Liet. matem. rink, **47**, *spec. nr.*, 2007, 364–368

Discriminant analysis of the equicorrelated Gaussian observations

Kęstutis DUČINSKAS, Jurgita NEVERDAUSKAITĖ (KU)
e-mail: kestutis.ducinskas@ku.lt

Abstract. In this paper the problem of classification of an observation into one of two Gaussian populations with different means and common variance is considered in the case when equicorrelated training sample is given. Unknown means and common variance are estimated from training sample and these estimators are pluged in the Bayes discriminant function. The maximum likelihood estimators are used. The approximation of the expected error rate associated with Bayes plug-in discriminant function is derived. Numerical analysis of the accuracy of that approximation for various values of correlation is presented.

Keywords: equicorrelation, Bayes discriminant function, actual error rate, expected error rate.

Introduction

Discriminant analysis (DA) sometimes called supervised classification traditionally assumes that observations to be classified and observations in training sample are independent. However, in practical situations with temporally and spatially distributed data this is usually not the case. Data that are close together in time or space, are likely to be correlated, at best equicorrelated [4, 5]. Equicorrelation arises naturally from physical and biological considerations [1, 2]. Thus, to include temporal or spatial dependencies in the classification problem is very important.

In this paper, we consider the performance of the plug-in linear Bayes discriminant function (PBDF) when the parameters are estimated from training samples as realizations of a equicorrelated Gaussian random process. We use the maximum likelihood (ML) estimators of unknown parameters of means and common variance assuming that the correlation is known.

1. The main concepts and definitions

The model of $Z(s)$ in population Ω_l is

$$Z(s) = \beta_l' x(s) + \varepsilon(s), \quad s \in D \subset I,$$

where $x(s)$ is a $q \times 1$ vector of non random regressors whose first element is 1 and β_l is a $q \times 1$ vector of parameters, $l = 1, 2$ and I is index set. The error term $\{\varepsilon(s)\colon\ s \in D \subset I\}$ is zero-mean stationary spatial Gaussian random process with covariance function defined by model for all $s, u \in D$

$$cov\{\varepsilon(s), \varepsilon(u)\} = \begin{cases} \sigma^2\rho & \text{if } s \neq u, \\ \sigma^2 & \text{if } s = u, \end{cases}$$

where σ^2 is constant variance and ρ is the intraclass and the interclass correlation. Consider the problem of classification of the observation $Z^0 = Z(s_0)$, with $s_0 \in D$, into one of two populations specified above. Assume that training sample T is also given. Since the observation Z^0 is equicorrelated with observations from training sample, we have to deal with conditional means and variance

$$\mu^0_{lT}(s_0;\beta) = E(Z^0/T;\Omega_l), \quad \sigma^2_{0T}(\sigma^2) = V(Z^0/T;\Omega_l), \quad l = 1,2. \tag{1}$$

Suppose that we observe the training sample $T' = (T'_1, T'_2)$, where T_l is the $n_l \times 1$ vector of n_l observations of $Z(s)$ from $\Omega_l, l = 1,2$. Then T is the $n \times 1$ vector, where $n = n_1 + n_2$. Assume that $2q \times 1$ parameter vector $\beta' = (\beta'_1, \beta'_2)$ and σ^2 are unknown and ρ is known.

Let $\hat\beta$ and $\hat\sigma^2$ be the estimators of β and σ^2, respectively, based on T. Denote the $2q \times 1$ vector of parameters by $\Psi' = (\beta', \sigma^2)$ and denote the vector of their estimators by $\hat\Psi' = (\hat\beta', \hat\sigma^2)$. Let π_1, π_2 be prior probabilities of Ω_1 and Ω_2.

