THE BAYES DECISION RULE STABILITY UNDER INACCURATELY DETERMINED CLASS PRIOR PROBABILITIES

E.E. Zhuk
Belarusian State University
Minsk, BELARUS
e-mail: zhukee@mail.ru

Abstract

The Bayes decision rule stability is investigated under inaccurately determined class prior probabilities. The risk bias is analytically evaluated by the risk asymptotic expansion method. The obtained results are illustrated for the well known Fisher model.

1 Introduction: mathematical model and classification problem

Let random observations $x \in R^N$ from $L \geq 2$ classes $\{\Omega_1, \dots, \Omega_L\}$ be registered in R^N $(N \geq 1)$. According to the mathematical model [1, 2] observation $x \in R^N$ belongs to the class with unknown random class index $d^o \in S$ ($S = \{1, \dots, L\}$ is the set of class indices):

$$\mathbf{P}\{d^o = i\} = \pi_i^o > 0, \ i \in S; \tag{1}$$
$$\sum_{i \in S} \pi_i^o = 1,$$

where $\{\pi_i^o\}_{i \in S}$ are class prior probabilities [1, 2]. Under the fixed index $d^o = i$ $(i \in S)$ observation x from the class Ω_i is described by the conditional probability density function [1, 2]:

$$p_i(x) \geq 0, \ x \in R^N : \quad \int_{R^N} p_i(x) dx = 1, \ i \in S. \tag{2}$$

The statistical classification problem consists in the construction of the decision rule (DR) [1, 2]: $d = d(x) : R^N \to S$, which is a statistical estimator for unknown class index $d^o \in S$ of observation $x \in R^N$.

As the efficiency measure of the DR $d = d(x) \in S$, $x \in R^N$, the risk (the expected losses) is used [1, 2]: $r = r(d) = \mathbf{E}\{w_{d^o, d(x)}\}$, where $W = (w_{ij})_{i,j \in S}$ is the loss matrix: w_{ij} is the loss value when observations from the class Ω_i $(d^o = i)$ are classified to the class Ω_j $(d(x) = j)$. It is well known [1, 2], the Bayes DR (BDR):

$$d_o(x) = d(x; \pi^o) = \arg\min_{j \in S} f_j(x; \pi^o), \quad x \in R^N; \tag{3}$$

$$f_j(x; \pi^o) = \sum_{i \in S} \pi_i^o w_{ij} p_i(x), \quad j \in S,$$

has the minimum risk:

$$r_o = r(d_o) = \int_{R^N} \min_{j \in S} f_j(x; \pi^o) dx. \tag{4}$$

Often in practice the class prior probabilities $\{\pi_i^o\}_{i \in S}$ from (1) are inaccurately determined (for example, the so-called "expert judgements" [1]):

$$\pi_i = \pi_i^o + \varepsilon_i, \ i \in S; \tag{5}$$
$$\sum_{i \in S} \pi_i = 1,$$

where mistakes $\{\varepsilon_i\}_{i \in S}$ satisfy the conditions:

$$-\pi_i^o < \varepsilon_i < 1 - \pi_i^o, \ i \in S; \tag{6}$$
$$\sum_{i \in S} \varepsilon_i = 0,$$

and are characterized by the value

$$\varepsilon_+ = \max_{i \in S} |\varepsilon_i|, \tag{7}$$

which is named the mistake level. If $\varepsilon_+ = 0$, then $\pi_i = \pi_i^o$, $i \in S$, and the class prior probabilities are accurately determined.

Let us investigate the stability (in the sense of the risk) of the DR (3) under inaccurately determined class prior probabilities (5), (6).

2 Asymptotic investigations of the risk

Let the class prior probabilities $\pi = (\pi_1, \ldots, \pi_L)'$ from (5), (6) be used in the BDR $d(\cdot; \pi^o)$ from (3) instead of their true values $\pi^o = (\pi_1^o, \ldots, \pi_L^o)'$ from (1) ("$'$" is the transposition symbol). The efficiency of such DR $d(\cdot; \pi)$ is characterized by the following risk [1, 3]:

$$r(d) = \mathbf{E}\{w_{d^o, d(x;\pi)}\} = R(\pi, \pi^o) = \sum_{i \in S} \pi_i^o \sum_{j \in S} w_{ij} \int_{R^N} \prod_{\substack{k \in S \\ k \neq j}} U\left(f_{kj}(x; \pi)\right) p_i(x) dx, \tag{8}$$

where

$$f_{kj}(x; \pi) = f_k(x; \pi) - f_j(x; \pi), \quad x \in R^N, \quad k \neq j \in S, \tag{9}$$

$\{f_j(\cdot; \pi)\}_{j \in S}$ are determined in (3) and $U(z) = \{1, z \geq 0; \ 0, z < 0\}$ is the unit function.

