LIETUVOS MATEMATIKOS RINKINYS. LMD DARBAI
52 tomas, 2011, 315–320

ISSN 0132-2818
www.mii.lt/LMR/

Linear discriminant analysis of spatial Gaussian data with estimated anisotropy ratio

Lina Dreižienė

Department of Statistics, Klaipėda University
H. Manto 84, Klaipėda LT 92294
E-mail: l.dreiziene@gmail.com

Abstract. The paper deals with a problem of classification of Gaussian spatial data into one of two populations specified by different parametric mean models and common geometric anisotropic covariance function. In the case of an unknown mean and covariance parameters the Plug-in Bayes discriminant function based on ML estimators is used. The asymptotic approximation of expected error rate (AER) is derived in the case of unknown mean parameters and single unknown covariance parameter i.e., anisotropy ratio.

Keywords: Bayesian discriminant function, actual risk, expected risk, anisotropy ratio, Gaussian random field.

Introduction

In the case of completely specified populations and known loss function, an optimal classification rule in the sense of minimum risk is based on Bayesian discriminant function (BDF) [4]. In the practical situations the complete statistical description of populations usually is not possible. It is possible to estimate unknown parameters and plug-in them into BDF when using training sample. Plug-in BDF is called PBDF. The expressions for the expected error rate (ER) are very complicated even for the simplest forms of PBDF, therefore, asymptotic approximations of the ER are especially important.

Many authors have investigated the performance of the PBDF when parameters are estimated from training samples consisting of dependent observations (see e.g., [5]). Plug-in approach to discrimination for feature observations having elliptically contoured distributions is implemented in [2]. Šaltytė and Dučinskas [7] derived the asymptotic approximation of the expected error rate when classifying the observation of a scalar Gaussian random field into one of two classes with different regression mean models and common variance. However, the correlations between observations to be classified and training sample were assumed equal zero in the all publications listed above. This assumption is not correct in all situations, especially in cases where the locations of observations to be classified are close to ones of training sample. The first extension of above mentioned approximation to the case where spatial correlations between Gaussian observations to be classified and observations in training sample are not assumed equal zero is done in [3]. Here only the trend parameters and variance (parameter of covariance function) are assumed unknown. The extension of the latter approximation to the case of complete parametric uncertainty (all means and covariance function parameters are unknown) was implemented in [4]. In the present paper we derive closed form approximation of expected error rate in the case of estimated mean parameters and estimated anisotropy ratio.

1 The main concepts and definitions

The main objective of this paper is to classify the observations of Gaussian random field (GRF) $\{Z(s)\colon s \in D \subset R^p\}$ into one of two populations. Suppose that the model for observation Z(s) in population Ω_j is

$$Z(s) = x'(s)\beta_j + \varepsilon(s), \tag{1}$$

where $x(s)$ is a $q \times 1$ vector of non random regressors and is a vector of parameters, $j = 1, 2$. The error term is generated by zero-mean stationary GRF with covariance function defined by model for all $s, u \in D$

$$\operatorname{cov}\big\{\varepsilon(s), \varepsilon(u)\big\} = C(s - u; \theta), \tag{2}$$

where $\theta \in \Theta$ is a $p \times 1$ parameter vector, Θ being an open subset of R^p.

For the given training sample, consider the problem of classification of the $Z_0 = Z(s_0)$ into one of two populations when the training sample T is given and

$$x'(s_0)\beta_1 \neq x'(s_0)\beta_2, \quad s \in D. \tag{3}$$

Denote by $S_n = \{s_i \in D;\ i = 1, \dots, n\}$ the set of locations where training sample $T' = (Z(s_1), \dots, Z(s_n))$ is taken, and call it the set of training locations. We shall assume the deterministic spatial sampling design and all analyses are carried out conditional on S_n.

S_n is partitioned into union of two disjoint subsets, i.e., $S_n = S^{(1)} \cup S^{(2)}$, where $S^{(j)}$ is the subset of S_n that contains n_j locations of feature observations from Ω_j, $j = 1, 2$.

This is the case where spatial classified training data is collected at fixed locations (stations).

The $n \times 2q$ design matrix of the training sample T denoted by X is specified by

$$X = X_1 \oplus X_2,$$

where symbol $\oplus$ denotes the direct sum of matrices and X_j is the $n_j \times q$ matrix of regressors for observations from Ω_j, $j = 1, 2$.

