AUSTRIAN JOURNAL OF STATISTICS
Volume 40 (2011), Number 1 & 2, 25–36

Supervised Classification of the Scalar Gaussian Random Field Observations under a Deterministic Spatial Sampling Design

Kestutis Ducinskas and Lina Dreiziene
Klaipeda University, Lithuania

Abstract: Given training sample, the problem of classifying a scalar Gaussian random field observation into one of two populations specified by different parametric mean models and common parametric covariance function is considered. Such problems are usually called as supervised classification or contextual classification problems. This paper concerns with classification procedures associated with Bayes Discriminant Function (BDF) under deterministic spatial sampling design. In the case of parametric uncertainty, the ML estimators of unknown parameters are plugged in the BDF. The actual risk and the approximation of the expected risk (AER) associated with aforementioned plug-in BDF are derived. This is the extension of the previous one to the case of complete parametric uncertainty, i.e., when all mean functions and covariance function parameters are unknown. Stationary geometrically anisotropic Gaussian random field with exponential covariance function sampled on regular 2-dimensional lattice is used for illustrative examples.

Keywords: Bayesian Discriminant Function, Covariance Function, Gaussian Random Field, Actual Risk, Training Labels Configuration.

1 Introduction

It is well known that in case of completely specified populations and known loss function, an optimal classification rule in the sense of minimum risk is based on BDF (Anderson, 2003). In practice, however, the complete statistical description of populations is usually not possible. Training sample is used for obtaining the estimators of statistical parameters which are then usually plugged in BDF. Actual risk and expected risk (ER) are considered as performance measures for plug-in version of BDF (PBDF). However, the expressions for the ER are very cumbersome even for the simplest forms of PBDF. This makes it difficult to build some qualitative conclusions. Therefore, asymptotic approximations of the expected error rate 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. Lawoko and McLachlan, 1985). The influence of dependence in data (stationary time series, Markov dependence, autoregressive dependence) on the performance of the PBDF is also considered by Kharin (1996).

Switzer (1980) was the first to treat classification of spatial data. Plug-in approach to discrimination for feature observations having elliptically contoured distributions is implemented in Batsidis and Zografos (2006). Saltyte and Ducinskas (2002) 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 Saltyte-Benth and Ducinskas (2005). However, in all publication listed above, the observations to be classified are assumed independent from training samples. This is unrealistic assumption particularly when the locations of observations to be classified are close to ones of training sample.

The first extensions of above approximation to the case when spatial correlations between Gaussian observations to be classified and observations in training sample are not assumed equal zero is done in Ducinskas (2009). Here only the trend parameters and scale parameter of covariance function is assumed unknown. The extension of the latter approximation to the case of complete parametric uncertainty (all means and covariance function parameters are unknown) is implemented in the present paper. We focus on the maximum likelihood (ML) estimators, since the inverse of the information matrix associated with likelihood function of training sample well approximates the covariance matrix of these estimators. The asymptotic properties of ML estimators showed by Mardia and Marshall (1984) under increasing domain asymptotic framework and subject to some regularity conditions are essentially exploited. The asymptotic results sometimes yield useful approximations to finite-sample properties. The simulated annealing algorithm can be used in searching the optimal spatial sampling design.

By using the proposed AER, the performance of the PBDF is numerically analyzed in the case of stationary Gaussian random field on 2-dimentional regular lattice with exponential covariance function. The dependence of the values of obtained AER on the statistical parameters such as the range parameter, the anisotropy ratio and Mahalanobis distance, is investigated.

By applying the proposed criterion, the numerical comparison of two training labels configurations (TLC) is carried. That gives us the strong arguments for suggestion to include derived formulas of error rates in the geospatial data mining (Shekhar, Schrater, Vatsavai, Wu, and Chawla, 2002). The proposed DF could also be considered as the extension of widely used Bayesian methods to the restoration of image corrupted by spatial Gaussian noise (Cressie, 1993, Ch. 7.4).

