Nonlinear Analysis: Modelling and Control, 2001, v. 6, No. 2, 15-28

Quadratic Discriminant Analysis of Spatially Correlated Data

K. Dučinskas, J. Šaltytė
Klaipėda University
H.Manto 84, 5808 Klaipėda, Lithuania
[duce, jsaltyte]@gmf.ku.lt

Received: 13.09.2001
Accepted: 27.09.2001

Abstract

The problem of classification of the realisation of the stationary univariate Gaussian random field into one of two populations with different means and different factorised covariance matrices is considered. In such a case optimal classification rule in the sense of minimum probability of misclassification is associated with non-linear (quadratic) discriminant function. Unknown means and the covariance matrices of the feature vector components are estimated from spatially correlated training samples using the maximum likelihood approach and assuming spatial correlations to be known. Explicit formula of Bayes error rate and the first-order asymptotic expansion of the expected error rate associated with quadratic plug-in discriminant function are presented. A set of numerical calculations for the spherical spatial correlation function is performed and two different spatial sampling designs are compared.

Keywords: Bayesian classification rule, quadratic discriminant function, training samples, expected error rate, asymptotic expansion.

1 Introduction

In application areas like pattern recognition or geostatistics the data of interest are often spatially correlated. In numerous applications it is assumed, that data follow a Gaussian random field model. Therefore the discriminant analysis (DA) of spatially correlated Gaussian data is of great importance. When classes are completely specified, an optimal classification rule in the sense of minimum classification error (or error

rate) is the Bayesian classification rule. The expressions for the error rate are cumbersome for the DA based on linear discriminant function (LDF). Of course, they are even more delicate for the DA associated with quadratic disriminant function (QDF) [1]. Moreover, in practice the complete description of classes usually is not possible and the training samples are required for the estimation of probabilistic characteristics of each class. Therefore asymptotic expansions of the expected error rate are particularly important while evaluating the performance of certain discriminant functions and comparing different designs for training samples.

2 The problem and the model

Let $\{Z(\mathbf{s}): \mathbf{s} \in D \subset \Re^2\}$ be a univariate Gaussian random field having different stationary means and factorised covariance matrices in populations Ω_1 and Ω_2. Then the model of $Z(\mathbf{s})$ in population Ω_l is

$$Z(\mathbf{s}) = \mu_l + \varepsilon_l(\mathbf{s}),$$

where μ_l is a mean and $\{\varepsilon_l(\mathbf{s}): \mathbf{s} \in D \subset \Re^2\}$, is a zero-mean stationary Gaussian random field with covariance defined by a parametric model $\operatorname{cov}\{\varepsilon_l(\mathbf{t}), \varepsilon_l(\mathbf{s})\} = \sigma(\mathbf{h}; \boldsymbol{\theta}_l)$, where $\mathbf{h} = \mathbf{t} - \mathbf{s}$, $\mathbf{t}, \mathbf{s} \in D$, and $\boldsymbol{\theta}_l \in \boldsymbol{\Theta}$ is a $q \times 1$ parameter vector, $\boldsymbol{\Theta}$ being an open subset of $\Re^q$, l=1,2. The spatial covariance function in Ω_l is $\operatorname{cov}\{\varepsilon_l(\mathbf{t}), \varepsilon_l(\mathbf{s})\} = c(\mathbf{h}; \boldsymbol{\theta}_l)\sigma_l^2$, where $c(\mathbf{h}; \boldsymbol{\theta}_l)$ is the spatial correlation function and $\sigma_l^2 = \sigma(\mathbf{0}; \boldsymbol{\theta}_l)$, l=1,2. It is assumed that the function $c(\mathbf{h}; \boldsymbol{\theta}_l)$ is positive definite [2]. Assume that, for all $\mathbf{t}, \mathbf{s} \in D$, $\mathbf{t} \neq \mathbf{s}$,

$$\operatorname{cov}\{\varepsilon_1(\mathbf{t}), \varepsilon_2(\mathbf{s})\} = 0.$$

Consider the problem of supervised classification [3] of the observation $Z(\mathbf{r}) \in \boldsymbol{Z} \subset R$ with $\mathbf{r} \in D_0 \subset D$ into one of two populations specified above. In effect, classification rule divides the feature space $\boldsymbol{Z}$ into two mutually exclusive and exhaustive assignment regions U_1 and

