BULGARIAN ACADEMY OF SCIENCES
SPACE RESEARCH IN BULGARIA, 1
Sofia · 1978

Optimization of a Parallelepiped Discriminant Function at a Multidimensional Analysis

T. K. Yanev

When effecting a classificational analysis of stochastic subjects O_j, $j=1, \ldots N$, presented in the multidimensional space $S=\{x_i\}$, $i=1, \ldots, n$, the choice of a criterion is a basic stage for the quality of the deciding rule (discriminant function). Often the sum of the error probabilities from the 1st and 2nd genders serves as such a criterion. The minimum value of the sum is attained by the Bayes criterion of the minimum risk. As we know [1], this procedure, though well elaborated methodologically, entails considerable calculation difficulties, especially when the number of classes C_k, $k=1, \ldots, M$, formed by O_j, is considerable. In this case, even the linearized Bayes procedures, or the linear discriminant functions lose their effect in general, because a great number of linear discriminant functions are needed [1].

The simplest approach is, of course, the one in which the discriminant function is defined by simple in form volumes in the multidimensional space S. Such could be, by way of example, multidimensional parallelepipeds, spheres, ellipsoids, and the like. In this instance, the deciding rules are simplified considerably, and in the case of the multidimensional parallelepiped they are reduced to a system of simple inequations. On account of such a simplification of the deciding rule, the value of the risk function R increases.

This paper treats the problem of optimization of the parameters of the constant limits (discriminant functions) of the classes in space S, in order to obtain a minimum value of R (Naturally, this minimum value is higher than the value which could be obtained by the Bayes discriminant procedure).

As we know [3], in the general case the risk function has the following form:

$$(1) \qquad R=\int \sum_{k=1}^{M}\left[\sum_{m=1}^{M} p_m C_{mk} f(X/\mu_m)\delta(\gamma_k/X)\, dX\right],$$

where: $X(x_1, \dots, x_n)$ is the current vector-realization, p_m is the a priori probability of the m-class subject appearance, $f(X/\mu_m)$ is the density of the probability of the m-class subject appearance in the volume $dX = \prod_{i=1}^{n} dx_i$,

$$C = \begin{vmatrix} C_{11} & \dots & C_{1M} \\ \dots & \dots & \dots \\ C_{M1} & \dots & C_{MM} \end{vmatrix}$$

is a payment-off matrix with elements C_{km}, equal to the price of the error when relating the x to the m-class, at the k-class subject availability.

$\delta(\gamma_k/X)$ is a unit function, equal respectively to:

1, when X is in the region Γ_k of the k-class,

0, when X is in another region.

It is usually accepted that $C_{mm}=0$, and that condition will be taken into consideration in this paper later on.

Under the Bayes criterion of the minimum risk, regions Γ_k depend on the index m, i. e. $\Gamma_k = \Gamma_k(m)$, and are determined by the equation:

$$(2) \qquad p_k C_{mk} f(X/\mu_k) = p_m C_{km} f(X/\mu_m).$$

Equation (2) shows that the minimum risk is attained by the introduction of flexible limits, depending on the pair of indices (k, m) of the comparable classes. This condition makes the analysis complicated because, in the general case, regions (m) are strongly nonlinear and multidimensional.

At constant limits and normal distribution $f/x/\mu/$, which we accept further on, equation (1) has the form:

$$(3)\ R = \sum_{j=1}^{M} \sum_{k=1}^{M} C_{kj} p_k + \left[\frac{1}{(\sqrt{2\pi})^n} \sum_{j=1}^{M} \int_{a_{1j}}^{b_{1j}} \cdots \int_{a_{nj}}^{b_{nj}} \sum_{\substack{k=1 \\ k \neq j}}^{M} \left\{ p_k C_{kj} \frac{e^{-\frac{1}{2}(X-\mu_k)^T K_k^{-1} (X-\mu_k)}}{\sqrt{|K_k|}} \right. \right.$$

$$\left. \left. -(M-1) C_{jk} p_j \frac{e^{-\frac{1}{2}(X-\mu_j)^T K_j^{-1} (X-\mu_j)}}{\sqrt{|K_j|}} \right\} dX \right] = T_1 + T_2,$$

