LMD mokslo darbai, 372–376
1998, Vilnius

The risk of classification based on observations of anisotropic Gaussian random fields

J. Šaltytė, K. Dučinskas (KU)

1. Introduction

Let Z^2 be the 2-dimensional infinite integer lattice and let D denote a finite rectangular lattice within Z^2. Let $r = (r^1, r^2)$ be any point in D and assume that there are n points in D so that we can write $D = \{r(i), i = 1, \ldots, n\}$. Suppose that any point $r \in D$ can be assigned to one of two classes Ω_1, Ω_2 with positive prior probabilities π_1, π_2, respectively, where $\pi_1 + \pi_2 = 1$. The class of the point r is given by the random 2-dimensional vector $Y_r^T = (Y_{1r}, Y_{2r})$ of zero-one variables. The ith component of Y is defined to be one or zero according as an class of point r is or not Ω_i $(i = 1, 2)$. Then $Y_r \sim Mult_2(1; (\pi_1, \pi_2))$.

Suppose a p-dimensional observation $X_r \in \mathbf{X} \subset R^p$ can be made at each point $r \in D$. A decision is to be made as to which class the randomly chosen point $r \in D$ is assigned on the basis of observed value of X_r. The observed value of X, Y are denoted by x and y, respectively.

Let $X_r = \sum_{i=1}^{2} Y_{ir}\mu_i + \varepsilon_r$, where $\mu_1, \mu_2 \in R^p$, $\mu_1 \neq \mu_2$ and the noise $\varepsilon_r = (\varepsilon_r^1, \ldots, \varepsilon_r^p)$ are the realisations of the zero-mean stationary spatially correlated random process.

The first assumption is that this process is Gaussian with locally spatial anisotropic covariance. Hence, the common class-conditional covariance between any two observations X_r and X_s at points $r, s \in D$ can be factored as $\mathrm{cov}(X_r, X_s) = \rho(r - s)\Sigma$, $(r \neq s)$, where $\rho(\cdot)$ is the anisotropic correlation function, $\rho(0) = 1$ and $\Sigma = \mathrm{cov}\,(\varepsilon_r, \varepsilon_r)$ $(k, l = 1, \ldots, p)$.

Let the set of D points in vicinity of r denoted as $N_r = \{r(1), \ldots, r(m)\}$ then represents the neighbourhood of any point $r \in D$. Let X_{N_r} contains the observations on these points in the prescribed neighbourhood of point r, that is $X_{N_r} = (X_{r(1)}^T, \ldots, X_{r(m)}^T)^T$.

The second assumption about the joint distribution of X_{N_r} assumes local spatial continuity; that is, if $Y_{ir} = 1$, then $Y_{ir(j)} = 1$ with the probability close to 1, $i = 1, 2$, $j = 1, \ldots, m$. Note that this assumption will hold within the boundaries for each class but will not be true near the class boundaries. Its practical implication will not be discussed here.

Then the mean vector for $X_r^+ = \left(X_r^T, X_{N_r}^T\right)^T$ is

$$\mu_i^+ = E\left\{X_r^+ / Y_{ir} = 1\right\} \cong 1_{m+1} \otimes \mu_i \quad (i = 1, 2), \tag{1}$$

where 1_{m+1} is the $(m + 1)$-dimensional vector of ones, and $\otimes$ is the Kronecker delta (Cressie, 1993; Haslet & Horgan, 1987). The covariance matrix of X_r^+, given that r

belongs to Ω_i is

$$\Sigma^+ = C \otimes \Sigma, \tag{2}$$

where C is the anisotropic spatial correlation matrix of order $(m+1) \times (m+1)$, whose (k,l)th element is $\rho\,(r\,(k-1) - r\,(l-1))$ $(k, l = 1, \ldots, m+1)$ and $r_{(0)} = r$.

Presmoothing of the data is accomplished by implementing the assignment of point r on the basis of the value x_r^+ of the augmented vector X_r^+ (Cressie, 1993). Under the assumptions above, the ith class-conditional distribution of X_r^+ is $(m+1) \times p$-variate normal with mean (1) and covariance matrix (2).

