

Effect of anisotropy coefficient on error rates of linear discriminant functions

Kęstutis DUČINSKAS, Lina DREIŽIENĖ (KU)

e-mail: kestutis.ducinkas@ku.lt, budrikaite.lina@gmail.com

Abstract. Paper deals with statistical classification of spatial data as a part of widely applicable statistical approach to pattern recognition. Error rates in supervised classification of Gaussian random field observation into one of two populations specified by different constant means and common stationary geometric anisotropic covariance are considered. Formula for the exact Bayesian error rate is derived. The influence of the ratio of anisotropy to the error rates is evaluated numerically for the case of complete parametric certainty.

Keywords: geometric anisotropy, anisotropy ratio, Gaussian random field, equicorrelation, Bayesian discriminant function, actual error rate, expected error rate.

Introduction

The extension of classical statistical classification techniques in spatial data analysis is a problem of practical interest. Supervised statistical classification sometimes called discriminant analysis (DA) (see, e.g., McLachlan, [4]) traditionally assumes that observations to be classified and classified observations from training sample are independent. However, in practical situations with spatially distributed data this is usually not the case. Data that are close together in space, are likely to be correlated. Thus, to include spatial dependencies in the classification problem is very important.

When populations are completely specified (case of complete parametric certainty), an optimal classification rule in the sense of minimum misclassification probability is the Bayesian classification rule (BCR) (Anderson [1]). Switzer [5]) was the first to treat classification of spatial data, a work that was expended by Mardia [3]. However, neither of these authors analysed the error rate of classification.

In this paper we derived formula for Bayes error rate assuming that training sample observations and an observation to be classified are spatially correlated.

1. Statistical models for spatial population

In this paper, we consider the performances of BCR and plug-in BCR with parameters estimators from training sample.

Suppose that observation to be classified and training sample are considered as observations of a Gaussian random field $\{Z(s): s \in D \subset R_m\}$.

The model of $Z(s)$ in population Ω_l is

$$Z(s) = \mu_l + \varepsilon(s).$$

μ_l – unknown constant mean ($l = 1, 2$). The error term is generated by zero-mean stationary Gaussian random field $\{\varepsilon(s): s \in D\}$ with covariance function defined by model for all $s, u \in D$, $\text{cov}\{\varepsilon(s), \varepsilon(u)\} = C(s - u)$, where $C()$ is the covariance function of known parametric structure.

Geostatisticians usually use semivariogram as measure of spatial variation instead of covariance. On the assumption of stationarity the semivariogram is defined as: $\gamma(h) = C(0) - C(h)$.

There are three common used parameters of the stationary and isotropic semivariogram. If $\lim_{|h| \rightarrow \infty} \gamma(|h|) = \gamma_\infty < \infty$, then γ_∞ is called sill of semivariogram. The nugget effect shows the pure random variation in population density or it may be associated with sampling error. If $\gamma(|h|) \rightarrow \theta_0 > 0$ when $|h| \rightarrow 0$, then θ_0 is called nugget effect. The third parameter is the range of influence parameter, that defines the distance within which observations are correlated.

In many applications are found empirical evidence of directional effects in the covariance structure. The simplest way to deal with is by introducing geometric anisotropy into the assumed covariance structure. Geometric anisotropy means that models of the covariances (or semivariograms) have the same nugget, sill but different ranges in to perpendicular directions (Wackernagel, [6]).

Geometric anisotropy means that the correlation is stronger in one direction than it is in the other directions. If one plots the directional ranges in 2D case they would fall on the edge of an ellipse, where major and minor axes of ellipse correspond to the largest and shortest ranges of directional semivariograms.

Algebraically, it adds to the isotropic model two more parameters: the anisotropy angle ϕ and anisotropy ratio

$$\lambda = \frac{\text{Range}_{\max}}{\text{Range}_{\min}} > 1. \quad (1)$$

In this paper we restrict our attention to the nuggetless model of covariance, i.e., $C(h) = \sigma^2 r(h)$, where σ^2 is the variance (sill) and $r(h)$ is the spatial correlation function.

Procedures of fitting of the geometrically anisotropic semivariogram models to the environmental data can be easily realised by software system R package Gstat (see [2]).

2. Bayesian error rate in complete parametric certainty

Consider the problem of classification of the observation $Z^0 = Z(s_0)$, with $s_0 \in D$, into one of two populations specified above with given training sample T. Training sample T is specified by $T' = (T'_1, T'_2)$, where T'_l is the vector formed by n_l observations of $Z(s)$ from Ω_l , $l = 1, 2$. Let $n_1 + n_2 = n$.

Denote by R the correlation $n \times n$ matrix for vector T and by r_0 the vector of correlations between Z^0 and T. Since the observation Z^0 is correlated with training

sample, we have to deal with conditional means μ_{lT}^0 and variance σ_{0T}^2 that are defined by

$$\mu_{lT}^0 = E(Z^0/T; \Omega_l) = \mu_l + \alpha'(T - M), \quad l = 1, 2 \quad (2)$$

and

$$\sigma_{0T}^2 = V(Z^0/T; \Omega_l) = \sigma^2 K, \quad (3)$$

where

$$\alpha' = r'_0 R^{-1}, \quad M' = (\mu_1 1'_{n1}, \mu_2 1'_{n2}), \quad (4)$$

$$K = 1 - r'_0 R^{-1} r_0 \quad (5)$$

with 1_n specified as n vector of ones.

In this chapter we assume all covariance parameters are fixed except anisotropy ratio λ defined in (1).

Then spatial correlation function $r(h)$ may be considered as function of λ and it is obvious that K is also the function of λ , i.e., $K = K(\lambda)$.

