

Correct Classification Rates in Multi-Category Discriminant Analysis of Spatial Gaussian Data

Lina Dreiziėnė^{1,2}, Kęstutis Dućinskas¹, Laura Paulionienė¹

¹Department of Mathematics and Statistics, Klaipėda University, Klaipėda, Lithuania

²Institute of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania

Email: l.dreiziene@gmail.com, kestutis.ducinskas@ku.lt, saltyte.laura@gmail.com

Received 28 December 2014; accepted 17 January 2015; published 22 January 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper discusses the problem of classifying a multivariate Gaussian random field observation into one of the several categories specified by different parametric mean models. Investigation is conducted on the classifier based on plug-in Bayes classification rule (PBCR) formed by replacing unknown parameters in Bayes classification rule (BCR) with category parameters estimators. This is the extension of the previous one from the two category cases to the multi-category case. The novel closed-form expressions for the Bayes classification probability and actual correct classification rate associated with PBCR are derived. These correct classification rates are suggested as performance measures for the classifications procedure. An empirical study has been carried out to analyze the dependence of derived classification rates on category parameters.

Keywords

Gaussian Random Field, Bayes Classification Rule, Pairwise Discriminant Function, Actual Correct Classification Rate

1. Introduction

Much work has been done concerning the error rates in two-category discrimination of uncorrelated observations (see e.g. [1]). Several methods for estimations of the error rates in discriminant analysis of spatial data have been recently proposed (see e.g. [2] [3]).

The multi-category problem, however, has very rarely been addressed because most of the methods proposed for two categories do not generalize. Schervish [4] considered the problem of classification into one of three

How to cite this paper: Dreiziėnė, L., Dućinskas, K. and Paulionienė, L. (2015) Correct Classification Rates in Multi-Category Discriminant Analysis of Spatial Gaussian Data. *Open Journal of Statistics*, 5, 21-26.

<http://dx.doi.org/10.4236/ojs.2015.51003>

known normal populations by single linear discriminant function. Techniques for multi-category probability estimation by combining all pairwise comparisons are investigated by several authors (see e.g. [5]). Empirical comparison of different methods of error rate estimation in multi-category linear discriminant analysis for multivariate homoscedastic Gaussian data was performed by Hirst [6]. Bayesian multiclass classification problem for correlated Gaussian observation was empirically studied by Williams [7]. The novel model-free estimation method for multiclass conditional probability based on conditional quintile regression functions is theoretically and numerically studied by Xu [8]. Correct classification rates in multi-category classification of independent multivariate Gaussian observations were provided by Schervish [9]. We generalize results of above to the problem of classification of multivariate spatially correlated Gaussian observations.

We propose the method of multi-category discriminant analysis essentially exploiting the Bayes classification rule that is optimal in the sense of minimum misclassification probability in case of complete statistical certainty (see [10], chapter 6). In practice, however, the complete statistical description of populations is usually not possible. Then having training sample, parametric plug-in Bayes classification rule formed by replacing unknown parameters with their estimators in BCR is being used.

Šaltytė and Dučinskas [11] 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. This result was generalized to multivariate spatial-temporal regression model in [12]. However, the observations to be classified are assumed to be independent from training samples in all publication listed above. The assumption of independence for the classification of scalar GRF observations was removed by Dučinskas [2]. Multivariate two-category case has been considered in Dučinskas [13] and Dučinskas and Dreiženė [14]. Formulas for the error rates for multiclass classification of scalar GRF observation are derived in [15]. The authors of the above papers have been focused on the maximum likelihood (ML) estimators because of tractability of the covariance matrix of these estimators. In the present paper, we extend the investigation of the performance of the PBCR in multi-category case. The novel closed form expressions for the actual correct classification rate (ACCR) are derived.

By using the derived formulas, the performance of the PBR is numerically analyzed in the case of stationary Gaussian random field on the square lattice with the exponential covariance function. The dependence of the correct classification rate and ACCR values on the range parameter is investigated.

The rest of the paper is organized as follows. Section 2 presents concepts and notions concerning BCR applied to multi-category classification of multivariate Gaussian random field (MGRF) observation. Bayes probability of correct classification is derived. In Section 3, the actual correct classification rate incurred by PBCR is considered and its closed-form expression is derived. Numerical examples, based on simulated data, are presented in Section 4, in order to illustrate theoretical results. The effect of the values of range parameter on the values of ACCR is examined.