2 Main Concepts and Definitions

The main objective of this paper is to classify the observations of Gaussian random field (GRF)

$$\{Z(s) : s \in D \subset \mathbb{R}^p\} \ .$$

The model of 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 β_j is a $q \times 1$ vector of parameters, $j = 1, 2$. The error term is generated by a zero mean stationary GRF $\{\varepsilon(s) : s \in D\}$ with covariance function defined by model for all $s, u \in D$

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

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

In the case when covariance function parameters are known, model (1), (2) is called an universal kriging model (Cressie, 1993, Ch. 3).

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

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

Denote by $S_n = \{s_i \in D; i = 1, \ldots, n\}$ the set of locations where training sample $T' = (Z(s_1), \ldots, Z(s_n))$ is taken, and call it the set of training locations (STL). It specifies the spatial sampling design or spatial framework for training sample (see Shekhar et al., 2002.

We shall assume the deterministic spatial sampling design and all analyses are carried out conditional on S_n.

Assume that each training sample realization $T = t$ and S_n are arranged in the following way. The first n_1 components are observations of $Z(s)$ from Ω_1 and remaining $n_2 = n - n_1$ components are the observations of $Z(s)$ from Ω_2. So 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 $S^{(j)}$, $j = 1, 2$.

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

$$X = X_1 \oplus X_2 \,,$$

where the 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, $j = 1, 2$.

As it follows, we assume that STL S_n and TLC ξ are fixed. This is the case, when spatial classified training data are collected at fixed locations (stations).

So the model of 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 $N_n(0, \Sigma(\theta))$.

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

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

and variance

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

where

$$x_0' = x'(s_0) \,, \qquad \alpha_0' = c_0'\Sigma^{-1} \,.$$

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 conditional likelihoods.

Then BDF is specified by McLachlan (2004)

$$W_t(Z_0, \Psi) = \left(Z_0 - \frac{1}{2}(\mu^0_{1t} + \mu^0_{2t})\right)\left(\mu^0_{1t} - \mu^0_{2t}\right)/\sigma^2_0 + \gamma\,, \tag{6}$$

where $\gamma = \log(\pi^*_1/\pi^*_2)$ and $\Psi = (\beta', \theta')'$.

Here $\pi^*_j = \pi_j(L(j, 3-j) - L(j,j))$, $j = 1,2$, where π_1, π_2 ($\pi_1 + \pi_2 = 1$) are prior probabilities of the populations Ω_1 and Ω_2, respectively.

Note that in (6) the prior probabilities π_1, π_2 can be sometimes replaced by estimates $\hat{\pi}_j = n_j/n$, $j = 1,2$.

So BDF allocates the observation in the following way: Classify observation Z_0 given $T = t$ to population Ω_1 if $W_t(Z_0, \Psi) \geq 0$ and to population Ω_2, otherwise.

Definition 1. The risk for the BDF $W_t(Z_0, \Psi)$ is defined as

$$P(\Psi) = \sum_{i=1}^{2}\sum_{j=1}^{2} \pi_i L(i,j) P_{ij}\,,$$

where, for $i,j = 1,2$,

$$P_{ij} = P_{it}((-1)^j W_t(Z_0, \Psi) < 0)\,.$$

Here, for $i = 1,2$, the probability measure P_{it} is based on the conditional distribution of Z_0 given $T = t$, Ω_i specified in (4), (5). As it follows, the risk $P(\Psi)$ will be called Bayes risk.

Note that under the condition (3), the squared Mahalanobis distance between marginal distributions of Z_0 and the squared Mahalanobis distance between conditional distributions of Z_0 given $T = t$ are specified by $\Delta^2 = (\mu^0_1 - \mu^0_2)^2/C(0)$ and $\Delta^2_0 = (\mu^0_{1t} - \mu^0_{2t})^2/\sigma^2_0$, respectively.

From (4), (5) it is easy to derive that

$$\Delta^2_0 = \Delta^2 C(0)/\sigma^2_0\,.$$

Thus, Δ_0 does not depend on realizations of T.

