Spatial Statistics 2015: Emerging Patterns

Error rates in multi-category classification of the spatial multivariate Gaussian data

Lina Dreižienė[ab]*, Kęstutis Dučinskas[b]

[a]*Institute of Mathematics and Informatics, Vilnius University, Akademijos str. 4, LT-08663 Vilnius, Lithuania*
[b]*Department of Mathematics and Statistics, Klaipėda University, H.Manto str. 84, LT-92294 Klaipėda, Lithuania*

Abstract

The problem of classifying a spatial multivariate Gaussian data into one of several categories specified by different regression mean models is considered. The classifier based on plug-in Bayes classification rule (PBCR) formed by replacing unknown parameters in Bayes classification rule (BCR) with category parameters estimators is investigated. This is the extension of the previous one from the two category case to the multiple category case. The novel close-form expressions for the Bayes misclassification probability and actual error rate associated with PBCR are derived. These error rates are suggested as performance measures for the classifications procedure.

The three-category case with feature modelled by bivariate stationary Gaussian random field on regular lattice with exponential covariance function is used for the numerical analysis. Dependence of the derived error rates on category parameters is studied.

Keywords: Gaussian random field; Bayes classification rule; Pairwise discriminant function; Actual error rate

1. Introduction

Misclassification probabilities and error rates in two-category discriminant analysis of spatial data have been recently investigated by several authors (see e.g. Šaltytė-Benth and Dučinskas [1], Dučinskas [2], Batsidis and Zografos [3]). Multi-category linear discriminant analysis of spatial data generated by univariate Gaussian random

---

* Corresponding author. Tel.: +370-685-44253;
 E-mail address: l.dreiziene@gmail.com

field (GRF) is considered in Dučinskas et al. [4]. Correct classification rates for linear discriminant analysis of multivariate GRF observations for relatively large category number (comparing with dimensionality of GRF) is derived in Dreižienė et al. [5]. In the present paper the above investigation is extended to the error rates for two cases of dimensional structure. We propose the method of multi-category discriminant analysis essentially exploiting the BCR. Parametric PBCR formed by replacing unknown parameters with their estimators in BCR is being used. The novel close-form expressions for the overall probability of misclassification and overall actual error rate associated with PBCR are derived. These error rates are suggested as performance measures for the classification procedure.

2. The main concepts and definitions

Suppose that every observation of p-variate GRF $\{Z(s) : s \in D \subset R^2\}$ belongs to one of L disjunctive categories, say $\Omega_1, \ldots, \Omega_L$. The model of observation $\mathbf{Z}(s)$ in category Ω_l, $l = 1, \ldots, L$ is $\mathbf{Z}(s) = \mathbf{B}'_l \mathbf{x}(s) + \boldsymbol{\varepsilon}(s)$, where $\mathbf{x}(s)$ is a $q \times 1$ vector of non random regressors and $\mathbf{B}_l$ is a $q \times p$ matrix of parameters.

The error term is generated by p-variate zero-mean stationary GRF $\{\boldsymbol{\varepsilon}(s) : s \in D\}$ with covariance function defined by model for all $s, u \in D$ is $\text{cov}\{\boldsymbol{\varepsilon}(s), \boldsymbol{\varepsilon}(u)\} = r(s-u)\boldsymbol{\Sigma}$, where $r(s-u)$ is the spatial correlation function and $\boldsymbol{\Sigma}$ is the variance–covariance matrix with elements $\{\sigma_{ij}\}$.

Consider the problem of classification of the vector of observation of $\mathbf{Z}$ at location s_0 denoted by $\mathbf{Z}_0 = \mathbf{Z}(s_0)$ into one of L populations specified above with given joint training sample $\mathbf{T}$.

Joint training sample is specified by $p \times n$ matrix $\mathbf{T}' = (\mathbf{T}'_1, \ldots, \mathbf{T}'_L)$, where $\mathbf{T}_l$ is the $n_l \times p$ matrix of n_l observations of $\mathbf{Z}(\cdot)$ from Ω_l, $l = 1, \ldots, L$, $n = \sum_{l=1}^{L} n_l$. The $n \times Lq$ design matrix corresponding to $\mathbf{T}$ is specified by $\mathbf{X} = \bigoplus_{l=1}^{L} \mathbf{X}_l$, where symbol $\oplus$ denotes the direct sum of matrices and $\mathbf{X}_l$ is the $n_l \times q$ matrix of regressors for $\mathbf{T}_l$, $l = 1, \ldots, L$.