2. The Main Concepts and Definitions

The main objective of this paper is to classify a single observation of MGRF $\{Z(s) : s \in D \subset \mathbb{R}^2\}$ into one of L categories, say $\Omega_1, \dots, \Omega_L$.

The model of observation $Z(s)$ in category Ω_l ($l=1, \dots, L$) is

$$Z(s) = \mu_l(s; B_l) + \varepsilon(s).$$

Here μ_l represents a mean component and B_l is a matrix of parameters. The error term is generated by p -dimensional zero-mean stationary GRF $\{\varepsilon(s) : s \in D\}$ with covariance function defined by model for all $s, u \in D$

$$\text{cov}\{\varepsilon(s), \varepsilon(u)\} = r(s-u)\Sigma,$$

where $r(s-u)$ is the spatial correlation function and Σ is the variance-covariance matrix with elements $\{\sigma_{ij}\}$. So we have deal with so called intrinsic covariance model (see [16]).

Consider the problem of classification of the vector of observation of Z at location s_0 denoted by $Z_0 = Z(s_0)$ into one of L populations specified above with given joint training sample T . Joint training

sample T is stratified training sample, specified by $n \times p$ matrix $T = (T_1', \dots, T_L')'$, where T_l is the $n_l \times p$ matrix of n_l observations of $Z(\bullet)$ from Ω_l , $l = 1, \dots, L$, $n = \sum_{l=1}^L n_l$.

Then the model of T is

$$T = M(B) + E,$$

where $B' = (B_1', \dots, B_L')$ is the matrix of category means parameters and E is the $n \times p$ matrix of random errors that has matrix-variate normal distribution *i.e.*

$$E \sim N_{n \times p}(0, R \otimes \Sigma).$$

Here R denotes the spatial correlation matrix among components (rows) of T . In the rest of the paper the realization (observed value) of training sample T will be denoted by t .

Denote by r_0 the vector of spatial correlations between Z_0 and observations in T and set $\alpha_0 = R^{-1}r_0$, $\rho = 1 - r_0' \alpha_0$, $\mu_l^0 = \mu_l(s_0)$, $l = 1, \dots, L$.

Notice that in category Ω_l , the conditional distribution of Z_0 given $T = t$ is Gaussian, *i.e.*

$$(Z_0 | T = t, \Omega_l) \sim N_p(\mu_{lt}^0, \Sigma_{0l}),$$

where conditional means μ_{lt}^0 are

$$\mu_{lt}^0 = E(Z_0 | T = t; \Omega_l) = \mu_l^0 + (t - M(B))' \alpha_0, \quad l = 1, \dots, L \quad (1)$$

and conditional covariance matrix Σ_{0l} is

$$\Sigma_{0l} = V(Z_0 | T = t; \Omega_l) = \rho \Sigma. \quad (2)$$

The marginal and conditional squared Mahalanobis distances between categories Ω_k and Ω_l ($k, l = 1, \dots, L$) for observation taken at location $s = s_0$ are specified respectively by

$$\Delta_{kl}^2 = (\mu_k^0 - \mu_l^0)' \Sigma^{-1} (\mu_k^0 - \mu_l^0),$$

and

$$d_{kl}^2 = (\mu_{kt}^0 - \mu_{lt}^0)' \Sigma_{0l}^{-1} (\mu_{kt}^0 - \mu_{lt}^0) = \Delta_{kl}^2 / \rho.$$

It is easy to notice that d_{kl} does not depend on realizations of T and depends only on their locations.

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.

There is no loss of generality in focusing attention on category L , since the numbering of the categories is arbitrary. Let the set of population parameters is denoted by $\Psi = \{B, \Sigma\}$. Set $r = L - 1$.

Denote the log ratio of conditional densities in categories Ω_L and Ω_l by

$$W_{Ll}(Z_0, \Psi) = (Z_0 - (\mu_{Ll} + \mu_{ll})/2)' \Sigma_l^{-1} (\mu_{Ll} - \mu_{ll}) + \gamma_{Ll}, \quad (3)$$

where $\gamma_{kl} = \ln(\pi_k / \pi_l)$, $k, l = 1, \dots, L$.