Note, $r_o = R(\pi^o, \pi^o)$ is the risk (4) of the BDR $d(\cdot; \pi^o)$ from (3) (under $\pi = \pi^o$), and $r_o = R(\pi^o, \pi^o) \leq R(\pi, \pi^o)$, $\forall \, \pi$.

Introduce the notations ($k \neq j \in S$; $x \in R^N$):

$$\tilde{\Gamma}_{kj}(\pi) = \{x : \ f_{kj}(x; \pi) = 0\}; \quad U_{jk}(x; \pi) = \prod_{\substack{i \in S \\ i \neq j, i \neq k}} U\left(f_{ij}(x; \pi)\right); \tag{10}$$

$$\Gamma_{kj}(\pi) = \tilde{\Gamma}_{kj}(\pi) \cap \{x : \ U_{jk}(x; \pi) = 1\}$$

is the fragment of the surface $\tilde{\Gamma}_{kj}(\pi) \subset R^{N-1}$, which belongs to the domain $\{x : U_{jk}(x; \pi) = 1\} \subset R^N$ ($\Gamma_{12}(\pi) := \tilde{\Gamma}_{12}(\pi)$ under the case of two classes, $L = 2$).

Theorem. *Let the class characteristics $\{\pi_i^o, p_i(\cdot)\}_{i\in S}$ be determined (1), (2) and the following surface integrals be finite $(t, s, k, j \in S,\ k \neq j)$:*

$$J_{tskj}(\pi^o) = \int_{\Gamma_{kj}(\pi^o)} p_t(x)p_s(x)\left|\nabla_x f_{kj}(x;\pi^o)\right|^{-1} dS_{N-1} < +\infty, \tag{11}$$

where the notations (9), (10) are used and $\nabla_x f_{kj}(x;\pi^o) \in R^N$ is the vector of first order partial derivatives w.r.t. $x \in R^N$.

Then under the small mistake level (7) for the prior probabilities (5), (6): $\varepsilon_+ \to 0$, the risk $R(\pi,\pi^o)$ from (8) allows the asymptotic representation:

$$R(\pi,\pi^o) = r_o + \frac{1}{2}\sum_{t,s\in S}\sum_{j=2}^{L}\sum_{k=1}^{j-1}(w_{tk}-w_{tj})(w_{sk}-w_{sj})J_{tskj}(\pi^o)\varepsilon_t\varepsilon_s + O(\varepsilon_+^3), \tag{12}$$

where r_o is the risk (4) of the BDR (3).

Corollary. *If under the conditions of the theorem the loss matrix has the form: $W = (w_{ij})_{i,j\in S}$: $w_{ij} = \{0, i = j;\ 1, i \neq j\}$, then the DR $d(\cdot;\pi)$ has the form:*

$$d(x;\pi) = \arg\max_{i\in S}\{\pi_i p_i(x)\}, \quad x \in R^N, \tag{13}$$

and its risk $R(\pi,\pi^o) = \mathbf{P}\{d(x;\pi) \neq d^o\}$ is the classification error probability ($\varepsilon_+ \to 0$).

$$R(\pi,\pi^o) = r_o + \frac{1}{2}\sum_{j=2}^{L}\sum_{k=1}^{j-1}\left(J_{kkkj}(\pi^o)\varepsilon_k^2 + J_{jjkj}(\pi^o)\varepsilon_j^2 - 2J_{kjkj}(\pi^o)\varepsilon_k\varepsilon_j\right) + O(\varepsilon_+^3), \tag{14}$$

where

$$r_o = 1 - \int_{R^N}\max_{i\in S}\{\pi_i^o p_i(x)\}dx \tag{15}$$

is the risk $r_o = R(\pi^o,\pi^o) = \mathbf{P}\{d(x;\pi^o) \neq d^o\}$ of the BDR $d(\cdot;\pi^o)$ and $\{J_{tskj}(\pi^o)\}$ is the surface integrals (11) under $f_{kj}(x;\pi^o) = \pi_j^o p_j(x) - \pi_k^o p_k(x)$, $x \in R^N$, $k \neq j \in S$.

In practice the results (12), (14) of the theorem and its corollary allow to evaluate the risk bias $R(\pi,\pi^o) - r_o = O(\varepsilon_+^2) > 0$, which describes the effects of the mistakes (6) in the class prior probabilities (5).