Let $p_i(x_r)$ and $p_i^+(x_r^+)$ denote the probability densities of x_r and x_r^+, respectively, when $Y_{ir} = i$. Let $d(\cdot)$ denotes a classification rule, where $d(x_r) = i$ implies that point r with observation $X_r = x_r$ is to be assigned to class Ω_i $(i = 1, 2)$. Similarly let $d^+(x_r^+)$ is classification rule based on augmented observation $X_r^+ = x_r^+$.

The losses of classification when a point from class Ω_i is allocated to class Ω_j are denoted by $L(i, j)$. Then the risks of classification based on rules $d(\cdot)$ and $d^+(\cdot)$ can be expressed as

$$R\,(d\,(\cdot)) = \sum_{i=1}^{2} \pi_i \int\limits_{\mathbf{X}} L\,(i, d\,(x))\, p_i\,(x)\, dx$$

and

$$R^+ = R\left(d_B^+\,(\cdot)\right) = \sum_{i=1}^{2} \pi_i \int\limits_{\mathbf{X}^{m+1}} L\left(i, d^+\,(x)\right) p_i^+\,(x)\, dx,$$

respectively. Then Bayes classification rules (BCR) $d_B\,(\cdot)$ and $d_B^+\,(\cdot)$ minimising $R\,(d\,(\cdot))$ and R^+, respectively, are defined as

$$d_B\,(x) = \arg\max_{i=1,2} l_i\, p_i(x), \qquad d_B^+(x^+) = \arg\max_{i=1,2} l_i\, p_i^+(x^+),$$

where $l_i = \pi_i(L(3-i, i) - L(i, i))$, $i = 1, 2$.

2. The risk of the classification

Consider the special case of anisotropic spatial correlation function

$$\rho(h; t, \alpha) = \exp\Big(-\alpha\sqrt{t^2 h_1^2 + h_2^2}\Big), \tag{3}$$

where $h^T = (h_1, h_2)$, $t > 0$, $t \neq 1$, $\alpha > 0$. It is obvious that in the case $t = 1$ it becomes isotropic correlation function.

Anisotropic spatial correlation function defined in (3) is applicable to the situation, when the behavior of the process X_r in the N-S direction is different from that in W-E direction, i.e. when $t < 1$, the correlation between observations in the N-S direction decreases faster than that in the W-E direction, whereas in the case of $t > 1$ decreasing of correlation in the W-E direction is faster than in the direction N-S.

Consider three situations based on the positions of classified point r in D for the first-order neighbourhood scheme (Dučinskas, Šaltytė, 1998).

Situation A. The point r and all first-order neighbours are inside D.

Situation B. The point r is on the boundary of D and three first-order neighbours are inside D.

Situation C. The point r is at the corner of D and two first-order neighbours are inside D.

Assume, that X_{rA}^{+}, X_{rB}^{+}, X_{rC}^{+} and R_{0A}^{+}, R_{0B}^{+}, R_{0C}^{+} denote vectors of augmented observations and Bayes classification risks for above situations A, B and C, respectively. Let $\Delta^2 = (\mu_1 - \mu_2)^T\, \Sigma^{-1}\, (\mu_1 - \mu_2)$ is the square of Mahalanobis distance and $\gamma_1 = \ln\left(\frac{l_1}{l_2}\right)$.

THEOREM 1. *Let $d_B^{+}\left(X_{rA}^{+}\right)$ is used for the classification of $r \in D$ in the situation* A. *Then R_{0A}^{+} is equal*

$$R_{0A}^{+} = \sum_{i=1}^{2}\left(\pi_i L\,(i,1) - (-1)^i\, l_i \Phi\left((-1)^i\, \frac{\Lambda_A \Delta}{2} - \frac{\gamma_1}{\Lambda_A \Delta}\right)\right),$$

where

$$\begin{aligned}\Lambda_A^2 = {} & (4e^{(-\alpha(t+2))} - e^{(-2\alpha(t+1))} + e^{-2\alpha} - 8e^{\left(-\alpha\left(1+\sqrt{t^2+1}\right)\right)} + 8e^{\left(-\alpha\sqrt{t^2+1}\right)} + e^{(-2\alpha t)} \\ & + 4e^{(-\alpha(1+2t))} - 8e^{\left(-\alpha\left(t+\sqrt{t^2+1}\right)\right)} - 5 + 4e^{(-\alpha)} - 8e^{(-\alpha(t+1))} \\ & + 4e^{\left(-2\alpha\sqrt{t^2+1}\right)} + 4e^{(-\alpha t)})/(3e^{(-2\alpha(t+1))} + e^{(-2\alpha t)} \\ & + e^{(-2\alpha)} + 4e^{\left(-2\alpha\sqrt{t^2+1}\right)} - 8e^{\left(-\alpha\left(t+\sqrt{t^2+1}+1\right)\right)} - 1).\end{aligned}$$

