ASSIGNMENT OF MULTIVARIATE SAMPLES TO THE FIXED CLASSES BY THE MAXIMUM LIKELIHOOD METHOD AND ITS RISK

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

Abstract

The problem of statistical assignment of samples of identically distributed multivariate observations to the classes determined by their probability distributions is considered. The decision rule by the maximum likelihood method is proposed and its risk is analytically evaluated. The obtained results are illustrated for the well known Fisher model.

1 Introduction: mathematical model and problem of assignment of multivariate samples

Let $L \geq 2$ classes $\{\Omega_1, \ldots, \Omega_L\}$ be determined in the observation space R^N ($N \geq 1$). According to the traditional mathematical model [2, 3] the class Ω_i contains identically distributed random observations $x \in R^N$ described by the probability density function:

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

which determines Ω_i and is called a conditional density [2] of the class Ω_i ($i \in S$, $S = \{1, \ldots, L\}$ is the set of class indices).

Usually the statistical classification problem consists in the construction of the decision rule (DR) [2, 3]: $d = d(x): \quad R^N \to S$, which is a statistical estimator for unknown class index $d^o \in S$ of observation $x \in R^N$. It is supposed that classified observation belongs to one of the classes $\{\Omega_i\}_{i \in S}$ and is described by one of the densities (1).

But, often in practice [2, 4] we need to assign to the classes $\{\Omega_i\}_{i \in S}$ observations with a density

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

which is distinguished from the densities (1).

In this paper the following more common problem is investigated. Let a random sample $X = \{x_1, ..., x_n\}$ of size n consists of independent in total random observations $x_t \in R^N$, $t = \overline{1, n}$, described by some density (2). The problem consists in assignment of this sample to one of the classes $\{\Omega_i\}_{i \in S}$ determined by the densities (1).

2 Maximum likelihood method and its risk

To solve the assignment problem the maximum likelihood method can be used [1, 2, 3]:

$$D(X) = \arg\max_{i\in S} P_i(X); \tag{3}$$

$$P_i(X) = \prod_{t=1}^{n} p_i(x_t), \quad i \in S,$$

where the maximum likelihood DR (MLDR) (3): $D = D(X):\ R^{nN} \to S$, assigns the sample X to the class with index $D(X) \in S$. The functions $\{P_i(X)\}_{i\in S}$ from (3) are the conditional likelihood functions [2, 3], which mean conditional probability densities of sample X evaluated for the classes $\{\Omega_i\}_{i\in S}$.

Note, the unconditional ("real") likelihood function of the sample X is

$$P(X) = \prod_{t=1}^{n} p(x_t),$$

where $p(\cdot)$ is the density (2).

Now let us define the generalization of the traditional risk [2, 3] as an efficiency measure of the MLDR (3). Introduce the following notations: Ω_0 is the additional $(L+1)$-th class which contains observations from the sample X and is determined by the density (2) ($S_0 = \{0\} \cup S = \{0, 1, \ldots, L\}$ is the extended set of class indices);

$$R_{ij} = R(\Omega_i, \Omega_j) \geq 0, \quad i, j \in S_0, \tag{4}$$

are any interclass distances [2] (R_{ij} is the nearness measure for the classes Ω_i and Ω_j); $U(z) = \{1, z \geq 0;\ 0, z < 0\}$ is the unit function.

The risk of the MLDR (3) means the maximum probability to do not assign the sample X to one of the classes from $\{\Omega_i\}_{i\in S}$ which are the most close to the class Ω_0:

$$r = 1 - \min_{i\in I} P^{(i)}, \quad I = \{k:\ R_{0k} = \min_{j\in S} R_{0j}\}, \tag{5}$$

where

$$P^{(i)} = \mathbf{P}\{D(X) = i\} = \int_{R^N} \prod_{\substack{k\in S\\ k\neq i}} U\left(P_i(X) - P_k(X)\right) P(X) dX, \quad i \in S, \tag{6}$$

mean the probabilities to assign by the MLDR (3) the sample to the classes $\{\Omega_i\}_{i\in S}$.

If all distances $\{R_{0i}\}_{i\in S}$ between the class Ω_0 and the classes $\{\Omega_i\}_{i\in S}$ are distinguished then the risk (5) is simplified:

$$r = 1 - P^{(k)}, \quad k = \arg\min_{j\in S} R_{0j}. \tag{7}$$

The smaller is the risk r from (5) or (7) ($0 \leq r \leq 1$), the more efficiency is the MLDR (3) (the greater is the probability to assign by the MLDR (3) the sample X to the most closed class).

