LMD konf. darbai, 193–197
1997, Vilnius

Classification of points in 2-dimensional lattice based on realisations of Gaussian random fields

K. Dučinskas (KU), J. Šaltytė (KU)

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 $D = \{r(i),\ i = 1, \dots, n\}$. Suppose that any point $r \in D$ can be assigned to one of two classes 1, 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 i-th component of Y is defined to be one or zero according as the class of point r is i or not $(i = 1, 2)$.

$$\text{Then } Y_r \sim \text{Mult}_2\left(1; (\pi_1, \pi_2)\right). \tag{1}$$

Suppose a p-dimensional observation $X_r \in R^p$ 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

$$X_r = \sum_{i=1}^{2} Y_{ir}\mu_i + \varepsilon_r, \tag{2}$$

where $\mu_1, \mu_2 \in R^p$, $\mu_1 \neq \mu_2$ and the noise $\varepsilon_{r(1)}, \dots, \varepsilon_{r(n)}$ are the realinations of the zero – mean stationary spatially correlated random process.

The first assumption is that this process is Gaussian with locally spatial isotropic 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

$$\text{cov}(X_r, X_s) = \rho(d)\Sigma, \quad (r \neq s), \tag{3}$$

where $\rho(\cdot)$ is the isotropic correlation function, $\rho(0) = 1$, and d is the Euclidean distance between points r and s, $\Sigma = \text{cov}(\varepsilon_r, \varepsilon_r)$.

Let the set of points of D in vicinity of r, denoted as $N_r = \{r_1, \dots, r_m\}$, represent the neighbourhood of any point $r \in D$.

Then let X_{N_r} contain the observations at these points in the prescribed neighbour of pixel r, that is

$$X_{N_r} = \left(X_{r_1}^T, \dots, X_{r_m}^T\right)^T.$$

For example, $m = 4$ for the first-order neighbourhood of adjacent points, while $m = 8$ for the second-order neighbourhood including also diagonally adjacent points.

The second assumption about the joint distribution of $X_{r(1)}, \ldots, X_{r(n)}$ assumes local spatial continuity of the neighbourhood in that if point r belongs to i, then so does every neighbour.

Thus, letting

$$X_r^+ = \left(X_r^T, X_{N_r}^T\right)^T \tag{4}$$

we have

$$\mu_i^+ = E\{X_r^+ \mid Y_{ir} = 1\} = 1_{m+1} \otimes \mu_i \quad (i = 1, 2), \tag{5}$$

where $\otimes$ is the Kronecker delta, and 1_{m+1} is the $(m+1)$-dimensional vector of ones. The covariance matrix of X_r^+, given that r belongs to i is

$$\Sigma^+ = R \otimes \Sigma, \tag{6}$$

where R is the spatial correlation matrix of order $(m+1) \times (m+1)$, whose (i, j)-th element is $\rho(\|r_{j-1} - r_{i-1}\|)$, denoting $r = r_0$, $(i, j = 0, \ldots, m)$. The spatial correlation matrix R have to be positive definite for any number of observations in D.

Presmoothing of the data is accomplished by implementing the assignment of point r on the basis of the value x_r^+ of the augmented vectors X_r^+ (see e.g. [1]). Under the assumptions above, the i-th class conditional distribution of X_r^+ is $(m+1) \times p$-variate normal with mean (5) and covariance matrix (6).

Let $p_i(x_r)$ and $p_i^+(x_r^+)$ denote the probability densities of x_r and x_r^+, respectively, when $c(r) = i$.

Let $d(\cdot)$ denote a classification rule, where $d(x_r) = i$ implies that point r with observation $X_r = x_r$ is to be assigned to the class 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 i is allocated to class j is denoted by $L(i, j)$. Then the risks of classification based on rules $d(\cdot)$ and $d^+(\cdot)$ can be expressed as

$$R = R\big(d(\cdot)\big) = \sum_{i=1}^{2} \pi_i \int_X L\big(i, d(x)\big) p_i(x)\, dx$$

and

$$R^+ = R\big(d^+(\cdot)\big) = \sum_{i=1}^{2} \pi_i \int_{X^{m+1}} L\big(i, d^+(x)\big) p_i^+(x)\, dx,$$

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

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

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

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

Risk of classification R_0 for the BCR $d_B(\cdot)$ is equal (see e.g. [2])

$$R(d_B(\cdot)) = R_0 = \sum_{i=1}^{2} \left(\pi_i L(i,1) - (-1)^i l_i \Phi\left((-1)^i \frac{\Delta}{2} - \frac{\gamma_1}{\Delta} \right) \right) \tag{7}$$

where $\Delta^2 = (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2)$, $\gamma_1 = \ln(l_1/l_2)$.

Consider three situations based on positions of classified point r in D for the first-order neighbourhood scheme (see Figure 1).

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 on the boundary of D and two first-order neighbours are inside D.

Let 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.