$$(3a)\ I_2 = \frac{1}{(\sqrt{2\pi})^n} \sum_{j=1}^{M} \sum_{\substack{k=1 \\ k \neq j}}^{M} C_{kj} p_k \int_{a_{1j}}^{b_{1j}} \cdots \int_{a_{nj}}^{b_{nj}} \frac{\exp\left[-\frac{1}{2}(X-\mu_k)^T K_K^{-1} (X-\mu_k)\right]}{\sqrt{|K_k|}} dX,$$

$$(3b)\ I_1 = \sum_{j=1}^{M} \sum_{\substack{k=1 \\ k \neq j}}^{M} C_{kj} p_k - (M-1) \sum_{j=1}^{M} \int_{a_{1j}}^{b_{1j}} \cdots \int_{a_{nj}}^{b_{nj}} \frac{\exp\left[-\frac{1}{2}(X-\mu_j)^T K_j^{-1} (X-\mu_j)\right]}{\sqrt{|K_j|}} dx,$$

K — covariational matrix.

In the general case, the limits a_{ij} and b_{ij} of the multidimensional parallelepiped can change independently on one another. At the R optimization by (3) this means the introduction of $2n$ parameters of the optimization. Further on we will accept the following limiting conditions:

$$
a_{ij} = \mu_{ij}(1-q), \qquad (4)
$$
$$
b_{ij} = \mu_{ij}(1+q),
$$

i. e. we accept that the limiting surface of the j-class is a parallelepiped which is centrally symmetric to the point described by the tip of the vector, μ_{ij}, and its magnitude changes with one and the same coefficient of proportion q along all the axes of the space S. Under this condition, it is seen in (3b) that I_1 is a monotonously decreasing function of q, when the values of μ_{ik}, K_k, M are fixed, because the normal distribution is positively defined in the multidimensional volume $(-\infty, \infty)^n$ and to each dq corresponds an increase of the integration region volume of the integral in (3b), and $dI_1 < 0$ is accordingly obtained.

Analogously, we obtain from (3a) that I_2 is a monotonously increasing function of q.

Also, the following limit relations follow from (3a) and (3b):

$$
I_{1_{\min}} = \lim_{q\to\infty} I_1 = 0; \quad I_{2_{\max}} = \lim_{q\to\infty} I_2 = \sum_{j=1}^{M} \sum_{\substack{k=1 \\ j\neq k}}^{M} C_{kj} p_k
$$

and therefore, because of monotonous change of I_1 and I_2, and since

$$
R_{\max} = \lim_{q\to\infty} R = \lim_{q\to\infty} I_2 = R(q=0), \qquad (5)
$$

then relations (5) show that in the region $(-\infty, \infty)^n$ the risk function R has at least one minimum. The values of q, which correspond to $R_{\min}$, can be determined by the equation

$$
\partial R/\partial q = 0. \qquad (6)
$$

In the relatively simple case, when signs x_i describing the subjects O_j of a given class are independent on one another, i. e. when the covariational matrix of the class is diagonal, the multiple integrals in (3) are given as a product of one-fold integrals, and a possibility is offered for condition (6) to be obtained by differentiation under the sign of a one-fold integral. Then we obtain for (6):

$$
\frac{\partial R}{\partial q} = \frac{1}{(\sqrt{2\pi})^n} \sum_{j=1}^{M} \sum_{\substack{k=1 \\ k\neq j}}^{M} C_{kj} p_k \sum_{l=1}^{n} \left\{ \frac{-[\mu_{lj}(1+q)-\mu_{lk}]}{\sigma_{kl}^3} \right. \qquad (6a)
$$
$$
\times \prod_{\substack{l=1 \\ l\neq i}}^{n} \left(\frac{1}{\sigma_{kl}} \exp \cdot \left\{ -\frac{1}{2} [\mu_{lj}(1+q) - \mu_{lk}]^2 / \sigma_{kl}^2 \right\} \right) - \frac{\mu_{lj}(1-q)-\mu_{lk}}{\sigma_{kl}^3} \prod_{\substack{l=1 \\ l\neq i}}^{n} \left(\frac{1}{\sigma_{kl}} \exp \right.
$$

$$\times\left[-\frac{1}{2}\{\mu_{lj}(1-q)-\mu_{lk}\}/\sigma_{kl}^2\right]\Bigg\}\frac{1}{(\sqrt{2\pi})^n}\sum_{j=1}^{M}(M-1)C_{jt}\,p_j\sum_{i=1}^{n}2\left\{\frac{-\mu_{ij}q}{\sigma_{ji}^3}\prod_{\substack{l=1\\ l\neq j}}^{M}\frac{1}{\sigma_{jl}}\exp.\right.$$

$$\left.\times\left[-\frac{1}{2}\frac{(\mu_{ij}q)^2}{\sigma_{jl}^2}\right]\right\}.$$

It is obvious from (6a) that the direct determination of the dependence $R(q)$ by the equation (3) should probably prove to be easier, even in this simplified case, than the solution of (6a), or of (6) respectively. For that purpose it is appropriate to calculate the integrals in (3a) by the Monte Carlo method, which in this case offers considerable simplification, because the integration regions are simple volume-parallelepipeds orientated along the coordinate axes.