3 The case of two classes and the Fisher model

Let us consider the case of two classes ($L = 2$, $S = \{1,2\}$) when the DR $d(\cdot;\pi)$ from (13) is used.

Under $L = 2$ the prior probabilities from (5), (6):

$$\pi_1 = \pi_1^o + \varepsilon, \quad \pi_2 = \pi_2^o - \varepsilon = 1 - \pi_1^o - \varepsilon, \tag{16}$$

where π_1^o, $\pi_2^o = 1 - \pi_1^o$ is the true prior probabilities from (1) and $\varepsilon_1 = -\varepsilon_2 = \varepsilon$ ($\varepsilon_+ = |\varepsilon|$) is the mistake ($-\pi_1^o < \varepsilon < 1 - \pi_1^o$).

The DR $d(\cdot;\pi)$ from (13) may be rewritten in the form:

$$d(x;\pi) = U(G(x;\pi)) + 1, \quad x \in R^N, \tag{17}$$

where $G(x;\pi) = f_{12}(x;\pi) = (1-\pi_1)p_2(x) - \pi_1 p_1(x)$.

And for the risk $R(\pi,\pi^o) = \mathbf{P}\{d(x;\pi) \neq d^o\}$ of the DR (17) from (14), (11) and

$$\Gamma_{12}(\pi^o) := \tilde{\Gamma}_{12}(\pi^o) = \Gamma(\pi^o) = \{x:\ G(x;\pi^o) = 0\} = \left\{x:\ p_2(x) = \frac{\pi_1^o}{1-\pi_1^o}p_1(x)\right\}$$

we obtain ($\varepsilon_+ = |\varepsilon|$):

$$R(\pi,\pi^o) = r_o + \frac{1}{2(1-\pi_1^o)^2}\int_{\Gamma(\pi^o)} (p_1(x))^2 \left|\nabla_x G(x;\pi^o)\right|^{-1} dS_{N-1}\varepsilon^2 + O(|\varepsilon|^3), \tag{18}$$

where

$$r_o = 1 - \pi_1^o - \int_{R^N} G(x;\pi^o)U(G(x;\pi^o))dx \tag{19}$$

is the Bayes risk $r_o = R(\pi^o,\pi^o) = \mathbf{P}\{d(x;\pi^o) \neq d^o\}$ from (15).

Now let us illustrate the obtained results for the often meeting in applications Fisher model [1, 3], when the conditional densities from (2) are supposed multivariate normal (Gaussian): $p_i(x) = n_N(x|\mu_i,\Sigma)$, $x \in R^N$, $i \in S$, with the various mathematical mean vectors (the class "centers") $\mu_i = \mathbf{E}\{x|d^o = i\} \in R^N$, $i \in S$, and the common for all classes non-singular covariance matrix $\Sigma \coloneqq \mathbf{E}\{(x-\mu_i)(x-\mu_i)'|d^o = i\}$ $(i \in S)$.

For the Fisher model the Bayes risk (19) is easily evaluated [1, 3]:

$$r_o = \pi_1^o\Phi\left(-\frac{\Delta}{2} - \frac{h}{\Delta}\right) + (1-\pi_1^o)\Phi\left(-\frac{\Delta}{2} + \frac{h}{\Delta}\right), \quad h = \ln\left(\frac{\pi_1^o}{1-\pi_1^o}\right), \tag{20}$$

where $\Phi(\cdot)$ is the standard Gaussian distribution function with the density $\varphi(z) = n_1(z|0,1) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)$, $z \in R$, and $\Delta = \sqrt{(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)}$ is the so-called Mahalanobis interclass distance. In this case from the relation (18) we obtain:

$$R(\pi,\pi^o) = r_o + \frac{\varphi\left(\frac{\Delta}{2} + \frac{h}{\Delta}\right)}{2\pi_1^o(1-\pi_1^o)^2\Delta}\varepsilon^2 + O(|\varepsilon|^3). \tag{21}$$

From (20), (21) it is seen that the classification stability increases (the risk bias $R(\pi,\pi^o) - r_o$ decreases) under increasing the Mahalanobis distance Δ and decreasing the mistake level $\varepsilon_+ = |\varepsilon|$.

References

[1] Aivazyan S.A., Buchstaber V.M., Yenyukov I.S., Meshalkin L.D. (1989). *Applied Statistics: Classification and Dimensionality Reduction*. Finansy i Statistika, Moskow.

[2] Fukunaga K. (1990). *Introduction to Statistical Pattern Recognition*. Academic Press, New York.

[3] Kharin Yu.S., Zhuk E.E. (2005). *Mathematical and Applied Statistics*. BSU, Minsk.