U_2, where if $Z(\mathbf{r})$ falls in U_l, then the object is allocated to Ω_l, l=1,2. Under the assumption, that the populations are completely specified and for known prior probabilities of populations $\pi_1(\mathbf{r})$ and $\pi_2(\mathbf{r})$ ($\pi_1(\mathbf{r})+\pi_2(\mathbf{r})=1$), the Bayesian classification rule (BCR) $d_B(\cdot)$ minimising the probability of misclassification (PMC) is

$$d_B(\mathrm{z}(\mathbf{r}))=\arg\max_{\{l=1,2\}}\pi_l(\mathbf{r})p_l(\mathrm{z}(\mathbf{r})), \tag{1}$$

where $\pi_l(\mathbf{r})$ and $p_l(\mathrm{z}(\mathbf{r}))$ are a prior probability and probability density function of Ω_l, respectively, l=1,2. Then the corresponding QDF is

$$W(Z(\mathbf{r}),\phi)=\frac{1}{2}\ln\frac{\sigma_2^2}{\sigma_1^2}+\frac{1}{2}\sum_{l=1}^{2}(-1)^l\frac{(Z(\mathbf{r})-\mu_l)^2}{\sigma_l^2}+\gamma(\mathbf{r}), \tag{2}$$

where $\phi=(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)^T$ and $\gamma(\mathbf{r})=\ln\left(\frac{\pi_1(\mathbf{r})}{\pi_2(\mathbf{r})}\right)$.

The PMC of BCR, associated with QDF $W(Z(\mathbf{r}),\phi)$, usually called the Bayesian error rate, is

$$P_B^{\mathbf{r}}=\pi_1 P(W(Z(\mathbf{r}),\phi)\le 0/\Omega_1)+\pi_2 P(W(Z(\mathbf{r}),\phi)>0/\Omega_2), \tag{3}$$

where Ω_l under the sign of probability means that an object with observation $Z(\mathbf{r})$ belongs to the population Ω_l, l=1,2.

As it was already mentioned, in practical applications the parameters of density function are usually not known and must be estimated. Then the estimators of unknown parameters are found from the training samples $\mathbf{T}_1$ and $\mathbf{T}_2$ taken separately from Ω_1 and Ω_2, respectively. When estimators of unknown parameters are used, the plug-in version of BCR is obtained. The performance of the plug-in version of the BCR when parameters are estimated from training samples with independent observations is widely investigated by many authors [4]. However, it has been founded that the assumption of independence is frequently violated. Lawoko and McLachlan [5] for instance, investigated the performance of sample LDF when training samples follow a stationary autoregressive

process. In this paper we will consider the performance of the plug-in QDF when the parameters are estimated from training samples with spatially correlated observations. The maximum likelihood (ML) procedure for the estimation of unknown means and variances, assuming the spatial dependence parameter to be known, is used.

Suppose in region $D_1 \subset D$, $D_1 \cap D_0 = \varnothing$, we observe the stratified training sample $\mathbf{T} = \{\mathbf{T}_1, \mathbf{T}_2\}$ with $\mathbf{T}_l = \{Z_{l1},...,Z_{lN_l}\}$, where $Z_{l\alpha} = Z(\mathbf{s}_\alpha^l)$ denotes the α'*th* observation from Ω_l, l=1,2, $\alpha = 1,..., N_l$. Assume, that all points in D_0 are beyond the range of spatial correlation function ([6], ch.2) defined for points in D_1. Then $Z(\mathbf{r})$ is independent on $\mathbf{T}$.

Let $\hat{\mu}_l$ and $\hat{\sigma}_l^2$ be the ML estimators of μ_l and σ_l^2, l=1,2, respectively, based on $\mathbf{T}$. Then $\hat{\phi} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2)^T$ is the estimator of ϕ.