So the model of the training sample is

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

where $\beta = (\beta_1', \beta_2')'$ is a $2q \times 1$ vector of regression parameters and E is the $n \times 1$ vector of random errors that has multivariate Gaussian distribution $N_n(0, C(\theta))$.

Denote by $c_0(\theta)$ the covariance between Z_0 and T. Let t denote the realization of T.

For notational convenience, the argument θ in all its functions is now dropped.

Since Z_0 follows model specified in (1), the conditional distribution of Z_0 given $T = t$, Ω_j is Gaussian with mean μ_{lt}^0 and variance $\sigma_0^2(\theta)$

$$\mu_{lt}^0 = E(Z_0|T = t; \Omega_j) = x_0'\beta_j + \alpha_0'(t - X\beta), \quad j = 1, 2, \tag{4}$$

$$\sigma_0^2(\theta) = \operatorname{var}(Z_0|T = t; \Omega_j) = C(0) - c_0'C^{-1}c_0, \tag{5}$$

where $x_0' = x'(s_0)$ and $\alpha_0' = c_0'C^{-1}$.

Under the assumption of complete parametric certainty of populations and for known finite nonnegative losses $\{L(i,j),\ i,j = 1,2\}$, the BDF minimizing the risk of classification is formed by log ratio of the conditional likelihoods.

Then Bayes discriminant function (BDF) is

$$W_t(Z_0; \Psi) = \left(Z_0 - \frac{1}{2}\left(\mu_{1t}^0 + \mu_{2t}^0\right)\right)\left(\mu_{1t}^0 - \mu_{2t}^0\right)/\sigma_0^2 + \gamma, \tag{6}$$

here $\gamma = \ln(\pi_1^*/\pi_2^*)$, $\Psi = (\beta', \theta')'$, $\pi_j^* = \pi_j(L(j, 3-j) - L(j,j))$, $j = 1, 2$. π_1, π_2 are prior probabilities of the populations Ω_1 and Ω_2, respectively.

In the practical applications not all statistical parameters of populations are known. The PBDF is constructed by replacing parameters in the BDF with their estimators.

Let $\hat{\beta}$, $\hat{\theta}$ be the estimators of corresponding parameters from training sample T. Then the PBDF has the following form

$$W_t(Z_0; \hat{\Psi}) = \left(Z_0 - \hat{\alpha}_0'(t - X\hat{\beta}) - \frac{1}{2}x_0'H\hat{\beta}\right)\left(x_0'G\hat{\beta}\right)/\hat{\sigma}_0^2 + \gamma, \tag{7}$$

with $H = (I_q, I_q)$ and $G = (I_q, -I_q)$, where I_q denotes the identity matrix of order q.

Definition 1. The actual risk for PBDF $W_t(Z_0; \hat{\Psi})$ is defined as

$$P(\hat{\Psi}) = \sum_{i=1}^{2}\sum_{j=1}^{2} \pi_i L(i,j)\hat{P}_{ij}, \tag{8}$$

where for $i, j = 1, 2$

$$\hat{P}_{ij} = P_{it}\left((-1)^j W_t(Z_0; \hat{\Psi}) < 0\right). \tag{9}$$

The actual risk specified in (8), (9) for $W_t(Z_0; \hat{\Psi})$ specified in (7) is (see e.g. [4])

$$P(\hat{\Psi}) = \sum_{j=1}^{2} \left(\pi_j^* \Phi(\hat{Q}_j) + \pi_j L(j,j)\right), \tag{10}$$

and

$$\hat{Q}_j = (-1)^j \left((a_j - \hat{b}) \operatorname{sgn}\left(x_0'G\hat{\beta}\right) + \hat{\sigma}_0^2\gamma/\left|x_0'G\hat{\beta}\right|\right)/\sigma_0, \tag{11}$$

where for $j = 1, 2$ $a_j = x_0'\beta_j + \alpha_0'(t - X\beta)$, $\hat{b} = \hat{\alpha}_0'(t - X\hat{\beta} + x_0'H\hat{\beta})/2$.

The expectation of the actual risk with respect to the distribution of T is called the expected risk (ER) and is designated as $E_T(P(\hat{\Psi}))$.