THEOREM. *Let $d_B^+(X_{rA}^+)$ be used for the classification of $r \in D$ i. e. $c(r) = i$ iff $d_B^+(X_{rA}^+) = i$ $(i = 1, 2)$. Then R_{0A} is equal to*

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

where

$$k_A = 1 + \left(4(\rho(1) - 1)^2\right) / \left(1 + 2\rho\sqrt{2}) + \rho(2) - 4\rho^2(1)\right).$$

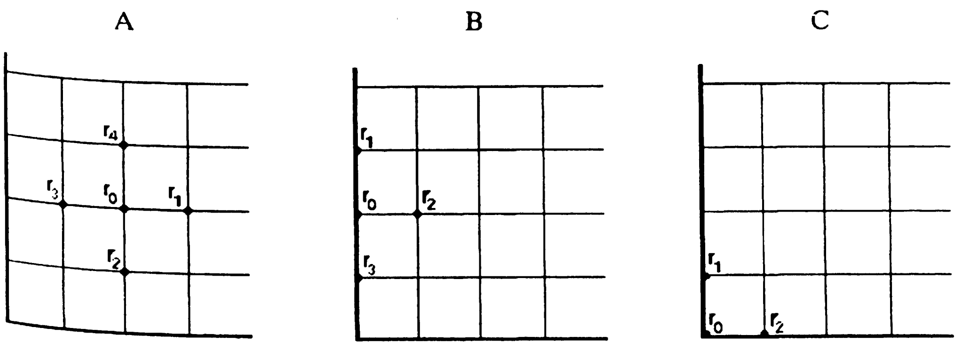

Figure 1. Three positions of point r in D.

Proof. Square of Mahalanobis distance between classes 1 and 2 based on augmented observation X_{rA}^{+} is

$$\Delta_A^2 = (\mu_1^+ - \mu_2^+)^T (\Sigma_A^+)^{-1} (\mu_1^+ - \mu_2^+) = (1_5 \otimes (\mu_1 - \mu_2))^T (R_A \otimes \Sigma)^{-1} 1_5 \otimes (\mu_1 - \mu_2),$$

where spatial correlation matrix R_A is

$$R_A = \begin{pmatrix} 1 & \rho(1) & \rho(1) & \rho(1) & \\ & 1 & \rho(\sqrt{2}) & \rho(2) & \rho(\sqrt{2}) \\ & & 1 & \rho(\sqrt{2}) & \rho(2) \\ & & & 1 & \rho(\sqrt{2}) \\ & & & & 1 \end{pmatrix}.$$

Using $(R_A \otimes \Sigma)^{-1} = R_A^{-1} \otimes \Sigma^{-1}$ and taking the inverse of R_A we complete the proof of the theorem.

Remark. In situation A our results coincides with Mardia [4] results derived for $\pi_1 = \pi_2$, $L(i, j) = 1 - \delta_{ij}$, where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

THEOREM 2. *For situations B and C the Bayes classification risks are*

$$R_{0B} = \sum_{i=1}^{2} \left(\pi_i L(i, 1) - (-1)^i l_i \Phi\left((-1)^i k_B \Delta/2 - \gamma_l/(k_B \Delta)\right)\right),$$

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

where

$$k_B = 1 + \frac{\left(3 + \rho(2) - 4\rho(\sqrt{2})(\rho(1) - 1\right)^2}{\left(1 + \rho(2) - 2\rho^2(\sqrt{2})\right) + 4\left(\rho(\sqrt{2}) - 3 - \rho(2)\right)\rho^2(1))\right)},$$

$$k_C = 1 + \frac{2(\rho(1) - 1)^2}{1 + \rho(\sqrt{2}) - 2\rho^2(1)}.$$

Proof. The poroof of Theorem 2 is similar to proof of theorem 1 only replacing R_A by

$$R_B = \begin{pmatrix} 1 & \rho(1) & \rho(1) & \rho(1) \\ & 1 & \rho(\sqrt{2}) & \rho(2) \\ & & 1 & \rho(\sqrt{2}) \\ & & & 1 \end{pmatrix}$$

and

$$R_C = \begin{pmatrix} 1 & \rho(1) & \rho(1) \\ & 1 & \rho(\sqrt{2}) \\ & & 1 \end{pmatrix}.$$

From positive definiteness of spatial correlation matrices it follows that

$$k_A > k_B > k_C. \tag{8}$$

For $0-1$ losses risk means the probability of misclassification P_0. Then from the results of Theorem 1, 2 and (8) it follows that $P_{0A} < P_{0B} < P_{0C}$, for the considered three situations.

The values of k_A, k_B and k_C for spatial correlation function $\rho(h) = \exp(-\alpha h)$ are presented in table 1.

Table 1.
Values of k_A, k_B and k_C for $\rho(h) = \exp(-\alpha h)$.

α	k_A	k_B	k_C
1	2.479601	2.193012	1.821796
2	3.81264	3.16568	2.462423
3	4.536285	3.673046	2.788971
4	4.8319	3.879958	2.921983
5	4.940122	3.956674	2.971643
6	4.978625	3.984367	2.989711
7	4.992317	3.994337	2.996257
8	4.997221	3.99794	2.998634
9	4.998989	3.999248	2.999501
10	4.999631	3.999725	2.999817

REFERENCES

[1] N. A. C. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, New York, 1993.

[2] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, John Willey & Sons, New York, 1992.

[3] J. Haslet and G. Horgan, Linear models in spatial discriminant analysis, *NATO ASI series F*, **30** (1987), 47–55.

Dvimatės gardelės taškų klasifikavimas pagal Gauso atsitiktinių laukų realizacijas

K. Dučinskas, J. Šaltytė

Straipsnyje pateiktis 2-matės baigtinės gardelės taškų klasifikavimo rizikos analitinės išraiškos, pagal izotropinių stacionarinių Gauso atsitiktinių laukų realizacijas. Duomenų priauginimui naudojama pirmos eilės kaimynų schema.