In population Ω_j, the conditional distribution of $W_t(Z_0, \Psi)$ given $T = t$ is normal with mean

$$\mathrm{E}_j(W_t(Z_0, \Psi)) = (-1)^{j+1}\Delta^2_0/2 + \gamma$$

and variance

$$\mathrm{var}_j(W_t(Z_0, \Psi)) = \Delta^2_0\,, \qquad j = 1,2\,.$$

By using the properties of normal distribution we obtain

$$P(\Psi) = \sum_{j=1}^{2}\left(\pi^*_j \Phi\left(-\Delta_0/2 + (-1)^j\gamma/\Delta_0 + \pi_j L(j,j)\right)\right)\,, \tag{7}$$

where $\Phi(\cdot)$ is the standard normal distribution function.

In practical applications not all statistical parameters of the populations are known. Then the estimators of the unknown parameters can be found from a training sample. When the estimators of the unknown parameters are plugged into the BDF, we obtain the plug-in BDF (PBDF). In this paper we assume that the true values of the parameters β and θ are unknown (complete parametric uncertainty).

Let $\hat{\beta}, \hat{\theta}$ be the estimators of the corresponding parameters from the training sample T and let $\tilde{\beta}, \tilde{\theta}$ be the realizations of these estimators based on $T = t$. Those realizations are usually called the estimates of the parameters.

We shall write hat above functions of parameter θ, if it is replaced by the estimator $\hat{\theta}$ and shall write tilde above functions of parameter θ, if it is replaced by the estimate $\tilde{\theta}$.

Then by using (4), (5) we get the estimate of the conditional mean

$$\tilde{\mu}_{jt}^0 = x_0'\tilde{\beta}_j + \tilde{\alpha}_0'(t - X\beta)\,, \qquad j = 1, 2$$

and the estimate of conditional variance

$$\tilde{\sigma}_0^2(\theta) = \tilde{C}(0) - \tilde{c}_0'\Sigma^{-1}\tilde{c}_0\,.$$

Put $\tilde{\Psi} = (\tilde{\beta}', \tilde{\theta}')'$. Then replacing the parameters by their estimates in (6) we form the PBDF as

$$W_t(Z_0, \tilde{\Psi}) = \left(Z_0 - \tilde{\alpha}_0'(t - X\tilde{\beta}) - \tfrac{1}{2}x_0'H\tilde{\beta}\right)\left(x_0'G\tilde{\beta}\right)\Big/\tilde{\sigma}_0^2 + \gamma\,, \tag{8}$$

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

Definition 2. The actual risk for BPDF $W_t(Z_0, \tilde{\Psi})$ is defined as

$$P(\tilde{\Psi}) = \sum_{i=1}^{2}\sum_{j=1}^{2} \pi_i L(i,j)\tilde{P}_{ij}\,, \tag{9}$$

where for $i, j = 1, 2$,

$$\tilde{P}_{ij} = P_{it}\left((-1)^j W_t(Z_0, \tilde{\Psi}) < 0\right)\,. \tag{10}$$

Lemma 1. The actual risk specified in (9), (10) for $W_t(Z_0, \tilde{\Psi})$ in (8) is

$$P(\tilde{\Psi}) = \sum_{j=1}^{2}\left(\pi_j^*\Phi(\tilde{Q}_j) + \pi_j L(j,j)\right)\,, \tag{11}$$

and

$$\tilde{Q}_j = (-1)^j\left((a_j - \tilde{b})\mathrm{sgn}(x_0'G\tilde{\beta})/\sigma_0 + \tilde{\sigma}_0^2\gamma/(\sigma_0|x_0'G\tilde{\beta}|)\right)\,, \tag{12}$$

where for $j = 1, 2$

$$a_j = x_0'\beta_j + \alpha_0'(t - X\beta)\,, \qquad \tilde{b} = \tilde{\alpha}_0'(t - X\tilde{\beta} + x_0'H\tilde{\beta}/2)\,. \tag{13}$$

Proof. In population Ω_j, the conditional distribution of $W_t(Z_0, \tilde{\Psi})$ given $T = t$ is normal with mean

$$\mathrm{E}_j(W_t(Z_0, \tilde{\Psi})) = (a_j - \tilde{b})x_0'G\tilde{\beta}/\tilde{\sigma}_0^2 + \gamma$$

and variance

$$\text{var}_j(W_t(Z_0, \tilde{\Psi})) = (x_0' G \tilde{\beta})^2 \sigma_0^2 / \tilde{\sigma}_0^4 \,, \qquad j = 1, 2 \,.$$

Then by using the properties of normal distribution and definition 2 we complete the proof of lemma 1.