Then the model of $\mathbf{T}$ is $\mathbf{T} = \mathbf{X}\mathbf{B} + \mathbf{E}$, where $\mathbf{B}' = (\mathbf{B}'_1, \ldots, \mathbf{B}'_L)$ is the $p \times Lq$ matrix of means parameters and $\mathbf{E}$ is the $n \times p$ matrix of random errors that has matrix-variate normal distribution i.e. $\mathbf{E} \sim N_{n \times p}(\mathbf{0}, \mathbf{R} \otimes \boldsymbol{\Sigma})$. Here $\mathbf{R}$ denotes the spatial correlation matrix among components (rows) of $\mathbf{T}$.

In the rest of the paper the realization (observed value) of training sample $\mathbf{T}$ will be denoted by $\mathbf{t}$.

Denote by $\mathbf{r}_0$ the vector of spatial correlations between $\mathbf{Z}_0$ and observations in $\mathbf{T}$, and set $\boldsymbol{\alpha}_0 = \mathbf{R}^{-1}\mathbf{r}_0$, $\rho = 1 - \mathbf{r}'_0 \boldsymbol{\alpha}_0$, $\boldsymbol{\mu}_l^0 = \mathbf{B}'_l \mathbf{x}_0$, $l = 1, \ldots, L$ and $\mathbf{M} = \mathbf{X}\mathbf{B}$.

Notice that in category Ω_l the conditional distribution of $\mathbf{Z}_0$ given $\mathbf{T} = \mathbf{t}$ is Gaussian, i.e. $(\mathbf{Z}_0 | \mathbf{T} = \mathbf{t}; \Omega_l) \sim N_p(\boldsymbol{\mu}_{lt}^0, \boldsymbol{\Sigma}_{0t})$, where conditional means $\boldsymbol{\mu}_{lt}^0$ are

$$\boldsymbol{\mu}_{lt}^0 = E(\mathbf{Z}_0 | \mathbf{T} = \mathbf{t}; \Omega_l) = \boldsymbol{\mu}_l^0 + (\mathbf{t} - \mathbf{M})' \boldsymbol{\alpha}_0, \quad l = 1, \ldots, L \tag{1}$$

and conditional covariance matrix $\boldsymbol{\Sigma}_{0t}$ is

$$\boldsymbol{\Sigma}_{0t} = V(\mathbf{Z}_0 | \mathbf{T} = \mathbf{t}; \Omega_l) = \rho \boldsymbol{\Sigma}. \tag{2}$$

Under the assumption of completely parametric certainty of populations and for known prior probabilities of populations π_l, $\sum_{l=1}^{L} \pi_l = 1$ Bayes rule minimizing the probability of misclassification is based on the logarithm of the conditional densities ratio.

Denote by $\boldsymbol{\Psi} = \{\mathbf{B}, \boldsymbol{\Sigma}\}$ the set of unknown parameters and denote the log ratio of conditional densities in categories Ω_k and Ω_l by

$$W_{kl}(\mathbf{Z}_0, \boldsymbol{\Psi}) = (\mathbf{Z}_0 - (\boldsymbol{\mu}_{kt} + \boldsymbol{\mu}_{lt})/2)' \boldsymbol{\Sigma}_t^{-1} (\boldsymbol{\mu}_{kt} - \boldsymbol{\mu}_{lt}) + \gamma_{kl} , \tag{3}$$

where $\gamma_{kl} = \ln(\pi_k / \pi_l)$.

These functions (3) will be called pairwise discriminant functions (PDF).

Then Bayes rule (see Anderson [6]) is defined as: classify $\mathbf{Z}_0$ to population Ω_k if for $l = 1,...,L$ $W_{kl}(\mathbf{Z}_0, \boldsymbol{\Psi}) \geq 0$.

Recall, that under the definition (see e.g. Dučinskas [2]) a conditional probability of misclassification due to aforementioned BCR is

$$P(\boldsymbol{\Psi}) = 1 - \sum_{k=1}^{L} \pi_k PC_k , \tag{4}$$

where for $k = 1,...,L$

$$PC_k = P_{0t}\left(W_{kl}(\mathbf{Z}_0; \boldsymbol{\Psi}) \geq 0, l = 1,\ldots,L, l \neq k \big| \Omega_l \right) \tag{5}$$

is the class conditional probability of correct classification of $\mathbf{Z}_0$ when it comes from Ω_k.

Let $\varphi_p(\cdot \,|\, \mathbf{M}, \mathbf{V})$ be the probability density function of the normally distributed random vector $\mathbf{U} \sim N_p(\mathbf{M}, \mathbf{V})$ and $\varphi_p(\cdot) = \varphi(\cdot \,|\, \mathbf{0}, \mathbf{I}_p)$.

Define two dimensional structure cases: CASE A for p<L and CASE B otherwise.

