Sankhyā : The Indian Journal of Statistics
1989, Volume 51, Series B, Pt. 3, pp. 287-294.

PATTERN RECOGNITION WHEN FEATURE VARIABLES ARE SUBJECT TO ERROR

By SOMESH DAS GUPTA
Indian Statistical Institute

SUMMARY. In the problem of pattern recognition the measurements of the feature variables are often subjected to noise. Various effects of this noise-vector are considered in this paper especially with respect to the changes in the Bayes risk. The results of Chaudhuri (1982), and Chaudhuri, Murthy and Dutta Majumder (1983) have been rigorized and extended following a precise formulation.

1. Introduction

Let U denote an experimental unit randomly selected from a population π, which is known to be one of two distinct populations π_1 and π_2. The basic problem of classifying U into π_1 or π_2 is generally based on some vector $\boldsymbol{X}$ of p feature measurements on U. However, in some cases $\boldsymbol{X}$ cannot be measured directly, and instead $\boldsymbol{Y} = \boldsymbol{X}+\boldsymbol{e}$ is available, where $\boldsymbol{e}$ denotes the noise-vector affecting the $\boldsymbol{X}$-value. In this paper we shall study the various effects of the noise-component.

Let f_i be the density of X with respect to the Lebesgue measure, when π_i obtains. A classification rule φ based on the available data d is given by $\varphi(d) = (\varphi_1(d), \varphi_2(d))$, where $\varphi_i(d)$ is the probability of deciding $\pi = \pi_i$, given d. If X-measurements were available, then one may consider the Bayes rule $\varphi^*(\boldsymbol{X})$ for the prior probability vector (p_1, p_2), where p_i is the prior probability that π_i obtains. The Bayes rule φ^*, under zero-one loss function, is given by

$$\varphi_1^*(\boldsymbol{x}) = \begin{cases} 1, & \text{if } p_1 f_1(\boldsymbol{x}) > p_2 f_2(\boldsymbol{x}) \\ 0, & \text{if } p_1 f_1(\boldsymbol{x}) < p_2 f_2(\boldsymbol{x}). \end{cases} \qquad \ldots \quad (1.1)$$

The risk of φ^* can be found to be equal to

$$R_{\boldsymbol{X}} \equiv \left(\frac{1}{2}\right)\left[1- \int_{R^p} |p_1 f_1(\boldsymbol{x})-p_2 f_2(\boldsymbol{x})|\, d\boldsymbol{x}\right]. \qquad \ldots \quad (1.2)$$

However, if only $\boldsymbol{Y}$-values are available then one would consider the Bayes rule $\psi^*(\boldsymbol{Y})$, provided the density g_i of $\boldsymbol{Y}$ under π_i is known. In that case the Bayes risk of ψ^* would be equal to

$$R_{\boldsymbol{Y}} \equiv \left(\frac{1}{2}\right)\left[1- \int_{R^p} |p_1 g_1(\boldsymbol{y})-p_2 g_2(\boldsymbol{y})|\, d\boldsymbol{y}\right]. \qquad \ldots \quad (1.3)$$

AMS (1980) *subject classification :* 62H30.
Key words and phrases : Bayes rule, features subjected to noise, effects of noise distribution.

In case g_i's are not known but f_i's are known, one may use the rule $\varphi^*(Y)$ (i.e., φ^* given in (1.1) with x being replaced by y). The Bayes risk of such a rule is given by

$$R_{Y:X} = p_1 \int_{R^p} \varphi_2^*(y)\, g_1(y)\, dy + p_2 \int_{R^p} \varphi_1^*(y)\, g_2(y)\, dy. \quad \ldots \quad (1.4)$$

The following problems will be considered in this paper.

(a) How does R_Y compare with R_X?

(b) How does $R_{Y:X}$ compare with R_Y and R_X?

(c) How does the noise-distribution (i.e., the distribution of e) influence the value of $R_{Y:X}$?

(d) Given the noise distribution, how does $R_{Y:X}$ change with changes in f_i's ?