Note, if the sample X contains observations from one of the classes $\{\Omega_i\}_{i\in S}$ then the assignment problem is the so-called problem of group classification [1] and the risk (5), (6) means the conditional error probability:

$$r = 1 - P^{(i)} = \mathbf{P}\{D(X) \neq i | d^o = i\}, \quad i \in S, \tag{8}$$

where $d^o = i$ is the class index of observations from the sample X ($p(\cdot) \equiv p_i(\cdot)$, $P(\cdot) \equiv P_i(\cdot)$).

3 The case of the Fisher model

Now let us consider the often meeting in applications Fisher model [2, 3], when the conditional densities of the classes $\{\Omega_i\}_{i\in S}$ from (1) are supposed multivariate normal (Gaussian):

$$p_i(x) = n_N(x|\mu_i, \Sigma) = \tag{9}$$
$$= (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu_i)'\,\Sigma^{-1}(x-\mu_i)\right), \; x \in R^N, \; i \in S,$$

where

$$\mu_i = \mathbf{E}\{x | d^o = i\} = \int_{R^N} x\, p_i(x) dx, \quad i \in S,$$

are the conditional mathematical mean N-vectors (the "centers" [2, 3] of the classes $\{\Omega_i\}_{i\in S}$) and

$$\Sigma = \mathbf{E}\{(x-\mu_i)(x-\mu_i)' | d^o = i\}, \quad i \in S,$$

is the common for all classes $\{\Omega_i\}_{i\in S}$ non-singular covariance $(N \times N)$-matrix ("$'$" is the transposition symbol).

The observations from the sample $X = \{x_t\}_{t=1}^n$ are supposed to be normal too:

$$p(x) = n_N(x|\mu, \Sigma), \quad x \in R^N, \tag{10}$$

where

$$\mu = \mu_0 = \int_{R^N} x\, p(x) dx \in R^N$$

is their mathematical mean vector (the "center" of the $(L+1)$-th class Ω_0) and Σ is their covariance matrix: $\int_{R^N}(x-\mu)(x-\mu)'\,p(x)dx = \Sigma$.

Under the Fisher model (9) as the interclass distances (4) let us define the Mahalanobis distances between the class "centers" [2, 3]:

$$R_{ij} = R(\Omega_i, \Omega_j) = \rho(\mu_i, \mu_j), \quad i, j \in S_0, \tag{11}$$

where

$$\rho(y, z) = \sqrt{(y-z)'\Sigma^{-1}(y-z)}, \quad y, z \in R^N, \tag{12}$$

is the Mahalanobis metric in R^N.

Theorem. *Under the Fisher model (9) the MLDR (3) takes the form:*

$$D(X) = \arg\min_{i \in S} \rho(\bar{x}, \mu_i), \quad X \in R^{nN}, \tag{13}$$

where $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$ *is the sample average.*

If observations from the sample $X = \{x_t\}_{t=1}^{n}$ *are independent in total with the density (10) and the Mahalanobis interclass distances (11), (12) are used then the risk (5), (6)*

$$r = 1 - \min_{i \in I} P^{(i)}, \quad I = \{k : \rho(\mu, \mu_k) = \min_{j \in S} \rho(\mu, \mu_j)\};$$

$$P^{(i)} = \int_{R^N} \prod_{\substack{k \in S \\ k \neq i}} U\left(\rho(x, \mu_k) - \rho(x, \mu_i)\right) n_N\left(x | \mu, \frac{1}{n}\Sigma\right) dx, \quad i \in S.$$

Note that the MLDR (13) obtained under the Fisher model is also known as the L-means DR [2, 3].

Corollary. *Under the conditions of the theorem in the case of two classes (*$L = 2$*):*

$$r = \Phi\left(-\sqrt{n}\frac{|\rho^2(\mu, \mu_1) - \rho^2(\mu, \mu_2)|}{2\rho(\mu_1, \mu_2)}\right), \tag{14}$$

where $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} \exp\left(-\frac{w^2}{2}\right) dw$, $z \in R$ *is the standard Gaussian distribution function.*

From the relation (14) it is seen that the risk r has the maximum ("worst") value when $\rho(\mu, \mu_1) = \rho(\mu, \mu_2)$: $r = \Phi(0) = 1/2$. If observations from the sample X belong to one of the classes Ω_1 or Ω_2 ($\mu = \mu_1$ or $\mu = \mu_2$) then the risk (14) is the conditional error probability (8):

$$r = \Phi\left(-\sqrt{n}\frac{\rho(\mu_1, \mu_2)}{2}\right),$$

where $\rho(\mu_1, \mu_2) = \sqrt{(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)}$ is the Mahalanobis interclass distance between the classes Ω_1 and Ω_2.

References

[1] Abusev R.A., Lumelsky Ya. P. (1987). *Statistical Group Classification*. PSU, Perm.

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

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

[4] Zhuk E.E., Kharin Yu.S. (1998). *Robustness in Cluster Analysis of Multivariate Observations*. BSU, Minsk.