The Mean Recognition Performance for Independent Distributions

ROBERT P. W. DUIN

Abstract—**Conditions are given for the mean recognition performance over a class of independent distributions to approach unity when the dimensionality is raised to infinity.**

I. Introduction

The mean recognition performance has been studied by several authors. Hughes [1] investigated a measurement space quantized into n measurement cells. For this space a two-class pattern recognition problem is defined by the sample size of each class, m_1,m_2, and by the set of probabilities $P=\{p_{ij}\}$ for the cells, $j=1,\cdots,n$, $i=1,2$. The mean recognition performance $\bar{P}_{cr}(n,m_1,m_2)$ is now defined as the probability of correct recognition averaged over all sets P and over all samples of size m_1,m_2. Hughes showed that when P has a uniform *a priori* probability distribution, $\bar{P}_{cr}$ has a maximum in n with m_1,m_2 fixed. If the number of cells is increased beyond that optimum, $\bar{P}_{cr}$ decreases. The optimal value of n, called the *optimal measurement complexity*, is then a function of m_1 and m_2. A small sample size m_1,m_2 results in a small optimal measurement complexity.

This result has been further investigated by Chandrasekaran and others [2]–[6] who called it the "peaking phenomenon." A somewhat different model was used in which $\bar{P}_{cr}$ was now studied as a function of the dimensionality of the measurement space instead of the number of cells. For the case of independent binary features, Chandrasekaran [4] showed that $\bar{P}_{cr}$ has no peaking and approaches unity if the dimensionality is raised to infinity. Again a uniform distribution was used for the parameters. In this paper we examine more general conditions for $\bar{P}_{cr}$ to approach one monotonically as the dimension is raised to infinity. The proof is partly based on the paper by Chandrasekaran and Jain [6].

Manuscript received September 13, 1976; revised August 14, 1977.

The author is with the Department of Applied Physics, Delft University of Technology, Delft, The Netherlands.

II. Formulation of the Conditions

Let c_1,c_2 be two classes with *a priori* probabilities p_1,p_2, with $p_1+p_2=1$. To discriminate between the classes, n measurements $x^1,\cdots,x^n$ are taken. Let $f(x^j|\theta_i^j)$ be the density of x^j where θ_i^j is the parameter value associated with the density function of x^j given class c_i. If we assume independent measurements, the Bayes decision function is given by

$$\begin{aligned}&\text{choose } c_1, \quad \text{if } p_1 \prod_{j=1}^{n} f(x^j|\theta_1^j) \geq p_2 \prod_{j=1}^{n} f(x^j|\theta_2^j),\\ &\text{choose } c_2, \quad \text{otherwise.}\end{aligned} \tag{1}$$

We will assume that $(\theta_1^1,\theta_2^1),\cdots,(\theta_1^n,\theta_2^n)$ are independent identically distributed (i.i.d.) with a probability density $G(\theta_1,\theta_2)$. To estimate θ_1^j,θ_2^j, an independent sample of the jth measurements is generated, and estimates $\hat{\theta}_1^j,\hat{\theta}_2^j$ are formed. If

$$r^j=\log\left(f(x^j|\hat{\theta}_1^j)/f(x^j|\hat{\theta}_2^j)\right), \qquad 1\leq j\leq n,$$

then, given class c_i, $r^1,\cdots,r^n$ is an i.i.d. sequence. Letting $d=\log(p_2/p_1)$ and using the decision rule

$$\begin{aligned}&\text{choose } c_1, \quad \text{if } \sum_1^n r^j \geq d,\\ &\text{choose } c_2, \quad \text{otherwise,}\end{aligned} \tag{2}$$

we see that

$$\bar{P}_{cr}=p_1P\left(\sum_1^n r^j \geq d|\text{class } 1\right)+p_2P\left(\sum_1^n r^j < d|\text{class } 2\right). \tag{3}$$

To show that $\bar{P}_{cr}$ approaches 1 monotonically, it suffices to show that

$$E_{\theta_1,\theta_2}E_xE_{x|c_1}, \qquad r^1>0, \tag{4}$$

$$E_{\theta_1,\theta_2}E_xE_{x|c_2}, \qquad r^1<0, \tag{5}$$

where E_{θ_1,θ_2} is the expectation over θ_1^1,θ_2^1, E_x is the expectation over the sample used to estimate θ_1^1,θ_2^1, and $E_{x|c_i}$ is the expectation over x^1 given class c_i.