In the model $Y = X+e$, it is generally assumed that X and e are independently distributed, and the distribution of e is the same for both π_1 and π_2. Instead of the above model, we shall consider Y in a more general set-up, in which the conditional density $h(y/x)$ of Y given $X = x$ is the same for both π_1 and π_2.

The basic object of this paper is to understand the role of the error component.

2. Comparisons of R_Y, R_X and $R_{Y:X}$: general case

Theorem 2.1 : $R_X \leqslant R_Y \leqslant R_{Y:X}$.

Proof : First note that

$$\begin{aligned} &\int |p_1 g_1(y) - p_2 g_2(y)|\, dy \\ &= \int \left| \int (p_1 f_1(x) - p_2 f_2(x)) h(y/x) dx \right| dy \\ &\leqslant \int\int | p_1 f_1(x) - p_2 f_2(x) | h(y/x) dx \, dy \\ &= \int |p_1 f_1(x) - p_2 f_2(x)|\, dx. \end{aligned}$$

The above result along with (1.2) and (1.3) yields

$$R_X \leqslant R_Y.$$

Since the Bayes risk of $\varphi^*(Y)$ cannot be smaller than the Bayes risk of the Bayes rule $\psi^*(Y)$ based on Y, it follows that

$$R_Y \leqslant R_{Y:X}.$$

Note: It may be noted that φ^* and ψ^* are different in general. The equality $R_X = R_Y$ can hold if, and only if $p_1 f_1(x) - p_2 f_2(x)$ has the same sign (a.e.). On the other hand, $R_Y = R_{Y:X}$ can occur if, and only if, φ^* is a Bayes rule based on Y. This means that the symmetric difference of the sets $p_1 f_1(y) > p_2 f_2(y)$ and $p_1 g_1(y) > p_2 g_2(y)$, and the symmetric difference of the sets $p_1 f_1(y) < p_2 f_2(y)$ and $p_1 g_1(y) < p_2 g_2(y)$ have (Lebesgue) measure zero. For example, consider the univariate case with $p_1 = p_2 = 0.5$, $X \sim N(\mu_i, \sigma^2)$ under π_i, $Y = X + e$ where $e \sim N(0, \tau^2)$ independetnly of X; in this case $\varphi^* = \psi^*$ and $R_Y = R_{Y:X}$.

3. Study of $R_{Y:X}$: general case

It can be seen that

$$R_{Y:X} = p_2 + \int [\int \varphi_2^*(y) h(y/x) dy] (p_1 f_1(x) - p_2 f_2(x))\, dx. \quad \ldots \quad (3.1)$$

Theorem 3.1: *Suppose $p = 1$, and*

(a) $h(y/x) = \frac{1}{\tau} k\left(\frac{y-x}{\tau}\right)$, *for $\tau > 0$ and some density function k,*

(b) $p_1 f_1(x) - p_2 f_2(x)$ *changes sign only once, and takes nonzero value a.e. Then $R_{Y:X}$ increases with τ.*

Proof: Without loss of generality, suppose

$$p_1 f_1(x) - p_2 f_2(x) \geqslant 0 \text{ iff } x \geqslant c,$$

for some c. Then

$$\int \varphi_2^*(y) h(y/x)\, dy = \int_{-\infty}^{c} \frac{1}{\tau} k\left(\frac{y-x}{\tau}\right) dy$$

$$= K\left(\frac{c-x}{\tau}\right), \quad \ldots \quad (3.2)$$

where K is the c.d.f. corresponding to k. Now note that, for $\tau_1 < \tau_2$

$$\left[K\left(\frac{c-x}{\tau_2}\right) - K\left(\frac{c-x}{\tau_1}\right)\right](c-x) \leqslant 0. \quad \ldots \quad (3.3)$$

Hence, for all x

$$\left[K\left(\frac{c-x}{\tau_2}\right) - K\left(\frac{c-x}{\tau_1}\right)\right](p_1 f_1(x) - p_2 f_2(x)) \geqslant 0. \quad \ldots \quad (3.4)$$

The desired result now follows from (3.1), (3.2) and (3.4).