Proof. Square of Mahalanobis distance between classes Ω_1 and Ω_2 based on augmented observation X_{rA}^{+} is

$$\left(\Delta_A^{+}\right)^2 = \left(\mu_1^{+} - \mu_2^{+}\right)^T \left(\Sigma_A^{+}\right)^{-1} \left(\mu_1^{+} - \mu_2^{+}\right) = (1_5 \otimes \Delta\mu)^T \left(C_A^{+} \otimes \Sigma\right)^{-1} (1_5 \otimes \Delta\mu),$$

where spatial correlation matrix C_A^{+} is

$$C_A^{+} = \begin{pmatrix} 1 & e^{(-\alpha t)} & e^{(-\alpha)} & e^{(-\alpha t)} & e^{(-\alpha)} \\ & 1 & e^{\left(-\alpha\sqrt{t^2+1}\right)} & e^{(-2\alpha t)} & e^{\left(-\alpha\sqrt{t^2+1}\right)} \\ & & 1 & e^{\left(-\alpha\sqrt{t^2+1}\right)} & e^{(-2\alpha)} \\ & & & 1 & e^{\left(-\alpha\sqrt{t^2+1}\right)} \\ & & & & 1 \end{pmatrix}.$$

Using the property $\left(C_A^{+} \otimes \Sigma\right)^{-1} = \left(C_A^{+}\right)^{-1} \otimes \Sigma^{-1}$ and taking an inverse of C_A^{+} the proof of the theorem is completed.

THEOREM 2. *Let $d_B^+(X_{rB}^+)$ and $d_B^+(X_{rC}^+)$ are used for the classification of $r \in D$ in the situations* B *and* C, *respectively. Then R_{0B}^+ and R_{0C}^+ are*

$$R_{0B}^+ = \sum_{i=1}^{2} \left(\pi_i L(i,1) - (-1)^i l_i \Phi \left((-1)^i \frac{\Lambda_B \Delta}{2} - \frac{\gamma_1}{\Lambda_B \Delta} \right) \right),$$

$$R_{0C}^+ = \sum_{i=1}^{2} \left(\pi_i L(i,1) - (-1)^i l_i \Phi \left((-1)^i \frac{\Lambda_C \Delta}{2} - \frac{\gamma_1}{\Lambda_C \Delta} \right) \right),$$

where

$$\begin{aligned}\Lambda_B^2 = {} & 2(e^{(-\alpha(t+2))} + e^{\left(-2\alpha\sqrt{t^2+1}\right)} + 2e^{\left(-\alpha\sqrt{t^2+1}\right)} - 2e^{(-\alpha(t+1))} + 2e^{(-\alpha)} - 2e^{\left(-\alpha\left(1+\sqrt{t^2+1}\right)\right)} \\ & - 2 - 2e^{\left(-\alpha\left(t+\sqrt{t^2+1}\right)\right)} + e^{(-2\alpha t)} + e^{(-\alpha t)})/(e^{(-2\alpha)} + e^{(-2\alpha(t+1))} + 2e^{\left(-2\alpha\sqrt{t^2+1}\right)} \\ & - 4e^{\left(-\alpha\left(t+\sqrt{t^2+1}+1\right)\right)} - 1 + e^{(-2\alpha t)})\end{aligned}$$

and

$$\begin{aligned}\Lambda_C^2 = {} & (-3 + e^{\left(-2\alpha\sqrt{t^2+1}\right)} + 2e^{(-\alpha t)} - 2e^{\left(-\alpha\left(1+\sqrt{t^2+1}\right)\right)} - 2e^{\left(-\alpha\left(t+\sqrt{t^2+1}\right)\right)} \\ & + 2e^{(-\alpha)} + e^{(-2\alpha)} + 2e^{\left(-\alpha\sqrt{t^2+1}\right)} \\ & - 2e^{(-\alpha(t+1))} + e^{(-2\alpha t)})/(-1 + e^{\left(-2\alpha\sqrt{t^2+1}\right)} \\ & + e^{(-2\alpha t)} - 2e^{\left(-\alpha\left(t+\sqrt{t^2+1}+1\right)\right)} + e^{(-2\alpha)}).\end{aligned}$$