We will prove that the conditions (4) and (5) are satisfied if $G(\theta_1,\theta_2)$ satisfies

$$G(\theta_1,\theta_2)=G(\theta_2,\theta_1), \qquad \text{for all } \theta_1,\theta_2 \tag{6}$$

and

$$\iint_{\theta_1\neq\theta_2} G(\theta_1,\theta_2)\, d\theta_1\, d\theta_2>0, \tag{7}$$

and if $R(x,\theta_1,\theta_2)=E_xr^1$ satisfies

$$R(x,\theta_1,\theta_2)>0, \qquad \text{if } f(x|\theta_1)>f(x|\theta_2), \tag{8}$$

$$R(x,\theta_1,\theta_2)=0, \qquad \text{if } f(x|\theta_1)=f(x|\theta_2), \tag{9}$$

$$R(x,\theta_1,\theta_2)<0, \qquad \text{if } f(x|\theta_1)<f(x|\theta_2). \tag{10}$$

Note that

$$R(x,\theta_1,\theta_2)=-R(x,\theta_2,\theta_1) \tag{11}$$

because of the definition of r. For the proof we write (4) as

$$\int_x\int_{\theta_1}\int_{\theta_2} R(x,\theta_1,\theta_2)f(x|\theta_1)G(\theta_1,\theta_2)\, d\theta_1\, d\theta_2\, dx>0.$$

Let $S=f(x|\theta_1)-f(x|\theta_2)$. The integrals over θ_1 and θ_2 can be split into a sum of three terms, one term with $S>0$, one with $S<0$, and one with $S=0$. The last of these terms is zero because

0018-9448/78/0500-0394$00.75

of (9). If we interchange θ_1 and θ_2 in the integral over $S<0$, we get, using (6) and (11),

$$\int_x \iint_{S>0} R(x,\theta_1,\theta_2)\{f(x|\theta_1)-f(x|\theta_2)\}G(\theta_1,\theta_2)\,d\theta_1\,d\theta_2\,dx>0.$$

All factors are positive because of (7) and (8), which means that this condition and thereby (4), are satisfied. In the same way (5) can be proved.

III. Discussion

The conditions (4) and (5) guarantee for independent distributions that the mean probability of correct recognition approaches unity monotonically. These conditions differ slightly from the ones given in [6], mainly because Chandrasekaran and Jain do not demand monotonic behavior. However, this is necessary in order to avoid peaking in $\bar{P}_{cr}$. We proved that our conditions are fulfilled if (6)–(10) are satisfied. The condition (6) is probably the most demanding one. Condition (7) simply requires that the classes differ in their statistical behavior and is trivial when $G(\theta_1,\theta_2)$ contains no impulses or other types of singularities. The conditions (6) and (7) include the common assumption that θ_1 and θ_2 are uniformly distributed over the same interval. The conditions (8)–(10) require that the expected value (over all sample sets) of the estimated discriminant function r have the same sign as the optimal one.

If all conditions are satisfied, $\bar{P}_{cr}$ approaches unity monotonically. Note, however, that this does not imply that in a particular problem peaking can be avoided.

References

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

[2] K. Abend, T. J. Harly, Jr., B. Chandrasekaran, and G. F. Hughes, "Comments 'On the mean accuracy of statistical pattern recognizers'," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 420–423, May 1969.

[3] L. Kanal and B. Chandrasekaran, "On dimensionality and sample size in statistical pattern classification," *Pattern Recognition*, vol. 3, pp. 225–234, 1971.

[4] B. Chandrasekaran, "Independence of measurements and the mean recognition accuracy," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 452–456, July 1971.

[5] B. Chandrasekaran and A. K. Jain, "Quantization complexity and independent measurements," *IEEE Trans. Comput.*, vol. C-23, pp. 102–106, Jan. 1974.

[6] B. Chandrasekaran and A. K. Jain, "Independence, measurement complexity, and classification performance," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-5, pp. 240–244, Mar. 1975.