Set $\mathbf{a}_{kl} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_l)/\rho$.

Lemma 1. The class conditional probability of correct classification due to BCR specified in (5) in the CASE A is specified by

$PC_k = \int_{B_k} \varphi_p(u) du$, where $B_k = \left\{ \mathbf{u} \in R^p : \mathbf{u}' \mathbf{a}_{kl} + |\mathbf{a}_{kl}|^2/2 + \gamma_{kl} \geq 0, l = 1,...,L, l \neq k \right\}$ and in the CASE B

$PC_k = \int_{R_{L-1}^+} \varphi_{L-1}(v \,|\, \mathbf{M}_k, \mathbf{V}_k) dv$, where $\mathbf{M}_k$ is the L-1 vector with elements

$$\left\{ (\boldsymbol{\mu}_k^0 + \boldsymbol{\mu}_l^0)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_k^0 + \boldsymbol{\mu}_l^0)/\rho + \gamma_{kl}, l = 1,...,L, l \neq k \right\}$$

and $\mathbf{V}_k$ is the $(L-1) \times (L-1)$ matrix with elements

$$\left\{ (\boldsymbol{\mu}_k^0 + \boldsymbol{\mu}_l^0)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_k^0 + \boldsymbol{\mu}_m^0)/\rho, l,m = 1,...,L, l \neq m \right\} .$$

Proof. CASE A. Probability measure P_{0t} is based on conditional distribution of $\mathbf{Z}_0$ given $\mathbf{T} = t, \Omega_k$ with means and variance-covariance matrix specified in (1), (2). In the above conditions, $\mathbf{Z}_0$ may be expressed in form $\mathbf{Z}_0 = \boldsymbol{\Sigma}_t^{1/2} \mathbf{U} + \boldsymbol{\mu}_{kt}$, where $\mathbf{U} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ and $\mathbf{I}_p$ denotes the p dimensional identity matrix.

After making the change of variables $\mathbf{u} \to \mathbf{Z}_0$ in (3)-(5), we complete the proof of CASE A.

CASE B. Set $\mathbf{W}_1(\mathbf{Z}_0, \boldsymbol{\Psi}) = (W_{12}(\mathbf{Z}_0, \boldsymbol{\Psi}),...,W_{1L}(\mathbf{Z}_0, \boldsymbol{\Psi}))'$,

$\mathbf{W}_k(\mathbf{Z}_0, \boldsymbol{\Psi}) = (W_{k1}(\mathbf{Z}_0, \boldsymbol{\Psi}),...,W_{kk-1}(\mathbf{Z}_0, \boldsymbol{\Psi}), W_{kk+1}(\mathbf{Z}_0, \boldsymbol{\Psi}),...,W_{kL}(\mathbf{Z}_0, \boldsymbol{\Psi}))', k = 2,...,L$.

Then $\mathbf{W}_k(\mathbf{Z}_0, \boldsymbol{\Psi}) \sim N_{L-1}(\mathbf{M}_k, \boldsymbol{\Sigma}_k)$ and proof of Lemma 1 is straightforward.

3. The error rates for plug-in BDF

In practical applications not all statistical parameters of populations are known. Then the estimators of unknown parameters can be found from training sample. When estimators of unknown parameters are plugged into BDF, the plug-in BDF is obtained. In this paper we assume that true values of parameters $\mathbf{B}$ and $\boldsymbol{\Sigma}$ are unknown. Let $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\Sigma}}$ be the estimators of $\mathbf{B}$ and $\boldsymbol{\Sigma}$ based on $\mathbf{T}$. Set $\hat{\boldsymbol{\Psi}} = \{\hat{\mathbf{B}}, \hat{\boldsymbol{\Sigma}}\}$.

Then replacing $\mathbf{\Psi}$ by $\hat{\mathbf{\Psi}}$ in (3) we get the plug-in BDF

$$\hat{W}_{kl} = W_{kl}\left(\mathbf{Z}_0, \hat{\mathbf{\Psi}}\right) = \left(\mathbf{Z}_0 - (\hat{\mathbf{\mu}}_{kt} + \hat{\mathbf{\mu}}_{lt})/2\right)' \hat{\mathbf{\Sigma}}_t^{-1}\left(\hat{\mathbf{\mu}}_{kt} - \hat{\mathbf{\mu}}_{lt}\right) + \gamma_{kl} ,$$

where $\hat{\mathbf{\mu}}_{kl} = \hat{\mathbf{B}}_k \mathbf{x}(s_0) + \left(\mathbf{t} - \mathbf{X}\hat{\mathbf{B}}\right)' \mathbf{\alpha}_0$.