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

Then Bayes rule (BR) (see [10], chapter 6) is given by:

$$\text{classify } Z_0 \text{ to population } \Omega_L \text{ if for } l = 1, \dots, r, W_{Ll}(Z_0, \Psi) \geq 0. \quad (4)$$

3. Probabilities and Rates of Correct Classification

Set $a_{kl} = \Sigma^{-1/2} (\mu_k - \mu_l) / \rho$ and set M as r -dimensional vector with the l -th components ($l = 1, \dots, r$) spe-

cified as $m_l = |a_{Ll}|^2 / 2 + \gamma_{Ll}$, and $V = (v_{lm}, l, m = 1, \dots, r)$ with $v_{lm} = a'_{Ll} a_{Lm}$.

Lemma 1. The conditional probability of correct classification for category L due to BCR specified in (4) is

$$PC_L(\Psi) = \int_{R_r^+} \varphi_r(w; M, V) dw.$$

Here $\varphi_r(\cdot)$ is the probability density function of r -variate normal distribution with mean vector M and variance-covariance matrix V .

Proof. Recall, that under the definition (see e.g. [4] [9]) a probability of correct classification due to aforementioned BCR is

$$PC_k = P_{0t} \left(W_{kl}(Z_0; \Psi) \geq 0, k, l = 1, \dots, L, l \neq k \mid \Omega_l \right). \quad (5)$$

It is the probability of correct classification of Z_0 when it comes from Ω_l . Probability measure P_{0t} is based on conditional distribution of Z_0 given $T = t$, Ω_k with means and variance-covariance matrix specified in (1), (2). Z_0 may be expressed in form

$$Z_0 = \Sigma_t^{1/2} U + \mu_{kt},$$

where $U \sim N_p(0, I_p)$, and I_p denotes the p dimensional identity matrix.

After making the substitution of variables I_p in (5) we obtain that

$$E(W_{Ll}(Z_0)) = m_l \text{ and } \text{Cov}(W_{Ll}(Z_0), W_{Lm}(Z_0)) = v_{lm}, \quad l, m = 1, \dots, r.$$

Set $W(Z_0, \Psi) = (W_{L1}(Z_0, \Psi), \dots, W_{L,L-1}(Z_0, \Psi))$, then probability of correct classification can be rewritten in the following way $PC_L(\Psi) = P(W(Z_0, \Psi) > 0)$.

After straightforward calculations we show that $W(Z_0, \Psi) \sim N_{L-1}(M, \Sigma)$. That completes the proof of lemma.

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 Bayes discriminant function (BDF), the plug-in BDF is obtained (PBDF). In this paper we assume that true values of parameters B and Σ are unknown.

Let \hat{B} and $\hat{\Sigma}$ be the estimators of B and Σ based on T . Set $\hat{\Psi} = \{\hat{B}, \hat{\Sigma}\}$.

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

$$W_{Ll}(Z_0; \hat{\Psi}) = \frac{1}{\rho} \left(Z_0 - (T - M(\hat{B}))' \alpha_0 - (\hat{\mu}_L + \hat{\mu}_l)/2 \right)' \hat{\Sigma}^{-1} (\hat{\mu}_L - \hat{\mu}_l) + \gamma_{Ll}.$$

Then the classification rule based on PBCR is associated with plug-in PDF (PPDF) in the following way: classify Z_0 to population Ω_k if for $l = 1, \dots, L$ $W_{kl}(Z_0; \hat{\Psi}) \geq 0$.

Definition 1. The actual correct classification rate incurred by PBCR associated with PPDF is

$$PC_k^{\hat{}} = P_{0t} \left(W_{kl}(Z_0; \hat{\Psi}) \geq 0, l = 1, \dots, L, l \neq k \mid \Omega_k \right).$$

Set $\hat{a}_{Ll} = \Sigma^{1/2} \hat{\Sigma}^{-1} (\hat{\mu}_L - \hat{\mu}_l) / \sqrt{\rho}$ and $\hat{b}_{Ll} = \left(\mu_L + \alpha'_0 (M(\hat{B}) - M(B)) - (\hat{\mu}_L + \hat{\mu}_l)/2 \right)' \hat{\Sigma}^{-1} (\hat{\mu}_L - \hat{\mu}_l) / \rho + \gamma_{Ll}$.