The above theorem can easily be generalized to the multivariate case as follows:

Theorem 3.2 : *Suppose*

(a) $p_1 f_1(x) - p_2 f_2(x) \geqslant 0$ *iff* $T(x) \geqslant c$ *for some real-valued statistic* T *and constant* c, $p_1 f_1(x) - p_2 f_2(x) \neq 0$ *a.e., and*

(b) *the conditional density of* $T(Y)$ *given* $X = x$ *is given by*

$$\frac{1}{\tau} k \left(\frac{T(y) - T(x)}{\tau} \right),$$

for some density function k *and* $\tau > 0$. *Then* $R_{Y:X}$ *increases with* τ.

Note 1 : Theorem 3.2 applies to the case when f_1 and f_2 are respective densities of $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$, and e is distributed as $N_p(0, \Gamma)$ independently of X. Then

$$T(x) = x' \Sigma^{-1}(\mu_1 - \mu_2),$$

and

$$\tau_2 = (\mu_1 - \mu_2)' \Sigma^{-1} \Gamma \Sigma^{-1} (\mu_1 - \mu_2).$$

Note 2 : Chaudhury, Murthy and Dutta Majumdar (1983) considered the problem of classifying U into one of $I (I \geqslant 2)$ populations, and proved that $R_X \leqslant R_Y$ *under the tacit assumption* $\varphi^* = \psi^*$. It may be noted that, except for some trivial or simple cases, φ^* and ψ^* are different. Nevertheless, the result $R_X \leqslant R_Y$ can be shown to be valid in general ; its proof is given below.

Let f_i and g_i be the densities of X and Y, respectively, under π_i ; let $h(y/x)$ be the conditional density of Y given $X = x$. Then the Bayes risk of φ^*, against the prior distribution $(p_1, \ldots, p_I)$ and under zero-one loss, can be shown to take the following form :

$$R_X = 1 - \int_{R^p} \max_{1 < i < I} (p_i f_i(x))\, dx.$$

On the other hand,

$$\begin{aligned} R_Y &= 1 - \int_{R^p} \max_{1 < i < I} (p_i\, g_i(y))\, dy \\ &= 1 - \int_{R^p} \max_{1 < i \leqslant I} \{p_i \textstyle\int f_i(x) h(y/x)\, dx\}\, dy \\ &\geqslant 1 - \int_{R^p} \left\{ \int_0 \max_{1 \leqslant i < I} (p_i f_i(x)) h(y/x)\, dx \right\} dy \\ &= R_X. \end{aligned}$$

When $I \geqslant 2$, the result $R_Y \leqslant R_{Y\,;\,X}$ follows from the argument used for the case $I = 2$.

4. Some examples

Example 4.1. Let f_1, f_2 and $h(./x)$ stand for the densities of $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$ and $N(x, \tau^2)$, respectively. For simplicity, assume $p_1 = p_2 = 0.5$. Then

$$R_X = \Phi(-\Delta/2),$$
$$R_Y = R_{Y:X} = \Phi(-\beta\Delta/2),$$

where $$\Delta = |\mu_1 - \mu_2|/\sigma, \beta = (1+\tau^2/\sigma^2)^{-1/2},$$

and Φ is the c.d.f. of $N(0, 1)$. For fixed β, it can be easily shown that $R_{Y:X} - R_X$ attains its maximum at

$$\Delta = \Delta_0(\beta) \equiv [8ln(1/\beta)/(1-\beta^2)]^{1/2}.$$

Moreover $\Delta_0(\beta)$ can be found to be a decreasing function of β. The following table gives the values of $R_{Y:X} - R_X$ at $\Delta = \Delta_0(\beta)$ for various values of β :