Definition 3. The expectation of the actual risk with respect to the distribution of T is called the expected risk (ER) and is designated as $E_T(P(\hat{\Psi}))$.

More comprehensive information about the actual and expected risks for the classification into an arbitrary number of populations you can find in Ducinskas (1997).

The ER is useful in providing a guide to the performance of the plug-in classification rule before it is actually formed from the training sample. The ER is the performance measure to the PBDF similar as the mean squared prediction error (MSPE) is the performance measure to the plug-in kriging predictor (see Diggle, Ribeiro, and Christensen, 2002. The approximations of MSPE for plug-in kriging were suggested in several previous papers (Zimmerman and Cressie, 1992; Abt, 1999). These approximations are used for spatial sampling design criterion for prediction (see Zimmerman, 2006; Zhu and Stein, 2006). These facts strengthen the motivation for the deriving of the AER associated with PBDF.

3 Asymptotic Expansion of the Expected Risk

We will use the maximum likelihood estimators (MLEs) of the parameters based on the training sample. The asymptotic properties of the MLEs established by Mardia and Marshall (1984) under increasing domain asymptotic framework and subject to some regularity conditions are essentially exploited. Hence, the MLE $\hat{\Psi}$ is weakly consistent and asymptotically Gaussian, i.e.

$$\hat{\Psi} \sim AN(\Psi, J^{-1}) \,. \tag{14}$$

Here the expected information matrix is given by

$$J = J_\beta \oplus J_\theta \,, \tag{15}$$

where

$$J_\beta = X' \Sigma^{-1} X \tag{16}$$

and the (i, j)th element of J_θ is

$$\text{tr}(\Sigma^{-1} \Sigma_i \Sigma^{-1} \Sigma_j)/2 \tag{17}$$

Henceforth, denote by MM conditions the regularity conditions of Theorem 1 from Mardia and Marshall (1984).

Using properties of the multivariate Gaussian distribution it is easy to prove that

$$\hat{\beta} \sim AN_{2q}(\beta, J_\beta^{-1}) \,, \tag{18}$$

and

$$\hat{\theta} \sim AN_k(\theta, J_\theta^{-1}) \,. \tag{19}$$

Let $P_\beta^{(k)}$, $P_\theta^{(k)}$, $k=1,2$ denote the kth order derivatives of $P(\hat{\Psi})$ with respect to $\hat{\beta}$ and $\hat{\theta}$ evaluated at point $\hat{\beta}=\beta, \hat{\theta}=\theta$ and let $P^2_{\beta\theta}$ denote the matrix of second derivatives of $P(\hat{\Psi})$ with respect to $\hat{\beta}$ and $\hat{\theta}$ evaluated at $\hat{\beta}=\beta, \hat{\theta}=\theta$.

Make the following assumption: (A1) The training sample T and the estimator $\hat{\theta}$ are statistically independent.

The restrictive assumption (A1) is exploited intensively by many authors (see Zimmerman, 2006; Zhu and Zhang, 2006), since Abt (1999) showed that finer approximations of MSPE considering the correlation between T and $\hat{\theta}$ do not give better results.

Let $A=\partial\hat{\alpha}_0/\partial\hat{\theta}'$ be the $n\times k$ matrix of partial derivatives evaluated at $\hat{\theta}=\theta$ and let $\varphi(\cdot)$ be the standard normal density function.