The plug-in rule $d_B(z(\mathbf{r}); \hat{\phi})$ is obtained by replacing the parameters in (1) with their estimators. Then the corresponding discriminant function $W(Z(\mathbf{r}), \hat{\phi})$, also known as the plug-in QDF, is

$$W(Z(\mathbf{r}), \hat{\phi}) = \frac{1}{2} \ln \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} + \frac{1}{2} \sum_{l=1}^{2} (-1)^l \frac{(Z(\mathbf{r}) - \hat{\mu}_l)^2}{\hat{\sigma}_l^2} + \gamma(\mathbf{r}).$$

Definition 1. The actual error rate for $d_B(z(\mathbf{r}); \hat{\phi})$ is defined as

$$P^{\mathbf{r}} = \pi_1 P(W(Z(\mathbf{r}), \hat{\phi}) \le 0 / \Omega_1) + \pi_2 P(W(Z(\mathbf{r}), \hat{\phi}) > 0 / \Omega_2).$$

Definition 2. The expectation of the actual error rate with respect to distribution of $\mathbf{T}$ designated as $E_T\{P^{\mathbf{r}}(\hat{\phi})\}$ is called the expected error rate (ER) for the $d_B(z(\mathbf{r}); \hat{\phi})$ and expected error regret (EER) is defined by $EER = E_T\{P^{\mathbf{r}}(\hat{\phi})\} - P_B^{\mathbf{r}}$.

The goal of this paper is to find asymptotic expansions of ER associated with plug-in QDF. K. Dučinskas [7] presented the asymptotic expansion for the case of arbitrary number of classes and regular class-conditional densities under the assumption, that observations in training samples are independent. K. V. Mardia [8] considered the problem of classifying the spatially distributed Gaussian observations with constant means and common covariance matrices (case of LDF), but he did not analyse the ER of PMC. In this paper we present the asymptotic expansion up to the order $O\left(N^{-1}\right)$, where $N = N_1 + N_2$, for the ER of classifying spatially distributed Gaussian observation with different means and different spatially factorised covariance matrices. Terms of higher order are omitted from the asymptotic expansion since their contribution usually is in generally negligible (see e.g. M. J. Schervish [9]). The ML estimators of means and the bias-adjusted ML estimators of the covariances are used in the plug-in version of the BCR. M. Taniguchi [10] has proved under sufficiently general assumptions, that ML estimators ensure minimum of EER up to $O\left(N^{-1}\right)$ for the considered classification rule. A set of calculations for a certain neighbourhood structures and spherical spatial correlation model is performed and values of P_B^{r} and approximation of EER are presented in tables.

3 Asymptotic expansion

The expectation vector and the covariance matrix of the vectorised training sample $\mathbf{T}_l$ defined by $\mathbf{T}_l^V = \left\{Z_{l1},...,Z_{lN_l}\right\}^T$ are

$$\boldsymbol{\mu}_l^V = \left(\mu_{l1},...,\mu_{lN_l}\right)^T \qquad \text{and} \qquad \boldsymbol{\Sigma}_l^V = \sigma_l^2 \mathbf{C}_l,$$

respectively, where $\mathbf{C}_l$ is known spatial correlation matrix of order $N_l \times N_l$, whose $\alpha\beta$ 'th element is $c\left(\mathbf{s}_\alpha^l - \mathbf{s}_\beta^l\right)$, $\alpha, \beta = 1,..., N_l$, l=1,2. Let

$$\mathbf{C}_l^{-1} = \left(c_l^{\alpha\beta}\right),\ c_l^{\bullet\bullet} = \sum_{\alpha,\beta=1}^{N_l} c_l^{\alpha\beta} \text{ and } c_l^{\alpha\bullet} = \sum_{\beta=1}^{N_l} c_l^{\alpha\beta},\ l=1,2.$$

Lemma. For l=1,2, the ML estimators of μ_l and σ_l^2, based on stratified training sample **T** are

$$\hat{\mu}_l = \frac{1}{c_l^{\bullet\bullet}} \sum_{\alpha=1}^{N_l} c_l^{\alpha\bullet} Z_{l\alpha} , \tag{4}$$

$$\hat{\sigma}_l^2 = \frac{1}{N_l} \sum_{\alpha,\beta=1}^{N_l} c_l^{\alpha\beta} (Z_{l\alpha} - \hat{\mu}_l)(Z_{l\beta} - \hat{\mu}_l). \tag{5}$$

Proof. The log-likelihood of $\mathbf{T}_l$, l=1,2, is

$$\ln L_l = const - \frac{1}{2}\left(N_l \ln \sigma_l^2 + \ln|\mathbf{C}_l|\right) - \frac{1}{2\sigma_l^2} \sum_{\alpha,\beta=1}^{N_l} c_l^{\alpha\beta} (Z_{l\alpha} - \hat{\mu}_l)(Z_{l\beta} - \hat{\mu}_l).$$

Solving the equations $\frac{\partial \ln L_l}{\partial \mu_l} = 0$ and $\frac{\partial \ln L_l}{\partial \sigma_l^2} = 0$, l=1,2, we complete the proof of lemma.