β	.01	.05	0.1	0.2	0.5	0.7	0.9
max $(R_{Y:X} - R_X)$	.48	.44	.40	.32	.16	.08	.02

The above result on maximisation of $R_{Y:X} - R_X$ was demonstrated by Chaudhuri (1982) only through numerical computation of the graph of $R_{Y:X} - R_X$ without explicit derivation of $\Delta_0(\beta)$. For a sensitivity analysis it seems more natural to consider

$$(R_{Y:X} - R_X)/R_X. \quad \ldots \quad (4.1)$$

It can be shown that the ratio (4.1) is approximately equal to the following for large Δ :

$$\frac{1}{\beta} \exp\{(1-\beta^2)\Delta^2/8\} - 1.$$

On the other hand, for τ^2/σ^2 close to 0 (or, β close to 1) the ratio (4.1) is approximately equal to

$$(1-\beta)(\Delta/2)\Phi'(-\Delta/2)/\Phi(-\Delta/2).$$

Example 4.2 : Let f_1 and f_2 represent the densities of $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Assume, for simplicity, $p_1 = p_2 = 0.5$. Suppose $\sigma_2 > \sigma_1$. Then

$$f_1(X) \geqslant f_2(x) \Longleftrightarrow c_1 \leqslant x \leqslant c_2,$$

for some c_1, c_2 depending on μ_i's and σ_i's. Then it follows from (3.1) that

$$R_{Y:X} = p_1 + \int\int \varphi_1^*(y)h(y/x)(p_2 f_2(x) - p_1 f_1(x)dy\, dx.$$

Hence

$$2R_{Y:X} = 1+\int_{-\infty}^{\infty}(f_2(x)-f_1(x))\left[\int_{c_1}^{c_2} h(y/x)\,dy\right]dx.$$

Suppose now

$$h(y/x) = \frac{1}{\tau}\,k\left(\frac{y-x}{\tau}\right),$$

where k is some density function, and $\tau > 0$. Chaudhuri, Murthy and Dutta Majumdar (1983) have claimed some monotonicity property of $R_{Y:X}$ with respect to τ. To see that the claim of Chaudhuri, Murthy and Dutta Majumdar (1983) is not generally true, let us specialize the distributions as follows :

Assume $\mu_1 = \mu_2 = 0$, $\sigma_2 > \sigma_1$, and $p_1 = p_2 = 0.5$. Then the Bayes rule φ^* is given by

$$\varphi_1^*(x) = \begin{cases} 1, & \text{if } x^2 < c \\ 0, & \text{otherwise} \end{cases}$$

where

$$c = 2.\,\frac{\sigma_2^2\,\sigma_1^2}{\sigma_2^2-\sigma_1^2}\,\ln(\sigma_2/\sigma_1).$$

Thus

$$\begin{aligned} R_{Y:X} &= P[Y^2 < c \mid Y \sim N(0,\sigma_2^2+\tau^2)]+P[Y^2 > c \mid Y \sim N(0,\sigma_1^2+\tau^2)] \\ &= G(c/(\sigma_2^2+\tau^2))+1-G(c/(\sigma_1^2+\tau^2)), \end{aligned}$$

where G is the c.d.f. of the chi-square distribution with 1 degree of freedom. The derivative of $R_{Y:X}$ with respect to τ^2 is positive if, and only if

$$\frac{\sigma_1^2\,\sigma_2^2}{(\sigma_1^2+\tau^2)\,(\sigma_2^2+\tau^2)} < \frac{3}{2}\,\ln\left(\frac{\sigma_2^2+\tau^2}{\sigma_1^2+\tau^2}\right).$$

The above inequality does not hold for all τ^2 and all $\sigma_1^2 < \sigma_2^2$.