In this paper we assume that all true values of parameters β and single parameter of covariance (anisotropy ratio) are unknown. So we will use estimates of these unknown parameters form PBDF.

2 The asymptotic approximation of ER with estimated parameters

We will use the maximum likelihood (ML) estimators of parameters based on the training sample. The asymptotic properties of ML estimators established by Mardia and Marshall [6] under increasing domain asymptotic (increasing domain asymptotic is based on a growing observation region) framework and subject to some regularity conditions are essentially exploited. Hence, the ML estimator $\hat{\Psi}$ is weakly consistent and asymptotically Gaussian [4].

We make the following assumptions:

(A1) Assume that for large n $\hat{\beta} \sim AN_{2q}(\beta, J_\beta^{-1})$ and $\hat{\theta} \sim AN_k(\theta, J_\theta^{-1})$.

Here $J_\beta = X'C^{-1}X$ and (i,j)-th element of J_θ is $\mathrm{tr}(C^{-1}C_iC^{-1}C_j)/2$.

(A2) Training sample T and estimator $\hat{\theta}$ are asymptotically uncorrelated (see e.g., [1]).

Note that sufficient conditions for (A1) is formulated in [6]. Under (A1), (A2) the AER is derived in [4]. The accuracy of such a type approximation is examined in [3].

We consider the case where $\theta = \lambda$, where λ is anisotropy ratio.

Let Δ_0^2 be squared Mahalanobis distance between conditional distributions of Z_0 given $T = t$.

Denote $\Lambda' = \alpha_0'X - x_0'(H/2 + \gamma G/\Delta_0^2)$ and $K_\beta = \Lambda' J_\beta^{-1}\Lambda$.

Lemma 1. *Suppose that observation Z_0 to be classified by PBDF and let assumptions* (A1), (A2) *hold. Then the approximation of ER in the case of estimated unknown mean parameters and estimated unknown anisotropy ratio is*

$$AER = P(\Psi) + \pi_1^*\varphi(-\Delta_0/2 - \gamma/\Delta_0)\Delta_0(K_\beta + K_\lambda)/\bigl(2\sigma_0^2\bigr), \tag{12}$$

where

$$K_\lambda = \mathrm{tr}\left(CBJ_\lambda^{-1}B'\right) + \gamma^2\bigl(\bigl(\hat{\sigma}_0^2\bigr)_\lambda^{(1)}\bigr)'J_\theta^{-1}\bigl(\hat{\sigma}_0^2\bigr)_\lambda^{(1)}/\bigl(\Delta_0^2\sigma_0^2\bigr), \tag{13}$$

here B is the first order partial derivative of $\hat{\alpha}_0$ evaluated at the point $\hat{\lambda} = \lambda$

$$B = \partial\hat{\alpha}_0/\partial\hat{\lambda} = -C^{-1}C_\lambda^{(1)}\alpha_0 + C^{-1}(c_0)_\lambda^{(1)}. \tag{14}$$

$(\hat{\sigma}_0^2)_\lambda^{(1)}$ *denotes the first order partial derivative of $\hat{\sigma}_0^2$* (5) *evaluated at the point $\hat{\lambda} = \lambda$*

$$\bigl(\hat{\sigma}_0^2\bigr)_\lambda^{(1)} = -\bigl(c_0'\bigr)_\lambda^{(1)}\alpha_0 + \alpha_0'C_\lambda^{(1)}\alpha_0 - \alpha_0'(c_0)_\lambda^{(1)}. \tag{15}$$

$(c_0)_\lambda^{(1)}$ *and $C_\lambda^{(1)}$ in* (14)–(15) *are the first order partial derivatives of c_0 and C evaluated at the point $\hat{\lambda} = \lambda$.*

Proof. The proof of lemma is similar to the proof of Theorem 1 [4] by replacing θ with anisotropy ratio λ. □

Remark 1. Suppose we have the case of exponential spatial correlation function

$$r(h) = \exp\Bigl\{-\sqrt{h_x^2 + \lambda^2 h_y^2}/\varphi\Bigr\}.$$

Then the expression of B is

$$B = \frac{\lambda\sigma^2}{\varphi} C^{-1}(R \circ H\alpha_0 - r_0 \circ H_0). \tag{16}$$

Here φ is range parameter, $\circ$ represents the Hadamard product.