Since, for l=1,2, $E\{\hat{\sigma}_l^2\} = \frac{N_l - 1}{N_l}\sigma_l^2$, further we will use the bias-adjusted ML estimator of variance $\tilde{\sigma}_l^2 = \frac{N_l}{N_l - 1}\hat{\sigma}_l^2$.

It can be easily shown that $\hat{\mu}_l$ for finite N have known exact distribution $\hat{\mu}_l \sim N\left(\mu_l, \frac{1}{c_l^{\bullet\bullet}}\sigma_l^2\right)$, l=1,2.

For simplicity we omit the superscript "**r**" in $P^{\mathrm{r}}(\tilde{\phi})$, here $\tilde{\phi}$ is $\hat{\phi}$ with ML estimator of σ_l^2, l=1,2, replaced by the bias-adjusted ML estimator. Put $\Delta\hat{\mu}_l = \hat{\mu}_l - \mu_l$, $\Delta\tilde{\sigma}_l^2 = \tilde{\sigma}_l^2 - \sigma_l^2$, l=1,2. Denote by

$$P_l^{(1)} = \frac{\partial P(\tilde{\phi})}{\partial \hat{\mu}_l}, \qquad P_{k,l}^{(2)} = \frac{\partial^2 P(\tilde{\phi})}{\partial \hat{\mu}_k \partial \hat{\mu}_l}, \qquad P_{\tilde{\sigma}_l^2}^{(1)} = \frac{\partial P(\tilde{\phi})}{\partial \tilde{\sigma}_l^2}, \qquad P_{\tilde{\sigma}_{k,l}^2}^{(2)} = \frac{\partial^2 P(\tilde{\phi})}{\partial \tilde{\sigma}_k^2 \partial \tilde{\sigma}_l^2},$$

$P^{(2)}_{l,\tilde{\sigma}_l^2} = \dfrac{\partial^2 P(\tilde{\phi})}{\partial \hat{\mu}_l \partial \tilde{\sigma}_l^2}$ the partial derivatives of $P(\tilde{\phi})$ up to the second order with respect to the corresponding parameters evaluated at $\hat{\mu}_l = \mu_l$ and $\tilde{\sigma}_l^2 = \sigma_l^2$, k,l=1,2.

With insignificant loss of generality we consider the parametric structure case, when $\mu_1 > \mu_2$, $\sigma_1^2 > \sigma_2^2$ and

$$(\mu_1 - \mu_2)^2 - (\sigma_1^2 - \sigma_2^2)\left(\ln\left(\frac{\sigma_2^2}{\sigma_1^2} \right) + 2\gamma(\mathbf{r}) \right) > 0 .$$

Threshold points $z_1(\mathbf{r})$ and $z_2(\mathbf{r})$, i.e. solutions of $W(z(\mathbf{r}),\phi) = 0$, and assignment regions U_1 and U_2 for this parametric structure case is shown in Figure 1, assuming $\pi_1 = \pi_2 = 0.5$ for simplicity.