It can be shown for this example that the Bayes rule φ^* with X replaced by Y is unique Bayes and admissible in the class of rules based on Y, when e is distributed as $N(0, \tau^2)$ independently of X. To see this, consider the following prior density of τ^2 when π_i obtains

$$c\,\sigma_i/(\sigma_i^2+\tau^2)^{3/2},$$

where c is a numerical constant. Then the unconditional density of Y under π_i becomes proportional to

$$(1/\sigma_i)\,\{1-\exp(-u_i^2)\}/u_i^2,$$

where

$$u_i^2 = (y-\mu_i)^2/2\,\sigma_i^2.$$

The desired result now follows by taking the prior probability p_i proportional to $(1/\sigma_i)$ and after noting that $\{1-\exp(-a)\}/a$ is decreasing function of $a > 0$.

5. The likelihood-ratio method

It may seem inappropriate to use Y in the Bayes rule determined from the known distributions of X. It may be proper to estimate the noise distribution based on observation on Y and incorporate this estimate in formulating a reasonable classification rule.

We demonstrate the above idea with the following simple case. Let f_i be the density of $N(\mu_i, \sigma^2)$, where μ_i's and σ^2 are known. Let $Y = X+e$, where e is distributed as $N(0, \tau^2)$ independently of X, with unknown $\tau > 0$. Next we derive the likelihood-ratio classification rule. First note that

$$\sqrt{2\pi} \sup_{\tau>0} g_i(y) = \sup_{\tau>0} \frac{1}{\sqrt{\tau^2+\sigma^2}} \exp\left\{-\frac{1}{2(\sigma^2+\tau^2)}(y-\mu_i)^2\right\}$$

$$= \begin{cases} |y-\mu_i|^{-1}\exp(-1/2), & \text{if } (y-\mu_i)^2/\sigma^2 > 1 \\ (1/\sigma)\exp\{-(y-\mu_i)^2/2\,\sigma^2\}, & \text{if } (y-\mu_i)^2/\sigma^2 \leqslant 1. \end{cases}$$

The general form of the likelihood-ratio rule can be expressed as follows. Classify into π_1 if, and only if

$$\sup_{\tau>0} g_1(y) > k \sup_{\tau>0} g_2(y),$$

where $0 < k < \infty$ is a constant. In particular, the above rule reduces to the following when $k = 1$. Classify into π_1 if, and only if

$$(y-\mu_1)^2 < (y-\mu_2)^2.$$

Now consider the above problem in the p-variate case. Then f_i stands for the density of $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, and $\boldsymbol{e}$ is assumed to be distributed as $N_p(\boldsymbol{0}, \tau^2 \boldsymbol{I}_p)$, independently of $\boldsymbol{X}$. The computation of the maximum likelihood estimate of τ^2 becomes very complicated. We suggest to use

$$\hat{\tau}_i^2 = \left[\frac{(\boldsymbol{Y}-\boldsymbol{\mu}_i)'(\boldsymbol{Y}-\boldsymbol{\mu}_i)-tr\,\boldsymbol{\Sigma}}{p}\right]_+$$

as an estimate of τ^2 when π_i obtains. Now one may consider the following classification rule : Classify π into π_1 if, and only if

$$(\boldsymbol{Y}-\boldsymbol{\mu}_1)'(\boldsymbol{\Sigma}+\hat{\tau}_1^2\boldsymbol{I}_p)^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}_1) < (\boldsymbol{Y}-\boldsymbol{\mu}_2)'(\boldsymbol{\Sigma}+\hat{\tau}_2^2\boldsymbol{I}_p)^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}_2).$$

It is believed that the above rule would be better than φ^* with $\boldsymbol{X}$ replaced by $\boldsymbol{Y}$.

REFERENCES

ANDERSON, T. W. (1958) : *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

CHAUDHURI, B. B. (1982) : Bayes' error and its sensitivity in statistical pattern recognition in noisy environment. *International Jour. Systems Sci.*, **13**, No. 5, 559-570.

CHOUDHURI, B. B., MURTHY, C. A. and DUTTA MAJUMDAR, D. (1983) : Bayes error probability for noisy and imprecise measurement in pattern recognition. *IEEE Trans. on Systems, Man and Cybernetics*, **13**, No. 1, 89-94.

Paper received : *July*, 1986.

Revised : *January*, 1989.