Proof. The proof of theorem is similar to the proof of Theorem 1 only replacing C_A^+ by

$$C_B^+ = \begin{pmatrix} 1 & e^{(-\alpha t)} & e^{(-\alpha)} & e^{(-\alpha)} \\ & 1 & e^{\left(-\alpha\sqrt{t^2+1}\right)} & e^{\left(-\alpha\sqrt{t^2+1}\right)} \\ & & 1 & e^{(-2\alpha)} \\ & & & 1 \end{pmatrix}$$

and

$$C_C^+ = \begin{pmatrix} 1 & e^{(-\alpha t)} & e^{(-\alpha)} \\ & 1 & e^{\left(-\alpha\sqrt{t^2+1}\right)} \\ & & 1 \end{pmatrix}.$$

Values of Λ_A^2, Λ_B^2, Λ_C^2 derived in Theorem 1 and Theorem 2 are presented in Table 1 and Table 2 for $t = 0.5$ and $t = 2$, respectively.

From Tables 1 and 2 it can be compared the risk of classification for different types of data augmentation and different cases of anisotropy.

Table 1. Values of Λ_A^2, Λ_B^2, Λ_C^2 for $\rho(h; 0.5, \alpha) = \exp\left(-\alpha\sqrt{0.25h_1^2 + h_2^2}\right)$.

α	Λ_A^2	Λ_B^2	Λ_C^2
1	2,124911	2,020455	1,62489
2	3,186919	2,852486	2,154622
3	3,939158	3,37405	2,504078
4	4,390459	3,659138	2,710289
5	4,646582	3,809888	2,829076
6	4,792082	3,891093	2,89812
7	4,876226	3,936291	2,938833
8	4,925738	3,962198	2,963113
9	4,955232	3,977369	2,977698
10	4,972939	3,986379	2,986496

Table 2. Values of Λ_A^2, Λ_B^2, Λ_C^2 for $\rho(h; 2, \alpha) = \exp\left(-\alpha\sqrt{4h_1^2 + h_2^2}\right)$.

α	Λ_A^2	Λ_B^2	Λ_C^2
1	3,186919	2,550691	2,154622
2	4,390459	3,456671	2,710289
3	4,792082	3,801187	2,89812
4	4,925738	3,926896	2,963113
5	4,972939	3,973084	2,986496
6	4,990073	3,990091	2,99504
7	4,996351	3,996353	2,998176
8	4,998658	3,998658	2,999329
9	4,999506	3,999506	2,999753
10	4,999818	3,999818	2,999909

REFERENCES

[1] Dučinskas K. and Šaltytė J., *Classification of points in 2-dimensional lattice based on realisations of Gaussian random fields*, Proceedings of the XXXVIII conference of Lithuanian Mathematical Society, Vilnius, 1998,

[2] Cressie N. A. C., *Statistics for Spatial Data*, Wiley, New York, 1993.

[3] Haslet J. and Horgan G., *Linear Models in Spatial Discriminant Analysis*, NATO ASI series, 1987.

[4] Mardia K. V., Spatial Discrimination and Classification Maps, *Comm. Statist. Theor. Meth.*, (1984),

[5] McLaclan G. J., *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, 1992.

Anizotropinių atsitiktinių Gauso laukų stebėjimų klasifikavimo rizika

J. Šaltytė, K. Dučinskas

Straipsnyje nagrinėjamas baigtinės dvimatės gardelės vidinių bei ribinių taškų klasifikavimo uždavinys pagal atsitiktinių lokaliai tolydžių Gauso laukų su lokaliai anizotropine kovariacija realizacijas. Išvestos analitinės formulės šių taškų klasifikavimo rizikai skaičiuoti pirmos eilės kaimynų schemoje.