The plug-in BDF is obtained by replacing the parameters in (1) with their estimators. Then PBDF to the classification problem specified above is

$$W(Z^0;\hat\Psi) = \Big(Z^0 - \frac{1}{2}(\hat\mu^0_{lT} + \hat\mu^0_{2T})\Big)(\hat\mu^0_{lT} - \hat\mu^0_{2T})/\hat\sigma^2_{0T} + \gamma, \tag{2}$$

where

$$\hat\mu^0_{lT} = \mu^0_{lT}(s_0;\hat\beta), \quad \hat\sigma^2_{0T} = \sigma^2_{0T}(\hat\sigma^2),$$
$$\gamma = ln(\pi_1/\pi_2).$$

In the considered case the actual error rate for $W(Z^0;\hat\Psi)$ can be rewritten as

$$P(\hat\Psi) = \sum_{l=1}^{2}\big(\pi_l \Phi(\hat Q_l)\big), \tag{3}$$

where $\Phi()$ is the standard normal distribution function, and

$$\hat Q_l = (-1)' \frac{(\mu^0_{lT} - \frac{1}{2}(\hat\mu^0_{lT} + \hat\mu^0_{2T}))(\hat\mu^0_{lT} - \hat\mu^0_{2T})/\hat\sigma^2_{0T} + \gamma}{\sqrt{(\hat\mu^0_1 - \hat\mu^0_2)^2\sigma^2_{0T}/(\hat\sigma^2_{0T})^2}}, \quad l = 1,2. \tag{4}$$

DEFINITION 1. The expectation of the actual error rate with respect to the distribution of T, designated as $E_T\{P(\hat\Phi)\}$, is called the expected error rate (EER).

Hence the EER for the considered problem of Z^0 classification by PBDF is

$$E_T(P(\hat\Phi)) = E_T\Big\{\sum_{l=1}^{2}\big(\pi_l \Phi(\hat Q_l)\big)\Big\}.$$

2. The proposed approximation

Suppose that the model of T is

$$T = X\beta + E,$$

where $\beta' = (\beta_1', \beta_2')$ and E is the n-vector of random errors that has multivariate Gaussian distribution $N_n(0, \sigma^2, R)$.

The ML estimator of β and bias adjusted ML estimator of σ^2 [3] are

$$\hat{\beta} = (X^T R^{-1} X)^{-1} X^T R^{-1} T, \tag{5}$$

$$\hat{\sigma}^2 = (T - X\hat{\beta}) R^{-1} (X - X\hat{\beta})/(n - 2q). \tag{6}$$

where R is the coreelation matrix of T.

The Mahalanobis distance between conditional distributios of Z^0 specified by is

$$\Delta_0 = \left|(\mu_{1T}^0 - \mu_{2T}^0)/\sigma_{0T}\right|.$$

Put $R_\beta = (X' R^{-1} X)^{-1}$, $\rho_0 = \rho/(1 - \rho + n\rho)$.

Denote the approximation of EER $E_T\{P(\hat{\Phi})\}$ by AER.

THEOREM 1. *Suppose that observation Z^0 is classified by PBDF defined in* (2) *and let ML estimators of parameters specified in* (5), (6) *be used. Then the approximations of EER based on Taylor series expansion is*

$$AER = P_B + \pi_1 \phi(-\Delta_0/2 - \gamma/\Delta_0) \times \left\{[\rho_0 X' 1_n - \gamma_1 \times x_0]' R_\beta [\rho_0 X' 1_n - \gamma_1 \times x_0] + 2\gamma^2/(n - 2q)\right\}/(2\Delta_0),$$

where P_B is Bayes error and

$$\gamma_1' = (\Delta_0/2 + \gamma/\Delta_0, \Delta_0/2 - \gamma/\Delta_0),$$
$$x_0 = x(s_0).$$

Proof. Taylor series expansion of the actual error rate given by formulas (3) an d (4) up to the second order derivatives about true values of parameters is used. Taking the expectation of it by distribution of training sample T the proof is completed.

Remark. In the case $\rho = 0$, the proposed approximation of EER agrees with the asymptotic approximations of EER for independent Gaussian observations case [5].

The order of the remainder term of the Taylor expansion depends on the sampling design of the training sample.