Let the covariance have form $C = \tau^2 I + \sigma^2 R$, where R is $n \times n$ matrix of correlations between components of T and (i,j)-th element is

$$R_{ij} = \exp\Big\{ -\sqrt{\big(h_x^{ij}\big)^2 + \lambda^2\big(h_y^{ij}\big)^2}/\varphi\Big\}, \quad i \neq j \text{ and } R_{ij} = 1 \text{ else.}$$

Let H be $n \times n$ matrix and (i,j)-th element of H is

$$H_{ij} = \big(h_y^{ij}\big)^2 / \sqrt{\big(h_x^{ij}\big)^2 + \lambda^2\big(h_y^{ij}\big)^2}, \quad i \neq j \text{ and } H_{ij} = 0 \text{ else.}$$

Here $h_x^{ij} = h_x^i - h_x^j$, $h_y^{ij} = h_y^i - h_y^j$.

Denote by H_0 $n \times 1$ matrix and (i,j)-th element of H_0 is

$$\big(h_y^{0j}\big)^2 / \sqrt{\big(h_x^{0j}\big)^2 + \lambda^2\big(h_y^{0j}\big)^2}$$

and r_0 is $n \times 1$ matrix of correlations between Z_0 and T, and (i,j)-th element is

$$\exp\Big\{ -\sqrt{\big(h_x^{0j}\big)^2 + \lambda^2\big(h_y^{0j}\big)^2}/\varphi\Big\}.$$

Here $h_x^{0j} = h_x^0 - h_x^j$, $h_y^{0j} = h_y^0 - h_y^j$, h_x^0 and h_y^0 are coordinates of s_0. Using formulas above $(c_0)_\lambda^{(1)}$ and $C_\lambda^{(1)}$ have the following expressions

$$C_\lambda^{(1)} = -\frac{\lambda\sigma^2}{\varphi} R \circ H, \qquad (c_0)_\lambda^{(1)} = -\frac{\lambda\sigma^2}{\varphi} r_0 \circ H_0.$$

References

[1] M. Abt. Estimating the prediction mean squared error in Gaussian stochastic processes with correlation structure. *Scand. J. Stat.*, **26**:563–578, 1999.

[2] A. Batsidis and K. Zografos. Discrimination of observations into one of two elliptic populations based on monotone training samples. *Metrica*, **64**:221–241, 2006.

[3] K. Dučinskas. Approximation of the expected error rate in classification of the Gaussian random field observations. *Stat. Probab. Lett.*, **79**:138–144, 2009.

[4] K. Dučinskas and L. Dreižienė. Supervised classification of the scalar Gaussian random field observations under a deterministic spatial sampling design. *Austrian J. Stat.*, **40**(1,2):25–36, 2011.

[5] C.R.O. Lawoko and G.L. McLachlan. Discrimination with autocorrelated observations. *Patt. Recogn.*, **18**(2):145–149, 1985.

[6] K.V. Mardia and R.J. Marshall. Maximum likelihood estimation of models for residual covariance in spatiall regression. *Biometrika*, **71**:135–146, 1984.

[7] J. Šaltytė and K. Dučinskas. Comparison of Ml and OLS estimators in discriminant analysis of spatially correlated observations. *Informatica*, **13**(2):297–238, 2002.

REZIUMĖ

Tiesinė diskriminantinė erdvinių Gauso duomenų analizė, naudojant anizotropijos koeficiento įvertinius
L. Dreižienė

Straipsnyje analizuojamas Gauso erdvinių duomenų klasifikavimo į vieną iš dviejų populiacijų, kurių vidurkio modeliai skirtingi, o geometriškai anizotropinė kovariacinė funkcija tokia pat, uždavinys. Naudojama tiesioginio pakeitimo Bajeso diskriminantinė funkcija (PBDF), kuri gaunama vietoj nežinomų parametrų tiesioginio pakeitimo būdu įstačius jų įvertinius. Straipsnyje pateikiama klasifikavimo rizikos aproksimacijos formulė atvejui, kai nežinomi vidurkio parametrai ir nežinomas anizotropijos koeficientas yra pakeisti jų įverčiais.

Raktiniai žodžiai: Bajeso diskriminantinė funkcija, klasifikavimo rizika, anizotropijos koeficientas, atsitiktinis Gauso laukas