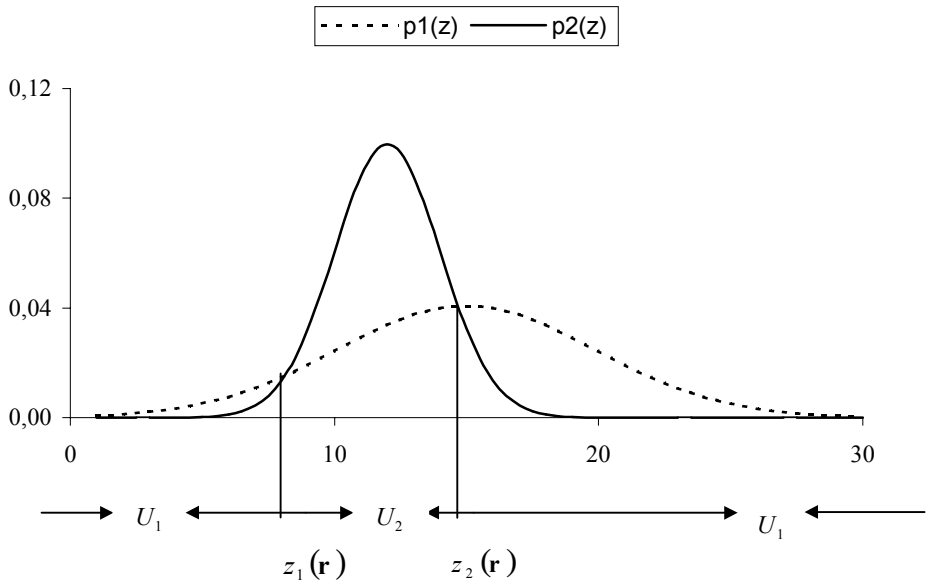


Fig. 1. Threshold points and assignment regions for the considered parametric structure case

On purpose to derive the asymptotic expansion of EER we need the following assumptions.

Assumption 1. Let $\lambda(\mathbf{C}_l)$ be the largest eigenvalue of $\mathbf{C}_l$, l=1,2. Suppose, that $\lambda(\mathbf{C}_l) < \kappa_l$, $0 < \kappa_l < \infty$, l=1,2.

Assumption 2. Assume, that $\frac{N_1}{N_2} \to \tau$, as $N_1, N_2 \to \infty$, $0 < \tau < \infty$.

Theorem. Suppose, that assumptions 1,2 hold for training samples $\mathbf{T}_1$, $\mathbf{T}_2$. Then the asymptotic expansion of the expected error rate for the $d_B\left(z(\mathbf{r}), \tilde{\phi}\right)$ is

$$E_T\left(P\left(\tilde{\phi}\right)\right) = P_B + \frac{1}{2}\sum_{l=1}^{2}\frac{\alpha_l}{c_l^{\bullet\bullet}} + \sum_{l=1}^{2}\frac{\beta_l}{N_l - 1} + O\left(N^{-2}\right), \quad (6)$$

where

$$P_B = \pi_1\Phi\left(\frac{z_2(\mathbf{r}) - \mu_1}{\sigma_1}\right) + \pi_2\Phi\left(\frac{z_1(\mathbf{r}) - \mu_2}{\sigma_2}\right) + \sum_{l=1}^{2}(-1)^l\pi_l\Phi\left((-1)^{l+1}\frac{z_l(\mathbf{r}) - \mu_l}{\sigma_l}\right) \quad (7)$$

$$\alpha_l = \pi_1\sum_{k=1}^{2}\frac{p_1(z_k(\mathbf{r}))\left(\frac{z_k(\mathbf{r}) - \mu_l}{2\sigma_l^2}\right)^2}{\left|\sum_{j=1}^{2}(-1)^j\frac{z_k(\mathbf{r}) - \mu_j}{2\sigma_j^2}\right|}, \quad (8)$$

$$\beta_l = \pi_1\sum_{k=1}^{2}\frac{p_1(z_k(\mathbf{r}))\left(\frac{(z_k(\mathbf{r}) - \mu_l)^2 - \sigma_l^2}{2\left(\sigma_l^2\right)^2}\right)^2}{\left|\sum_{j=1}^{2}(-1)^j\frac{z_k(\mathbf{r}) - \mu_j}{2\sigma_j^2}\right|}. \quad (9)$$

Proof. By a Taylor expansion of the $P\left(\tilde{\phi}\right)$ about the true values of parameters μ_1, μ_2, σ_1^2, σ_2^2 we have

$$P(\tilde{\phi}) = P_B + \sum_{l=1}^{2}\left(P_l^{(1)}\Delta\hat{\mu}_l + P_{\tilde{\sigma}_l^2}^{(1)}\Delta\tilde{\sigma}_l^2 + \frac{1}{2}P_{l,\tilde{\sigma}_l^2}^{(2)}\Delta\hat{\mu}_l\Delta\tilde{\sigma}_l^2 \right) + \\ + \frac{1}{2}\sum_{k,l=1}^{2}\left(P_{k,l}^{(2)}\Delta\hat{\mu}_k\Delta\hat{\mu}_l + P_{\tilde{\sigma}_{k,l}^2}^{(2)}\Delta\tilde{\sigma}_k^2\Delta\tilde{\sigma}_l^2 \right) + O_3, \tag{10}$$

where O_3 is the third and higher order terms of $\Delta\hat{\mu}_l$ and $\Delta\tilde{\sigma}_l^2$ and their products.