Lemma 2. The actual correct classification rate due to PBDR is

$$PC_L(\hat{\Psi}) = \int_{R_r^+} \varphi_r(w; \hat{M}, \hat{V}) dw,$$

where \hat{M} is r -dimensional vector with components $\hat{m}_l = \hat{b}_{Ll}$, $l = 1, \dots, r$ and $\hat{V} = (\hat{v}_{lm} = \hat{a}'_{Ll} \hat{a}_{Lm}, l, m = 1, \dots, r)$.

Proof. It is obvious that in population Ω_l the conditional distribution of BPDF $W_{L_l}(Z_0; \hat{\Psi})$ given $T = t$ is Gaussian, *i.e.*,

$$W_{L_l}(Z_0; \hat{\Psi}) \Big| T = t, \Omega_{L_l} \sim N(\hat{m}_l, \hat{v}_{ll}).$$

Set $W(Z_0, \hat{\Psi}) = (W_{L_1}(Z_0, \hat{\Psi}), \dots, W_{L_{L-1}}(Z_0, \hat{\Psi}))$, then probability of correct classification can be rewritten in the following way:

$$PC_L(\hat{\Psi}) = P(W(Z_0, \hat{\Psi}) > 0).$$

After straightforward calculations we show that $W(Z_0, \hat{\Psi}) \sim N_{L-1}(\hat{M}, \hat{V})$. That completes the proof of lemma.

4. Example and Discussions

Simulation study in order to compare proposed Bayes probability of correct classification rate and the actual correct classification rate incurred by PBCR was carried out for three class case ($L = 3$). Also the effect of the range parameter on these values is examined.

In this example, observations are assumed to arise from bivariate stationary Gaussian random field ($p = 2$) with constant mean and isotropic exponential correlation function given by $r(h) = \exp\{-h/\theta\}$, where θ is a parameter of spatial correlation (range).

Set $\mu_1 = -\mu_2 = 1_2$, $\mu_3 = 0_2$ and $\Sigma = I_2$.

Estimators of B and Σ have the following form:

$$\hat{B} = \hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3)' = \hat{\mu}_{ML} = (X'R^{-1}X)^{-1} X'R^{-1}T,$$

where X denotes design matrix of training sample T and is specified by $X = 1_4 \oplus 1_4 \oplus 1_4$ and

$$\hat{\Sigma} = (T - M(\hat{B}))' R^{-1} (T - M(\hat{B})) / (n - 3).$$

Considered set of training locations with indicated class labels is shown in **Figure 1**.

So we have small training sample sizes (*i.e.* $n_1 = n_2 = n_3 = 4$) and three different locations to be classified, furthermore we assume equal prior probabilities $\pi_l = 1/3$, $l = 1, 2, 3$.

Simulations were performed by geoR: a free and open-source package for geostatistical analysis included in statistical computing software R (<http://www.r-project.org/>). Each case was simulated 100 times (runs) and ACCR values are calculated by averaging ACCR over the runs. ACCR and CCR values are presented in **Table 1**. As might be expected ACCR values are lower than CCR. All values are increasing while range parameter is increasing. That means the stronger correlation gives better accuracy of proposed classification procedures.

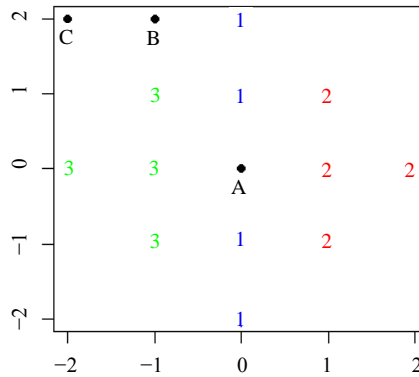


Figure 1. Locations of training sample: “1” are samples from population Ω_1 , “2” from Ω_2 , “3” from Ω_3 , A, B and C denotes the locations of observation to be classified.

Table 1. CCR and $\overline{\text{ACCR}}$ values.