Then the classification rule based on PBCR is associated with plug-in PDF in the following way: classify $\mathbf{Z}_0$ to population Ω_k if for $l = 1,...,L$ $W_{kl}\left(\mathbf{Z}_0, \hat{\mathbf{\Psi}}\right) \geq 0$.

Definition 1. The overall actual error rate incurred by PBCR associated with PPDF is

$$P_t\left(\hat{\mathbf{\Psi}}\right) = 1 - \sum_{k=1}^{L} \pi_k P\hat{C}_k ,$$

where, for $k = 1,...,L$ $P\hat{C}_k = P_{0t}\left(W_{kl}\left(\mathbf{Z}_0, \hat{\mathbf{\Psi}}\right) \geq 0, l = 1,...,L, l \neq k \big| \Omega_k\right)$ to be called class-conditional actual error rate.

Set $\hat{\mathbf{a}}_{kl} = \mathbf{\Sigma}^{1/2} \hat{\mathbf{\Sigma}}^{-1}(\hat{\mathbf{\mu}}_k - \hat{\mathbf{\mu}}_l)/\sqrt{\rho}$, $\hat{\mathbf{\mu}}_k = \hat{\mathbf{B}}_k \mathbf{x}(s_0)$, $\hat{\mathbf{b}}_{kl} = \left(\mathbf{\mu}_k + \mathbf{\alpha}_0' \mathbf{X}(\hat{\mathbf{B}} - \mathbf{B}) - (\hat{\mathbf{\mu}}_k + \hat{\mathbf{\mu}}_l)/2\right)\hat{\mathbf{\Sigma}}^{-1}(\hat{\mathbf{\mu}}_k + \hat{\mathbf{\mu}}_l)/\rho + \gamma_{kl}$.

Lemma 2. The defined conditional actual error rate has the following form for CASE A:

$$P\hat{C}_k = \int_{A_k} \varphi_p(u) du ,$$

where $A_k = \left\{\mathbf{u} \in R^p : \mathbf{u}'\hat{\mathbf{a}}_{kl} + \hat{b}_{kl} \geq 0, l = 1,...,L, l \neq k\right\}$. And for CASE B the conditional actual error rate is

$$P\hat{C}_k = \int_{R_{L-1}^+} \varphi_{L-1}\left(v \big| \hat{\mathbf{M}}_k, \hat{\mathbf{V}}_k\right) dv ,$$

where $R_{L-1}^+ = \left\{v \in R^{L-1}, v \geq 0\right\}$, $\hat{\mathbf{M}}_k$ is the (L-1) vector with elements $\left\{\hat{b}_{kl}, l = 1,...,L, l \neq k\right\}$ and $\mathbf{V}_k$ is the $(L-1) \times (L-1)$ matrix with elements $\left\{\hat{\mathbf{a}}_{kl}' \hat{\mathbf{a}}_{km}, l, m = 1,...,L, l, m \neq k\right\}$.

Proof. Lemma 2 could be proved analogically to the Lemma 1.

4. Example and discussions

To investigate the performance of the proposed plug-in Bayes rule a simulation study for the three-category case with feature modelled by bivariate stationary GRF on regular lattice with exponential covariance function given by $C(h) = \sigma^2 \exp\{-h/\theta\}$ is carried out. Here σ^2 is variance and θ is a parameter of spatial correlation. Dependence of the derived error rates on category parameters is studied.

The results of numerical analysis give us strong arguments to expect that the proposed formulas of error rates could be effectively used for the performance evaluation of the plug-in Bayes rules applied to multi-category classification of spatial Gaussian process observation for different dimensional structure cases.

References

[1] Šaltytė-Benth J, Dučinskas K. Linear discriminant analysis of multivariate spatial-temporal regressions. *Scandinavian Journal of Statistics*. 2005; 32, 281-294.
[2] Dučinskas K. Approximation of the expected error rate in classification of the Gaussian random field observations. *Statistics and Probability Letters*. 2009; 79, p. 138-144.
[3] Batsidis A and Zografos K. Errors of misclassification in discrimination of dimensional coherent elliptic random field observations. *Statistica Neerlandica*. 2011; 65, p 446-461.
[4] Dučinskas K, Dreižienė L, Zikarienė E. Multiclass classification of the scalar Gaussian random field observation with known spatial correlation function. *Statistics and Probability Letters*. 2015; 98, p. 107-114.
[5] Dreižienė L, Dučinskas K, Paulionienė L. Correct classification rates in multi-category discriminant analysis of spatial Gaussian data. *Open Journal of Statistics*. 2015; 5:1, p. 21-26.
[6] Anderson TW. An introduction to multivariate statistical analysis. Wiley, New York. 2003.