Theorem 1. Suppose that observation Z_0 is classified by BPDF and let conditions (MM) and assumption (A1) hold. Then the approximation of ER is

$$\mathrm{AER}=P(\Psi)+\pi_1^*\varphi(-\Delta_0/2-\gamma/\Delta_0)\Delta_0(K_\beta+K_\theta)/(2\sigma_0^2)\,, \tag{20}$$

where

$$K_\beta=\Lambda' J_\beta^{-1}\Lambda \tag{21}$$

$$K_\theta=\mathrm{tr}(\Sigma A J_\theta^{-1}A')+\gamma^2((\hat{\sigma}_0^2)_\theta^{(1)})'J_\theta^{-1}(\hat{\sigma}_0^2)_\theta^{(1)}/(\Delta_0^2\sigma_0^2) \tag{22}$$

$$\Lambda'=\alpha_0'X-x_0'(H/2+\gamma G/\Delta_0^2)\,. \tag{23}$$

Proof. Expanding $P(\hat{\Psi})$ in a Taylor series around $\hat{\beta}=\beta$ and $\hat{\theta}=\theta$, we have

$$\begin{aligned}P(\hat{\Psi}) &= P(\Psi)+(P_\beta^{(1)})'\Delta\hat{\beta}+P_\theta^{(1)}\Delta\hat{\theta}\\ &+\tfrac{1}{2}\left((\Delta\hat{\beta})'P_\beta^{(2)}\Delta\hat{\beta}+2(\Delta\hat{\beta})'P_{\beta\theta}^{(2)}\Delta\hat{\theta}+(\Delta\hat{\theta})'P_\theta^{(2)}(\Delta\theta)\right)+R_3\,,\end{aligned} \tag{24}$$

where $\Delta\hat{\beta}=\hat{\beta}-\beta$, $\Delta\hat{\theta}=\hat{\theta}-\theta$ and R_3 is the remainder term.

By using (11), the partial derivatives of $P(\hat{\Psi})$ evaluated at $\hat{\beta}=\beta$ and $\hat{\theta}=\theta$ are

$$P_\beta^{(1)}=\pi_1\varphi(Q_1)\sum_{j=1}^{2}Q_{j\beta}^{(1)}\,,\qquad P_\theta^{(1)}=\pi_1\varphi(Q_1)\sum_{l=j}^{2}Q_{j\theta}^{(1)}\,, \tag{25}$$

$$P_\beta^{(2)}=\pi_1\varphi(Q_1)\sum_{j=1}^{2}\left(Q_{j\beta}^{(2)}-Q_lQ_{j\beta}^{(1)}Q_{j\beta}^{(1)\prime}\right) \tag{26}$$

$$P_\theta^{(2)}=\pi_1\varphi(Q_1)\sum_{j=1}^{2}\left(Q_{j\theta}^{(2)}-Q_jQ_{j\theta}^{(1)}Q_{j\theta}^{(1)\prime}\right)\,. \tag{27}$$

After doing some algebra in (12), (13), we have

$$\begin{aligned}Q_{j\beta}^{(1)} &= (-1)^j\Lambda/\sigma_0\,,\\ Q_{j\theta}^{(1)} &= (-1)^j(-A'(T-X\beta)+\gamma(\sigma_0^2)_\theta^{(1)}/(\Delta_0\sigma_0))/\sigma_0\end{aligned}$$

and

$$\sum_{j=1}^{2}Q_{j\beta}^{(2)}=0\,,\qquad \sum_{j=1}^{2}Q_{j\theta}^{(2)}=0\,.$$

Application of the above formulae to (25)-(27) yields

$$P_{\beta}^{(1)} = P_{\theta}^{(1)} = 0\,, \qquad P_{\beta}^{(2)} = \pi_1\Delta_0\varphi(-\Delta_0/2 - \gamma/\Delta_0)\Lambda\Lambda'/\sigma_0^2 \tag{28}$$

and

$$P_{\theta}^{(2)} = \pi_1\Delta_0\varphi(Q_1)\left(-A'(T - X\beta) + \gamma(\sigma_0^2)_{\theta}^{(1)}/(\Delta_0\sigma_0)\right) \left(-A'(T - X\beta) + \gamma(\sigma_0^2)_{\theta}^{(1)}/(\Delta_0\sigma_0)\right)'/\sigma_0^2\,. \tag{29}$$

It is easy to show that all elements of the matrix $P_{\beta\theta}^{(2)}$ are finite.