θ	CCR			$\overline{\text{ACCR}}$		
	A (0,0)	B (-1,2)	C (-2,2)	A (0,0)	B (-1,2)	C (-2,2)
1	0.633777	0.634955	0.487689	0.544697	0.427956	0.404184
2	0.882716	0.839326	0.599707	0.636842	0.501045	0.483608
3	0.973217	0.948266	0.716064	0.725356	0.565855	0.480018
4	0.999777	0.955340	0.810298	0.773966	0.639186	0.553055

It's seen in **Figure 1**, that the closest location to be classified is location A and the farthest is location C. CCR and $\overline{\text{ACCR}}$ are largest for location A and smallest for location C. It can be concluded that better accuracy gives closer locations.

References

- [1] McLachlan, G.J. (2004) *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, New York.
- [2] Dučinskas, K. (2009) Approximation of the Expected Error Rate in Classification of the Gaussian Random Field Observations. *Statistics and Probability Letters*, **79**, 138-144. <http://dx.doi.org/10.1016/j.spl.2008.07.042>
- [3] Batsidis, A. and Zografos, K. (2011) Errors of Misclassification in Discrimination of Dimensional Coherent Elliptic Random Field Observations. *Statistica Neerlandica*, **65**, 446-461. <http://dx.doi.org/10.1111/j.1467-9574.2011.00494.x>
- [4] Schervish, M.J. (1984) Linear Discrimination for Three Known Normal Populations. *Journal of Statistical Planning and Inference*, **10**, 167-175. [http://dx.doi.org/10.1016/0378-3758\(84\)90068-5](http://dx.doi.org/10.1016/0378-3758(84)90068-5)
- [5] Wu, T.F., Lin, C.J. and Weng, R.C. (2004) Probability Estimates for Multi-Class Classification by Pairwise Coupling. *Journal of Machine Learning Research*, **5**, 975-1005.
- [6] Hirst, D. (1996) Error-Rate Estimation in Multiply-Group Linear Discriminant Analysis. *Technometrics*, **38**, 389-399. <http://dx.doi.org/10.1080/00401706.1996.10484551>
- [7] Williams, C.K.I. and Barber, D. (1998) Bayesian Classification with Gaussian Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, 1342-1351. <http://dx.doi.org/10.1109/34.735807>
- [8] Xu, T. and Wang, J. (2013) An Efficient Model-Free Estimation of Multiclass Conditional Probability. *Journal of Statistical Planning and Inference*, **143**, 2079-2088. <http://dx.doi.org/10.1016/j.jspi.2013.08.008>
- [9] Schervish, M.J. (1981) Asymptotic Expansions for the Means and Variances of Error Rates. *Biometrika*, **68**, 295-299. <http://dx.doi.org/10.1093/biomet/68.1.295>
- [10] Anderson, T.W. (2003) *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
- [11] Šaltytė, J. and Dučinskas, K. (2002) Comparison of ML and OLS Estimators in Discriminant Analysis of Spatially Correlated Observations. *Informatica*, **13**, 297-238.
- [12] Šaltytė-Benth, J. and Dučinskas, K. (2005) Linear Discriminant Analysis of Multivariate Spatial-Temporal Regressions. *Scandinavian Journal of Statistics*, **32**, 281-294. <http://dx.doi.org/10.1111/j.1467-9469.2005.00421.x>
- [13] Dučinskas, K. (2011) Error Rates in Classification of Multivariate Gaussian Random Field Observation. *Lithuanian Mathematical Journal*, **51**, 477-485. <http://dx.doi.org/10.1007/s10986-011-9142-4>
- [14] Dučinskas, K. and Dreižienė, L. (2011) Supervised Classification of the Scalar Gaussian Random Field Observations under a Deterministic Spatial Sampling Design. *Austrian Journal of Statistics*, **40**, 25-36.
- [15] Dučinskas, K., Dreižienė, L. and Zikariėnė, E. (2015) Multiclass Classification of the Scalar Gaussian Random Field Observation with Known Spatial Correlation Function. *Statistics and Probability Letters*, **98**, 107-114. <http://dx.doi.org/10.1016/j.spl.2014.12.008>
- [16] Wackernagel, H. (2003) *Multivariate Geostatistics: An Introduction with Applications*. Springer-Verlag, Berlin. <http://dx.doi.org/10.1007/978-3-662-05294-5>