Explicit formula (7) for P_B is obtained by straightforward calculations in (3) for considered parametric structure case.

Since $P(\tilde{\phi})$ is minimised at the true values of parameters, then, for l=1,2,

$$P_l^{(1)} = 0 \qquad \text{and} \qquad P_{\tilde{\sigma}_l^2}^{(1)} = 0\,. \tag{11}$$

Define the quantity $G(z(\mathbf{r})) = \sum_{l=1}^{2}(-1)^{l+1}\pi_l(\mathbf{r})p_l(z(\mathbf{r}))$. Note that (see K. Dučinskas, [11])

$$P_{l,l}^{(2)} = \sum_{l=1}^{2}\left(\partial_l G(z_k(\mathbf{r}))\right)^2 \left|\nabla_{z(\mathbf{r})}G(z_k(\mathbf{r}))\right|^{-1} \tag{12}$$

and

$$P_{\tilde{\sigma}_{l,l}^2}^{(2)} = \sum_{l=1}^{2}\left(\partial_{\tilde{\sigma}_l^2}G(z_k(\mathbf{r}))\right)^2 \left|\nabla_{z(\mathbf{r})}G(z_k(\mathbf{r}))\right|^{-1}, \tag{13}$$

where

$$\partial_l G(z_k(\mathbf{r})) = \sum_{l=1}^{2}(-1)^{l+1}\frac{z_k(\mathbf{r})-\mu_l}{\sigma_l^2},$$

$$\partial_{\tilde{\sigma}_l^2}G(z_k(\mathbf{r})) = \sum_{l=1}^{2}(-1)^{l+1}\frac{(z_k(\mathbf{r})-\mu_l)^2-\sigma_l^2}{2(\sigma_l^2)^2},$$

$$\nabla_{z(\mathbf{r})}G(z(\mathbf{r})) = \frac{\partial G(z(\mathbf{r}))}{\partial z(\mathbf{r})} = \sum_{l=1}^{2}(-1)^{l}\frac{z_k(\mathbf{r})-\mu_l}{\sigma_l^2}.$$

Using (4) and (5), for l=1,2, under the independence of estimators $\hat{\mu}_l$ and $\tilde{\sigma}_l^2$, we have (J. Šaltytė, [12])

$$E\{\Delta\hat{\mu}_l\} = E\{\Delta\hat{\mu}_1\Delta\hat{\mu}_2\} = E\{\Delta\tilde{\sigma}_l^2\} = \\ = E\{\Delta\tilde{\sigma}_1^2\Delta\tilde{\sigma}_2^2\} = E\{\Delta\hat{\mu}_l\Delta\tilde{\sigma}_l^2\} = 0, \tag{14}$$

$$E\{(\Delta\hat{\mu}_l)^2\} = \frac{1}{c_l^{\bullet\bullet}}\sigma_l^2, \quad E\{(\Delta\tilde{\sigma}_l^2)^2\} = \frac{2}{N_l - 1}(\sigma_l^2)^2. \tag{15}$$

Under the assumptions 1, 2 it follows, that the expectation with respect to the distribution of the training sample **T** of residual O_3 in (10) is of order $O\left(\frac{1}{N^2}\right)$. Hence, by substituting the estimators (4) and (5) to (10), taking the expectation of the right side of (10) and using (11)-(15), we complete the proof of the theorem.