Then by using assumption (A1) and replacing $\mathrm{E}_T(\Delta\hat{\theta}\Delta\hat{\theta}')$ by its asymptotic approximation J_{θ}^{-1}, we get the following approximation

$$\mathrm{E}\left((\Delta\hat{\theta})'P_{\theta}^{(2)}(\Delta\hat{\theta})\right) \cong \pi_1^*\Delta_0\varphi(-\Delta_0/2 - \gamma\Delta_0)\left(\mathrm{tr}(\Sigma A J_{\theta}^{-1}A') + \gamma^2\left((\hat{\sigma}_0^2)_{\theta}^{(1)}\right)' J_{\theta}^{-1}(\hat{\sigma}_0^2)_{\theta}^{(1)}/(\Delta_0^2\sigma_0^2)\right)/\sigma_0^2\,. \tag{30}$$

Then taking the expectation term by term on the righthand side of (24), using (7), (28) – (30) and replacing the moments of the estimators by the corresponding moments of the asymptotic distributions specified in (14) – (19), we complete the proof of Theorem 1.

Remark. If we consider a nuggetless covariance model and $\theta = \sigma^2$, then the approximation specified in (20) – (23) coincides with one derived in Ducinskas (2009).

This remark is proved by using $A = 0$ and $\sigma_0^2 \propto \sigma^2$ in (22).

4 Example and Discussion

A numerical example is considered to investigate the influence of the statistical parameters in the populations on the proposed AER in the finite (even small) training sample case. With an insignificant loss of generality the cases with $n_1 = n_2$, $\pi_1 = \pi_2 = 1/2$ and $L(i,j) = 1 - \delta_{ij}$, $i,j = 1,2$, are considered.

In this example, the observations are assumed to stem from a stationary Gaussian random field with constant mean and nuggetless covariance function given by $C(h) = \sigma^2 r(h)$, where σ^2 is the unknown variance (sill) and $r(h)$ is the spatial correlation function.

The exponential geometric anisotropic correlation function $r(h)$ with unknown anisotropy ratio λ and anisotropy angle $\varphi = \pi/2$ (see Diggle et al., 2002) specified by

$$r(h) = \exp\left\{-\sqrt{h_x^2 + \lambda^2 h_y^2}\Big/\alpha\right\}$$

is considered. Here α denotes the unknown range parameter. Hence, we have the case with $\theta = (\sigma^2, \lambda, \alpha)'$.

Assume that Δg is a regular two-dimensional lattice with unit spacing. Consider the case $s_0 = (1,1)$ and fixed STL S_8 contains 8 second-order neighbors of s_0.

Consider two TLC ξ_1, ξ_2 for the training sample specified by

$$\xi_1 = \left\{S^{(1)} = \{(1,2),(2,2),(2,1),(2,0)\}\,,\quad S^{(2)} = \{(1,0),(0,0),(0,1),(0,2)\}\right\}\,,$$

$$\xi_2 = \left\{S^{(1)} = \{(1,2),(2,1),(0,1),(1,0)\}\,,\quad S^{(2)} = \{(0,0),(0,2),(2,0),(2,2)\}\right\}\,.$$

They are presented in Figure 1.

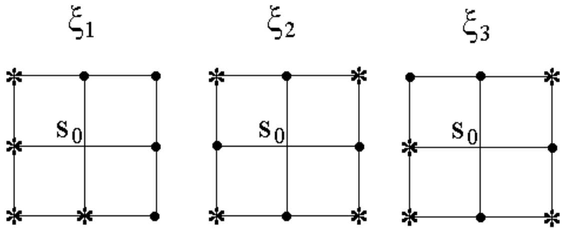

Figure 1: Two different TLC with $S^{(1)}$ and $S^{(2)}$ points and *, signed as respectively.