3. Numerical illustration and discussions

Numerical examples for comparison and evaluation of the accuracy of the derived approximations of EER is implemented for the constant means, i.e., $q = 1$ and $x(s) = 1$. Then $\mu_l(s) = \beta_l, l = 1, 2$ and $X = 1_{n_1} \oplus 1_{n_2}$.

Table 1. Values of AER, TER, AER/TER for the case $n_1 = n_2 = n_0$ and $\pi_1 = \pi_2 = 0.5$

ρ	$n_0 = 4$			$n_0 = 20$		
	AER	TER	AER/TER	AER	TER	AER/TER
	$\Delta = 0.2$			$\Delta = 0.2$		
0	0.46265	0.49164	0.94105	0.46067	0.48140	0.95694
0.1	0.46047	0.49027	0.93922	0.45854	0.47779	0.95970
0.2	0.45805	0.48801	0.93860	0.45603	0.47252	0.96511
0.3	0.45515	0.48463	0.93917	0.45301	0.46516	0.97388
0.4	0.45157	0.47944	0.94187	0.44927	0.45462	0.98923
0.5	0.44697	0.47100	0.94897	0.44445	0.43898	1.01247
0.6	0.44075	0.45613	0.96629	0.43795	0.41503	1.05522
	$\Delta = 1$			$\Delta = 1$		
0	0.31954	0.34720	0.92034	0.31074	0.31100	0.99917
0.1	0.30986	0.33004	0.93886	0.30133	0.29359	1.02636
0.2	0.29917	0.30482	0.98147	0.29042	0.27081	1.07242
0.3	0.28661	0.27300	1.04991	0.27750	0.24280	1.14292
0.4	0.27142	0.23424	1.15875	0.26188	0.20800	1.25899
0.5	0.25250	0.18784	1.34425	0.24245	0.16456	1.47338
0.6	0.22803	0.13253	1.72056	0.21744	0.11128	1.95401

The Mahalanobis distance between marginal distributions of Z^0 is specified by

$$\Delta = \left|(\beta_1 - \beta_2)/\sigma\right|.$$

With an insignificant loss of generality the cases with $n_1 = n_2 = n_0$, $\pi_1 = 0.5$. Computed values of proposed approximation AER is compared with theoretical values obtained by using the procedures of numerical integration of Maple 9.5. Denote these theoretical values by TER.

The values of AER, TER and AER/TER are given in Table 1 for various values of Δ and ρ with $n_0 = 4$ and 20.

From Table 1 it is evident for both $n_0 = 4$ and 20, the values of AER and TER decreases as Δ and ρ increases.

The reason of this effect is the increasing of the Mahalanobis distance Δ_0 when Δ and ρ increases.

It is also evident from Table 1 that deviation of AER from TER evaluated by and ratio AER/TER show the high accuracy of proposed approximation for $\Delta = 0.2$ and $\Delta = 1$ and all selected values of ρ.

References

1. P.S. Gill, S.G. Banneheka, T.B. Swartz, Test concerning equicorrelation matrices with grouped normal data, *Comm. Statist. Theory Methods*, **34**, 857–873 (2005).
2. R. Khattree, D.N. Nail, Estimation of interclass correlation under circular covariance, *Biometrica*, **81**(3), 612–616 (1994).
3. J.R. Magnus, H. Neudecher, *Matrix Differential Calculus and applications in Statistics and Econometrics*, Wiley, New York (2002).

4. K.V. Mardia, Spatial discrimination and classification maps, *Comm. Statist. Theory Methods*, **13**(18), 2181–2197 (1974).
5. G.J. McLachlan, *Discriminant Analysis and Statistical Patter Recognition*, Wiley, New York (2004).

REZIUMĖ

K. Dučinskas, J. Neverdauskaitė. Ekvikoreliuotų Gauso stebėjimų diskriminantinė analizė

Straipsnyje nagrinėjamas ekvikoreliuotų Gauso stebėjimų klasifikavimo uždavinys, kai klasės skiriasi tik regresiniais vidurkiais. Pateikta vidutinės klasifikavimo klaidos aproksimacija atvejui, kai nežinomi parametrai vertinami maksimalaus tikėtinumo metodu.