As the contribution of higher order terms in the presented asymptotic expansion is in generally negligible, for the evaluation of the performance of QDF the asymptotic EER is used. The asymptotic EER for the considered case is designated as

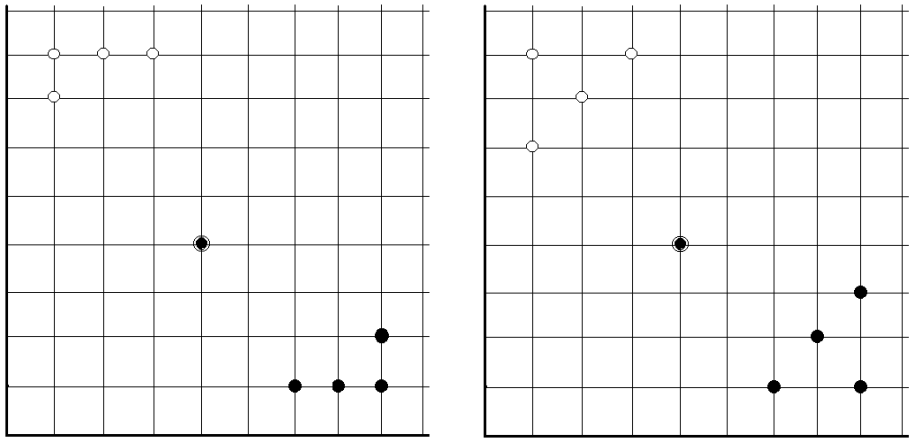
$$\text{AEER} = \frac{1}{2}\sum_{l=1}^{2}\frac{\alpha_l}{c_l^{\bullet\bullet}} + \sum_{l=1}^{2}\frac{\beta_l}{N_l - 1}.$$

Minimum of AEER could also be used as a criterion for optimal training sample design.

4 Numerical example

As an example consider the integer regular 2-dimensional lattice. We use two designs for training sample of size 4 for each class. Locations in design (A) are more compact than those in design (B) (Figure 2).

In this example without loss of generality we use the convenient canonical form of $\phi = (\Delta, 0, \sigma^2, 1)^T$, where $\Delta = \frac{\mu_1 - \mu_2}{\sigma^2}$ and $\sigma^2 = \frac{\sigma_1^2}{\sigma_2^2}$.



(A) (B)

Fig. 2. Training sample designs (locations from T_1 and T_2 are signed as ○ and ●, respectively; ◉ denotes the location $\mathbf{r}$)

Consider for both populations the spherical correlation function for observations $Z(\mathbf{s})$ and $Z(\mathbf{t})$ ([6], ch.2)

$$c(|\mathbf{h}|)=\begin{cases}\dfrac{\kappa_1}{\kappa_0+\kappa_1}\left(1-\dfrac{3}{2}\dfrac{|\mathbf{h}|}{\eta}+\dfrac{1}{2}\dfrac{|\mathbf{h}|^3}{\eta^3}\right), & 0\le|\mathbf{h}|\le\eta,\\ 1, & |\mathbf{h}|=0,\\ 0, & |\mathbf{h}|>\eta,\end{cases}$$

for nonnegative κ_0, κ_1, η. The nugget effect is κ_0 and the sill is $\kappa_0+\kappa_1$. For this model, observations more than η units apart are uncorrelated, so the range is η. Assume, that there is no nugget effect, i.e. $\kappa_0=0$, and range $\eta=3$.

In Table 1 the values of P_B and AEER are presented. Here also the comparison of obtained AEER with EER obtained under the assumption of independent observations (denoted by $\mathrm{AERR}_{\mathrm{ind}}$) is given. The ratio $\dfrac{\mathrm{AERR}}{\mathrm{AERR}_{\mathrm{ind}}}$ allows us to estimate the effect of spatial correlation on the

EER. In this table we consider only design (A). In Table 2 two experimental designs (A) and (B) (see Figure 2) are compared in terms of AEER.