For both TLC the values of AER specified in (20) – (23) are calculated for various parameter values λ, α. The results of the calculations with $\Delta = 1$ for ξ_1 are presented in Table 1 and for ξ_2 in Table 2.

Table 1: Values of AER for TLC ξ_1 with $\Delta = 1$ and various α and λ.

	α							
λ	0.6	0.8	1.2	1.6	2.0	2.4	2.8	3.2
1	0.328438	0.307256	0.270947	0.240968	0.215754	0.194206	0.175548	0.159223
2	0.334138	0.313996	0.281444	0.255017	0.232404	0.212669	0.195248	0.179737
3	0.334396	0.313624	0.280805	0.254897	0.232919	0.213723	0.196722	0.181530
4	0.333937	0.313140	0.280474	0.254719	0.232910	0.213868	0.196990	0.181895
5	0.333237	0.312551	0.280198	0.254600	0.232881	0.213904	0.197076	0.182018
6	0.332543	0.312004	0.279944	0.254496	0.232850	0.213913	0.197112	0.182072
7	0.331927	0.311540	0.279721	0.254400	0.232816	0.213911	0.197128	0.182101
8	0.331398	0.311157	0.279533	0.254314	0.232781	0.213903	0.197135	0.182116
9	0.330946	0.310840	0.279378	0.254238	0.232747	0.213892	0.197136	0.182124
10	0.330559	0.310576	0.279249	0.254173	0.232716	0.213879	0.197134	0.182128

Analyzing the contents of Tables 1 and 2 we can conclude that for both TLC ξ_1 and ξ_2 the AER increases with increasing anisotropy ratio λ and decreases with the decreasing range parameter α.

Now we numerically illustrate the comparison of two TLC, based on the minimum of AER criterion.

By the definition variable d represents the asymmetry population labels distribution in training sample. It is obvious that $d = 0$ for ξ_1, and $d = 4(\sqrt{2} - 1)$ for ξ_2. So we call ξ_1 and ξ_2 symmetric TLC and asymmetric TLC, respectively.

Table 2: Values of AER for TLC ξ_2 with $\Delta = 1$ and various α and λ.

	α							
λ	0.6	0.8	1.2	1.6	2.0	2.4	2.8	3.2
1	0.330108	0.310382	0.276753	0.248792	0.224998	0.204422	0.186405	0.170478
2	0.335293	0.316316	0.286035	0.261460	0.240264	0.221588	0.204943	0.189991
3	0.335625	0.316118	0.285660	0.261593	0.241000	0.222838	0.206599	0.191959
4	0.335198	0.315747	0.285613	0.261787	0.241399	0.223403	0.207291	0.192748
5	0.334506	0.315202	0.285500	0.261927	0.241684	0.223781	0.207734	0.193236
6	0.333814	0.314669	0.285326	0.261978	0.241861	0.224032	0.208032	0.193564
7	0.333199	0.314210	0.285141	0.261971	0.241960	0.224194	0.208232	0.193789
8	0.332669	0.313828	0.284971	0.261935	0.242008	0.224295	0.208368	0.193946
9	0.332217	0.313511	0.284823	0.261886	0.242026	0.224357	0.208458	0.194056
10	0.331831	0.313247	0.284698	0.261836	0.242026	0.224393	0.208519	0.194134

Table 3: Values of index η for various λ, α and Δ.

