MISCLASSIFICATION PROBABILITY BASED ON LINEAR DISCRIMINANT FUNCTION FOR SAR ERROR MODELS

K. Dučinskas, M. Karaliutė, I. Šimkienė
Department of Statistics
Klaipeda University, Lithuania
e-mail: k.ducinskas@gmail.com

Abstract

Given training sample, the problem of classifying Gaussian spatial data into one of two populations specified by spatial autoregressive model (SAR) with different mean functions is considered. This paper concerns classification procedures associated with Linear Discriminant Function (LDF) under deterministic spatial sampling design. In the case of complete parametric certainty, the overall misclassification probability associated with LDF is derived. Spatial weights based on inverse of Euclidean distance and the third order neighbourhood schemes on regular 2-dimensional lattice are used for illustrative examples. The effect of the prior distribution of class labels on the performance of proposed classification procedure is numerically evaluated.

Keywords: discriminant function, Covariance function, Gaussian random field, misclassification probability, Training labels configuration.

1 Introduction

Many authors (see e.g. Kharin (1996)) have investigated the performance of the linear discriminant function(LDF) in classification of dependent observations (Markov dependence, autoregressive models). Error rates in classification of spatially correlated Gaussian observations associated with LDF is considered in Ducinskas (2009). Here only the parametric models for spatial covariance belonging to Mattern class are considered. In this paper the extension of the latter investigation to the case of spatial Gaussian data specified by widely used SAR model is presented.

In this work an explicit expression for overall misclassification probability(OMP) is derived. By using the derived OMP, the performance of the BDF is numerically analyzed in the case of stationary Gaussian random field on 2-dimentional regular lattice. The dependence of the values of obtained OMP on the Mahalanobis distance for different spatial sampling designs and prior distributions for class labels is investigated. By applying the proposed criterion, the numerical comparison of some training labels configurations (TLC) is carried out.

2 The main concepts and definitions

The main objective of this paper is to classify the Gaussian random field (GRF) observation $Z(s_0)$ in population Ω_j $(j = 1, 2)$, when means satisfy the condition

$$\mu_1(s_0) \neq \mu_2(s_0), s_0 \in D.$$

Denote by $S_n = \{s_i \in D; i = 1, ..., n\}$ the set of locations where training sample $T' = (Z(s_1), ..., Z(s_n))$ is taken, and call it the set of training locations (STL). It specifies the spatial sampling design for training sample. Assume that 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 $j = 1, 2$. So each partition $\xi(S_n) = \{S^{(1)}, S^{(2)}\}$ with marked labels determines TLC.

For TLC $\xi(S_n)$, define the variable $d = |D^{(1)} - D^{(2)}|$, where $D^{(j)}$ is the sum of distances between the location s_0 and locations in $S^{(j)}$, $j = 1, 2$.

As it follows, we assume that STL S_n and TLC ξ are fixed. Denote by w_{ij} a spatial weight specifying the interconnection between locations s_i and s_j ($w_{ii} = 0$ and $w_{ij} \neq 0$ if $i \approx j$, and 0 otherwise) for $i, j = 0, ..., n$). Here $i \approx j$ denotes that location s_j is a neighbour of location s_i, and let W denote the n by n spatial weights matrix for S_n.

So the training sample T follows SAR error model

$$T = M + E,$$

where M is the vector of the training sample mean and E is the $n \times 1$ - vector of random errors with multivariate Gaussian distribution $N_n(0, V)$, with $V = \sigma^2[(I - \lambda W)'(I - \lambda W)]^{-1}$, and σ^2, λ are respectively the variance and autoregressive parameters. Let t denote the realization of T.

Assume that Z_0 follows SAR error model. Then the conditional distribution of Z_0 given $T = t$, Ω_j is Gaussian with mean $\mu^0_{lt} = E(Z_0|T = t; \Omega_j)$ and variance $\sigma^2_t = var(Z_0|T = t; \Omega_j)$.

Under the assumption of complete parametric certainty of populations the LDF minimizing the OMP is formed by log ratio of conditional likelihoods.

Then LDF is specified by (Anderson, 2003)

$$W_t(Z_0) = (Z_0 - \frac{1}{2}(\mu^0_{1t} + \mu^0_{2t}))(\mu^0_{1t} - \mu^0_{2t})/\sigma^2_t + \gamma,$$

where $\gamma = ln(\pi_1/\pi_2)$.

Here π_1, π_2 ($\pi_1 + \pi_2 = 1$) are prior probabilities of the populations Ω_1 and Ω_2 for observation at location s_0.They specified the prior distribution for class membership for observation at location s_0.

Definition 1 *The OMP for the LDF $W_t(Z_0, \Psi)$ is defined as*

$$PB = \sum_{i=1}^{2} \sum_{j=1, j\neq i}^{2} \pi_i P_{ij},$$

where for $i, j = 1, 2$, $P_{ij} = P_{it}((-1)^j W_t(Z_0) < 0)$.

Here, for $i, j = 1, 2$, the probability measure P_{it} is based on conditional distribution of Z_0 given $T = t, \Omega_i$. Squared Mahalanobis distance between conditional distributions of Z_0 given $T = t$ are denote by and $\Delta^2_0 = (\mu^0_{1t} - \mu^0_{2t})^2/\sigma^2_t$.

Set $\Delta = |\mu_1(s_0) - \mu_2(s_0)|$ and denote by w_0 the n vector of spatial weights between s_0 and S_n, i.e. $w'_0 = (w_{01}, w_{02}, ..., w_{0n})$.