Δ	P_B	AEER	AERR_{ind}	$\frac{\text{AERR}}{\text{AERR}_{\text{ind}}}$
0,5	0,292	0,00478	0,00474	1,00900
0,6	0,289	0,00414	0,00410	1,00921
0,7	0,287	0,00361	0,00357	1,00938
0,8	0,284	0,00317	0,00314	1,00948
0,9	0,281	0,00282	0,00279	1,00947
1,0	0,277	0,00254	0,00252	1,00934
1,1	0,273	0,00232	0,00230	1,00908
1,2	0,269	0,00214	0,00212	1,00870
1,3	0,265	0,00200	0,00199	1,00824
1,4	0,26	0,00189	0,00187	1,00772
1,5	0,255	0,00179	0,00177	1,00718
1,6	0,25	0,00170	0,00169	1,00663
1,7	0,244	0,00161	0,00160	1,00610
1,8	0,239	0,00153	0,00152	1,00561
1,9	0,233	0,00145	0,00144	1,00515
2,0	0,221	0,00127	0,00126	1,00474

Table 1. Values of the asymptotic expected error regret for the design (A), when $\pi_1 = 0.5$, $\mu_2 = 14$ and $\sigma^2 = 6$.

As it was expected, P_B and AEER are decreasing, when the difference between class-means increases (Tables 1 and 2). It is seen from the last column of Table 1, that the change in EER due to a spatial correlation is smaller for more separated populations. Hence, for close populations it is very important take into consideration the spatial dependence factor, when practical problems are solved. The last column in Table 2 confirms the advantage of design of wider spread locations.

Δ	$\mathrm{AEER}(\mathrm{A})$	$\mathrm{AEER}(\mathrm{B})$	$\frac{\mathrm{AEER}(\mathrm{A})}{\mathrm{AEER}(\mathrm{B})}$
0,5	0,00478	0,00477	1,00354
0,6	0,00414	0,00412	1,00362
0,7	0,00361	0,00359	1,00369
0,8	0,00317	0,00316	1,00373
0,9	0,00282	0,00281	1,00373
1,0	0,00254	0,00253	1,00367
1,1	0,00232	0,00231	1,00357
1,2	0,00214	0,00214	1,00343
1,3	0,00200	0,00200	1,00325
1,4	0,00189	0,00188	1,00304
1,5	0,00179	0,00178	1,00283
1,6	0,00170	0,00169	1,00261
1,7	0,00161	0,00161	1,00241
1,8	0,00153	0,00153	1,00221
1,9	0,00145	0,00144	1,00203
2,0	0,00136	0,00136	1,00187

Table 2. Comparison of the asymptotic expected error regrets for two different designs (A) and (B), when $\pi_1 = 0.5$, $\mu_2 = 14$ and $\sigma^2 = 6$.

5 References

1. Gupta P.L., Riley J.T., White T.J. "Misclassification probabilities for quadratic discrimination". *SIAM J.Sci.Stat.Comput.*, **7**(4), 1986.
2. Mardia K.V., Marshall R.J. "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression". *Biometrica*, **71**(1), 135-46, 1984.
3. Jain A.K., Duin R.P.W., Mao J. "Statistical Pattern Recognition: a Review". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(1), 4-38, 2000.
4. Wakaki H. "Comparison of Linear and Quadratic Discriminant Functions". *Biometrica*, 77, 227-229, 1990.

5. Lawoko C.R.O., McLachlan G.J. “Discrimination with Autocorrelated Observations”. *Pattern Recognition*, **18**(2), 145-149, 1985.
6. Cressie N.A.C. *Statistics for Spatial Data*. Wiley&Sons, 1993.
7. Dučinskas K. “An Asymptotic Analysis of the Regret Risk in Discriminant Analysis Under Various Training Schemes”. *Lith. Math. J.*, **37**(4), 337-351, 1997.
8. Mardia K.V. “Spatial Discrimination and Classification Maps”. *Commun. Statist. – Theor. Meth.*, **13**(18), 2181-2197, 1984.
9. Scherwish M.J. “Asymptotic Expansions for Correct Classification Rates in Discriminant Analysis”. *The Annals of Statistics*, **9**(5), 1002-1009, 1981.
10. Taniguchi M. “Higher order Asymptotic Theory for Discriminant Analysis in Exponential Families of Distributions”. *Journal of Multivariate Analysis*, **48**, 169-187, 1994.
11. Dučinskas K. “Second-order Expansion for the Expected Regret Risk in Classification of One-parametric Distributions”. *Lith. Math. J.*, **39**(2), 220-230, 1999.
12. Šaltytė J. *The Asymptotic Expansion of the Expected Risk for the LDA of Spatially Correlated Gaussian Observations*. PhD thesis. Klaipėda university, 2001.