	$\eta = \mathrm{AER}_1/\mathrm{AER}_2$								
	$\alpha = 0.6$			$\alpha = 1.0$			$\alpha = 3.0$		
λ	$\Delta = 0.5$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0.5$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0.5$	$\Delta = 1$	$\Delta = 2$
1	1.00217	1.00508	1.01322	1.00653	1.01576	1.04288	1.02511	1.06631	1.20093
2	1.00148	1.00346	1.00885	1.00492	1.01178	1.03171	1.02049	1.05339	1.16060
3	1.00158	1.00367	1.00937	1.00527	1.01262	1.03384	1.02068	1.05387	1.16256
4	1.00162	1.00378	1.00963	1.00558	1.01334	1.03571	1.02150	1.05602	1.16919
5	1.00164	1.00381	1.00971	1.00573	1.01370	1.03666	1.02223	1.05790	1.17486
6	1.00164	1.00382	1.00975	1.00580	1.01386	1.03710	1.02277	1.05930	1.17904
7	1.00165	1.00383	1.00978	1.00583	1.01393	1.03730	1.02315	1.06030	1.18201
8	1.00165	1.00384	1.00980	1.00584	1.01397	1.03741	1.02343	1.06101	1.18412
9	1.00165	1.00384	1.00982	1.00585	1.01399	1.03747	1.02362	1.06152	1.18561
10	1.00165	1.00385	1.00983	1.00585	1.01400	1.03751	1.02376	1.06188	1.18667

The comparison of two TLC is done by the values of index $\eta = \mathrm{AER}_2/\mathrm{AER}_1$, where AER_l is the approximation of ERR for ξ_l, $l = 1, 2$. The values of this index are given in Table 3.

Analyzing Table 3 we see that for all parametric structures $\eta \geq 1$. So we can conclude that the symmetric TLC ξ_1 is more optimal than the asymmetric TLC ξ_2 by the AER minimum criterion.

Hence the results of numerical analysis give us strong arguments to expect that the proposed approximation of the expected error rate could be effectively used for performance evaluation of the classification procedures and for the optimal designing of spatial training samples.

References

Abt, M. (1999). Estimating the prediction mean squared error in Gaussian stochastic processes with correlation structure. *Scandinavian Journal of Statistics*, *26*, 563-578.

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

Batsidis, A., and Zografos, K. (2006). Discrimination of observations into one of two elliptic populations based on monotone training samples. *Metrika*, *64*, 221-241.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. New York: Wiley.

Diggle, P. J., Ribeiro, P. J., and Christensen, O. F. (2002). *An Introduction to Model-Based Geostatistics* (Vol. 173).

Ducinskas, K. (1997). An asymptotic analysis of the regret risk in discriminant analysis under various training schemes. *Lithuanian Mathematical Journal*, *37*, 337-351.

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

Kharin, Y. (1996). *Robustness in Statistical Pattern Recognition*. Dordrecht: Kluwer Academic Publishers.

Lawoko, C. R. O., and McLachlan, G. L. (1985). Discrimination with autocorrelated observations. *Pattern Recognition*, *18*, 145-149.

Mardia, K. V., and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatiall regression. *Biometrika*, *71*, 135-146.

McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley.

Saltyte, J., and Ducinskas, K. (2002). Comparison of ML and OLS estimators in discriminant analysis of spatially correlated observations. *Informatica*, *13*, 297-238.

Saltyte-Benth, J., and Ducinskas, K. (2005). Linear discriminant analysis of multivariate spatial-temporal regressions. *Scandinavian Journal of Statistics*, *32*, 281-294.

Shekhar, S., Schrater, P. R., Vatsavai, R. R., Wu, W., and Chawla, S. (2002). Spatial contextual classification and prediction models for mining geospatial data. *IEEE Transactions on Multimedia*, *4*, 174-188.

Switzer, P. (1980). Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. *Journal of Mathematical Geology*, *12*, 367-376.

Zhu, Z., and Stein, M. L. (2006). Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological, and Environmental Statistics*, *11*, 24-44.

Zhu, Z., and Zhang, H. (2006). Spatial sampling design under infill asymptotic framework. *Environmetrics*, *17*, 323-337.

Zimmerman, D. L. (2006). Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, *17*, 635-652.

Zimmerman, D. L., and Cressie, N. (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics*, *44*, 27-43.

Authors' address:

Kestutis Ducinskas and Lina Dreiziene
Department of Statistics
Klaipeda University
H. Manto st. 84
Klaipëda LT 92294
Lithuania

E-mails: kestutis.ducinskas@ku.lt and l.dreiziene@gmail.com