The deciding rule at constant limits, defined by a multidimensional parallelepiped, is reduced to a verification of the inequation system:

(7) $$a_{ij}\leq x_i\leq b_{ij},\quad i=1,\ldots,n,\quad j=1,\ldots,M,$$

where $X(x_1,\ldots,x_n)$ is a vector-realization, subject to classification.

If (7) is satisfied for certain j and for all i, it is accepted that x belongs to the j-class of the multitude M.

The mean risk function (averaged for all the meanings of j) is determined under this deciding rule by equation (3); it can be obtained also by (3) only for the j-class, though without summing along index j.

If the change of the limits a_{ij} and b_{ij} of the multidimensional parallelepiped occurs depending on more than one parameter, then equation (6) is transformed into a system of partial derivatives of R towards these parameters.

Of course, the deciding rule with constant limits can be applied and can be optimized for other relatively simple limiting surfaces, as for example a multidimensional sphere, an ellipsoid, etc. In this case inequation (7) would become more complicated. Some complications would also appear in the calculation of R by (3). It is possible that the complexity of the deciding rule should become commensurable with the linear deciding rules.

On account of the rather great simplification of the deciding rule at a multidimensional parallelepiped, the risk R at the classificational analysis increases. If an acceptable risk R_{acc} is given, this risk in the general case will be realized with the greatest number of signs, i. e. the dimensionality of S would be greatest, under a deciding rule based on a multidimensional parallelepiped. This compromise would probably be acceptable at considerable data files, where it could happen that it was more profitable economically to measure more signs x_i but needing considerably lesser time of computing analysis. In particular, such a situation can appear when studying the natural formations of the Earth surface according to spectral reflective characteristics. In that case, to obtain data by a greater number of wavelengths would involve a single complication of the equipment for obtaining reflective characteristics (increase in the number of channels for obtaining spectral information). This single complication of the design would be compensated by its multiple applications in collecting and processing considerable data files.

Equation (3) shows that the risk function depends on the number M of the classes and increases with the increases of M. That is why it is convenient to introduce the quantity "relative risk" for the comparative analysis of different in volume sets of classes:

$$(8)\qquad \xi = R/\sum_{m=1}^{M}\sum_{\substack{k=1\\k\neq m}}^{M}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[p_m C_{mk} f(X/\mu_m)+p_k C_{km} f(X/\mu_k)]\,dx$$

$$= R/2\sum_{m=1}^{M}\sum_{\substack{k=1\\k\neq m}}^{M} C_{mk}\,p_m.$$

Taking into consideration equation (5), we find that ξ changes in the interval $0\leq\xi\leq 0.5$. The value $\xi=0.5$ corresponds to the maximum indeterminacy in the class identification. The same value is obtained not only when the region determining the limits of the classes is with a zero volume or, respectively, is infinitely great, but also when all the distributions are equal. In this case we also have complete indeterminacy.

The maximum indeterminacy is obtained also in the case when $M\to\infty$. This is due to the fact that in equation (3) the integral values tend toward zero, because in the constant volume of integration there is a part of the infinite normal multidimensional distribution which tends toward zero.

References

1. Libenson, M. H., A. Y. Hessin, B. A. Yanson. Automation of space image interpretation. Moscow, Energia, 1975.
2. Milenkiy, A. B. Signal classification in indeterminacy conditions. — Sovetskoe radio, Moscow, 1975.

Оптимизация параллелепипедной дискриминантной функции в многомерном анализе

Т. К. Янев

(Резюме)

При классификации стохастических объектов основным критерием точности классификации является функция риска. В настоящем работе исследована проблема минимизации функции риска, когда границы классов стохастических объектов принимаются за многомерные параллелепипеды. Процедура минимизации охватывает параметры этих параллелепипедов.