The OMP is useful in providing a guide to the performance of classification rule before it is actually formed from training sample. The OMP is the performance measure to the BDF similar as the mean squared prediction error (MSPE) is the performance measure to the kriging predictor (see Diggle et al 2002).

Make the following assumptions:

(A1) The set of locations S_n^0 form a clique of size $n+1$,

(A2) Spatial weights for S_n and S_n^0 are based on the Euclidean distance between different locations.

Lemma 1 *Suppose that observation Z_0 to be classified by BDF and let covariance matrix of T_0. Then under the assumptions (A1) and (A2), OMP takes the form*

$$PB = \sum_{j=1}^{2} (\pi_j \Phi(-\Delta_0/2 + (-1)^j \gamma/\Delta_0)),$$

where $\Phi(\cdot)$ is the standard normal distribution function and $\Delta_0^2 = \Delta^2(1+\lambda^2 w_0' w_0)/\sigma^2$.

Proof 1 *The proof of lemma is easily done by using the properties of normal distribution.*

3 Example and discussions

Numerical example is considered to investigate the influence of the statistical parameters of populations to the proposed LDF in the finite (even small) training sample case. With an insignificant loss of generality the cases with $n_1 = n_2$ are considered. We also suppose that assumptions (A1) and (A2) hold.

In this example, observations are assumed to arise from stationary Gaussian random field with constant mean. The spatial weights are specified by $w_{ij} = 1/d_{ij}$, where d_{ij} is the Euclidean distance between different locations.

Assume D is regular 2-dimentional lattice with unit spacing. We consider NN(3) spatial structure scheme for $s_0 = (2,2)$. STL for this scheme consists of 12 third-order neighbours of s_0 and is denoted by S_{12}. Set $M1 = \{i : i = 1, ..., n_1, i \approx 0\}$.

Two cases of prior probabilities are considered:

$$CN) \quad \pi_1 = \Big(\sum_{i \in M1} 1 \Big)/n, \quad CD) \quad \pi_1 = \Big(\sum_{i \in M1} 1/d_{0i} \Big) / \Big(\sum_{i=1,...,n} 1/d_{0i} \Big).$$

Case CN is based only on the number of neighbours, while case CD incorporated spatial adjacency (distances) also. So OMP is denoted PBN for the case CN and PBD for the case CD.

Consider two TLC ξ_1, ξ_2 for S_{12} specified by $\xi_1 = \{S^{(1)} = \{(1,3),(2,4),(2,3),(2,1),(3,2),(4,2)\}, \{S^{(2)} = \{(0,2),(1,2),(0,2),(3,3),(3,1),(1,1)\}\}$, $\xi_2 = \{S^{(1)} = \{(0,2),(1,2),(2,3),(2,1),(3,2),(4,2)\}, \{S^{(2)} = \{(2,4),(1,3),(1,1),(2,0),(3,3),(3,1)\}\}$.

The comparison of two cases of prior distribution for each TLC is done by the values of index $\eta = PBD/PBN$. The values PBD, PBN and index η for various values of

Table 1: Values of PBD, PBN and η for NN(3) neighbourhood schemee.

TLC	ξ_1			ξ_2		
Δ	PBD	PBN	η	PBD	PBN	η
0,1	0,4662	0,4847	0,9618	0,4336	0,4847	0,8947
0,2	0,4586	0,4694	0,9771	0,4324	0,4694	0,9211
0,3	0,4468	0,4542	0,9838	0,4267	0,4542	0,9396
0,4	0,4334	0,4390	0,9873	0,4176	0,4390	0,9513
0,5	0,4194	0,4239	0,9895	0,4065	0,4239	0,9591
0,6	0,4052	0,4089	0,9910	0,3944	0,4089	0,9645
0,7	0,3909	0,3941	0,9920	0,3817	0,3941	0,9685
0,8	0,3767	0,3794	0,9928	0,3686	0,3794	0,9716
0,9	0,3625	0,3649	0,9935	0,3553	0,3649	0,9740
1,0	0,3484	0,3505	0,9939	0,3421	0,3505	0,9759

parameter Δ are given in Table 1. The results of calculations with $\lambda = 0,3$, $\sigma^2 = 1$ for ξ_1 and ξ_2 are presented in Table 1. By the definition variable d represents the asymmetry population labels distribution in training sample. It is easy to obtain that $d = 2(\sqrt{2}-1)$ for ξ_1 and $d = 4(\sqrt{2}-1)$ for ξ_2. So ξ_1 is less asymmetric TLC than ξ_2.

Analyzing the contents of table 1. We can conclude that prior distribution based on distances to neighbours outperforms the one based only on numbers of neighbours, because PBD values are smaller than corresponding PBN values. Table also shows that for both neighbourhood schemes OMP decreases with the increasing of Δ. Values of index η in tables numerically illustrate the comparison of two cases of prior distributions. These values enable us to conclude that effect of the incorporation distances into prior distributions is stronger for more asymmetric TLC i.e. for ξ_2. Hence the results of numerical analysis give us strong arguments to expect that proposed derived formula of the OMP could be effectively used for performance evaluation of classification procedures and for the optimal designing of spatial training samples.

References

[1] Anderson T.W. (2003).*An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

[2] Diggle P.J., Ribeiro P.J., Christensen O.F. (2002). An introduction to model-based geostatistics. *Lecture notes in statistics*. Vol. **173**, pp. 43-86.

[3] Ducinskas K. (2009). Approximation of the expected error rate in classification of the Gaussian random field observations. *Statistics and Probability Letters*. Vol. **79**, pp. 138-144.

[4] Kharin Yu. (1996). *Robustness in Statistical Pattern Recognition*. Dordrecht: Kluwer Academic Publishers.