Under the assumption that the populations are completely specified and for known prior probabilities of populations π_1 and π_2 ($\pi_1 + \pi_2 = 1$), the Bayesian discriminant function (BDF) minimizing the probability of misclassification (PMC) is

$$W_T(Z^0) = \left(Z^0 - \frac{1}{2}(\mu_{1T}^0 + \mu_{2T}^0) \right) (\mu_{1T}^0 - \mu_{2T}^0) / \sigma_{0T}^2 + \gamma, \quad (6)$$

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

Denote by P_B the PMC of BCR, usually called Bayesian error rate. Put $\Delta = |\mu_1 - \mu_2|/\sigma$.

LEMMA 1. Bayesian error rate for BDF defined in (6) is

$$P_B = \sum_{l=1}^2 \pi_l \Phi(Q_l),$$

where

$$Q_l = -\Delta_0/2 + (-1)^l/\Delta_0, \\ \Delta_0 = \Delta/\sqrt{K}.$$

Proof. Proof of Lemma follows directly from (2)–(6) and definition of PMC (see, e.g., [4]) and properties of Gaussian distribution [1].

So we can consider P_B as the function of anisotropy ratio λ .

2. Numerical example

As an example we consider the case with D being integer regular 2-dimensional lattice. We assume that location of observation to be classified is $s_0 = (3, 3)$ and

$$D_1 = (0, 3), (0, 2), (1, 1), (2, 1),$$

$$D_2 = (4, 1), (5, 1), (6, 2), (6, 3).$$

Here D_i is the set of points in D , where training sample T_i is taken, $i = 1, 2$. So $n_1 = n_2 = 4$. Let $h' = (h_x, h_y)$.

Exponential geometric anisotropic correlation function $r(h)$ with anisotropy angle $\phi = \frac{\pi}{2}$ specified by $r(h) = \exp\{-\sqrt{h_x^2 + \lambda^2 h_y^2}/\alpha\}$ is considered.

Example was considered to evaluate numerically the effect of the anisotropy ratio on the error rates of classification in case of parametric certainty. With an insignificant loss of generality the case with $\pi_1 = \pi_2 = 0.5$ and $\alpha = 2$ was considered. Then the

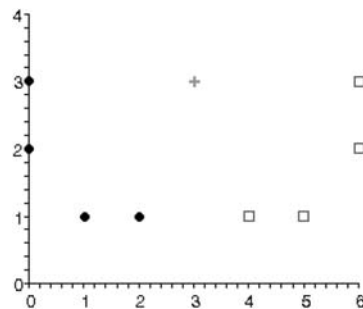


Fig. 1. Sampling points of T_1 (signed by „•“), T_2 (signed by „□“) and point s (signed „+“)

Table 1. Values of $P_B = \Phi(-\Delta_0/2)$ and $\eta = \Phi(-\Delta_0/2)/\Phi(-\Delta_0^*/2)$

λ	$\Phi(-\Delta_0/2)$			η		
	$\Delta = 0, 5$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0, 5$	$\Delta = 1$	$\Delta = 2$
1	0.390024	0.288247	0.131991	1	1	1
2	0.395177	0.297466	0.143799	1.013212	1.031984	1.089459
3	0.395852	0.298681	0.145394	1.014943	1.036200	1.101542
4	0.396114	0.299153	0.146015	1.015614	1.037834	1.106246
5	0.396249	0.299397	0.146338	1.015962	1.038683	1.108693
6	0.396314	0.299514	0.146493	1.016128	1.039090	1.109865
7	0.396343	0.299566	0.146561	1.016202	1.039269	1.110383
8	0.396355	0.299588	0.146590	1.016232	1.039344	1.110599
9	0.396360	0.299596	0.146601	1.016245	1.039374	1.110686
10	0.396362	0.299600	0.146606	1.016250	1.039386	1.110721

Bayesian error rate is specified by

$$P_B = \Phi(-\Delta_0/2)$$

Denote by $\Phi(-\Delta_0^*/2)$ the Bayes error rate for given sampling design with $\lambda = 1$.

Three values 0.5, 1 and 2 of Δ were chosen to represent weak, moderate and strong separation between populations. In Table 1 the values of $\Phi(-\Delta_0/2)$ and $\eta = \Phi(-\Delta_0/2)/\Phi(-\Delta_0^*/2)$ are presented. The ratio η can be used as relative measure of the influence or effect of the λ on Bayesian error rate.

The figures in Table 1 show that the effect of anisotropy ratio on Bayesian error rate is greater when separation between classes is stronger.

References

1. T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York (2003).
2. L. Budrikaitė, K. Dučinskas, Modelling of geometric anisotropic spatial variation, *Mathematical Modelling and Analysis*, 361–366 (2005).
3. K.V. Mardia, Spatial discrimination and classification maps, *Comm. Statist. Theory Methods*, **13**(18), 2181–2197 (1984).
4. G.J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York (2004).
5. P. Switzer, Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery, *Math. Geol.*, **12**(4) (1980).
6. H. Wackernagel, *Multivariate Geostatistics: an Introduction with Applications*, Springer-Verlag, Berlin Heidelberg (2003).

REZIUOMĖ

K. Dučinskas, L. Dreičienė. Anizotropiškumo santykio įtaka Gauso laukų klasifikavimo klaidai

Straipsnyje pateikiama stacionarių anizotropinių Gauso laukų stebėjimų klasifikavimo Bajeso klaidos išraiška. Skaitmeniškai tiriama jos priklausomybė nuo anizotropiškumo santykio dydžio.