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 $P = \{p_i\}$ for the cells, $i = 1, \ldots, n$. The mean recognition performance $\overline{P}_e(n, m_1, m_2)$ 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 priori probability distribution, $\overline{P}_e$ has a maximum in $n$ with $m_1, m_2$ fixed. If the number of cells is increased beyond that optimum, $\overline{P}_e$ 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 $\overline{P}_e$ 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 $\overline{P}_e$ 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 $\overline{P}_e$ to approach one monotonically as the dimension is raised to infinity. The proof is partly based on the paper by Chandrasekaran and Jain [6].

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 $x_1, \ldots, x_n$ are taken. Let $f(x_i | \theta)$ be the density of $x_i$ where $\theta$ is the parameter value associated with the density function of $x_i$ given class $c_i$. If we assume independent measurements, the Bayes decision function is given by

$$\begin{align*}
\text{choose } c_1, & \quad \text{if } p_1 \prod_{j=1}^{n} f(x_i | \theta_1^j) > p_2 \prod_{j=1}^{n} f(x_i | \theta_2^j), \quad (1) \\
\text{choose } c_2, & \quad \text{otherwise.}
\end{align*}$$

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

$$r^j = \log \left( \frac{f(x_i | \hat{\theta}_1)}{f(x_i | \hat{\theta}_2)} \right), \quad 1 \leq j \leq n,$$

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

$$\begin{align*}
\text{choose } c_1, & \quad \text{if } \sum_{1}^{n} r^j > d, \quad (2) \\
\text{choose } c_2, & \quad \text{otherwise,}
\end{align*}$$

we see that

$$\overline{P}_e = p_1 P \left( \sum_{1}^{n} r^j > d | \text{class 1} \right) + p_2 P \left( \sum_{1}^{n} r^j < d | \text{class 2} \right). \quad (3)$$

To show that $\overline{P}_e$ approaches 1 monotonically, it suffices to show that

$$E_{\theta_1, \theta_2} E_{x, x_1} \quad r^1 > 0, \quad (4)$$

and

$$E_{\theta_1, \theta_2} E_{x, x_1} r^1 < 0, \quad (5)$$

where $E_{\theta_1, \theta_2}$ is the expectation over $\theta_1, \theta_2$, $E_x$ is the expectation over the sample used to estimate $\theta_1, \theta_2$, and $E_{x_1}$ is the expectation over $x_i$ given class $c_i$.

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

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

and

$$\int \int G(\theta_1, \theta_2) \, d\theta_1 \, d\theta_2 > 0, \quad (7)$$

and if $R(x, \theta_1, \theta_2) = E_{x^1}$ satisfies

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

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

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

Note that

$$R(x, \theta_1, \theta_2) = -R(x, \theta_2, \theta_1) \quad (11)$$

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

$$\int_{x^1} \int \left( R(x, \theta_1, \theta_2) f(x | \theta_1) G(\theta_1, \theta_2) \right) d\theta_1 \, d\theta_2 \, dx > 0.$$

Let $S = f(x | \theta_1) - f(x | \theta_2)$. The integrals over $\theta_1$ and $\theta_2$ 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
of (9). If we interchange \( \theta_1 \) and \( \theta_2 \) in the integral over \( S < 0 \), we get, using (6) and (11),

\[
\int \int_{S > 0} R(x, \theta_1, \theta_2) \{ f(x|\theta_1) - f(x|\theta_2) \} G(\theta_1, \theta_2) \, d\theta_1 \, d\theta_2 \, dx > 0.
\]

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 \( P_{av} \). 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 \( \theta_1 \) and \( \theta_2 \) 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, \( P_{av} \) approaches unity monotonically. Note, however, that this does not imply that in a particular problem peaking can be avoided.

REFERENCES


