i = x^t w_i$, which are the same projections of the original vectors on the original eigenvectors z_i. Therefore, the declustering criterion[1] and all other criteria based on a Rayleigh quotient are insensitive to such transformations and scaling can be omitted.

Manuscript received August 30, 1978; revised November 16, 1978.
The author is with the Medische Informatica, Vrije Universiteit, Amsterdam, The Netherlands.

[1] J. Fehlauer and B. A. Eisenstein, *IEEE Trans. Comput.*, vol. C-27, pp. 261-266, Mar. 1978.

The Use of a Syntactic Shape Analyzer for Contour Matching

THEODOSIOS PAVLIDIS

Abstract—Description of contours in terms of complex arcs allows the use of simple algorithms for matching. An example of such matchings for island contours is included.

Index Terms—Contour matching, shape analysis.

A problem common to scene analysis and pattern recognition is the *matching* of objects with prototypes (e.g., [1], [2]). In its most general formulations this problem is equivalent to graph isomorphism or homeomorphism [3], [4] which in turn show exponential growth of the time of computation as a function of the size of the problem [5]. The purpose of this correspondence is to demonstrate that the proper prepro-

Manuscript received July 17, 1978; revised October 4, 1978. This work was supported by the National Science Foundation under Grants ENG76-16808 and ENG76-22311.
The author is with the Department of Electrical Engineering and Computer Science, Princeton University, Princeton, NJ 08544.

0162-8828/79/0700-0307$00.75 © 1979 IEEE