



Supervised linear classification of Gaussian spatio-temporal data

Marta Karaliūtė^a , Kęstutis Dučinskas^{a,b} 

^a*Institute of Data Science and Digital Technologies, Vilnius university*
Akademijos str. 4, LT-08412 Vilnius

^b*Faculty of Marine Technologies and Natural Sciences, Klaipeda University*
Herkaus Manto str. 84, LT-92294 Klaipeda
E-mail: mkaraliute@gmail.com; k.ducinskas@gmail.com

Received September 1, 2021; published online December 15, 2021

Abstract. In this article we focus on the problem of supervised classifying of the spatio-temporal Gaussian random field observation into one of two classes, specified by different mean parameters. The main distinctive feature of the proposed approach is allowing the class label to depend on spatial location as well as on time moment. It is assumed that the spatio-temporal covariance structure factors into a purely spatial component and a purely temporal component following AR(p) model. In numerical illustrations with simulated data, the influence of the values of spatial and temporal covariance parameters to the derived error rates for several prior probabilities models are studied.

Keywords: separable covariance function; AR(p) model; Bayes discriminant function

AMS Subject Classification: 62H11

Introduction

Spatial supervised classification is a problem of labeling observations based on feature information and information about spatial adjacency relationships with training sample. This problem has been studied by numerous authors (see e.g. Atkinson and Lewis [1]). However, these studies are usually based on the assumption of conditional independence of feature observations. Comprehensive overview of methods for statistical classification and discrimination of Gaussian spatial data is provided by Berrett and Calder [3]. The novel approach to classification of Gaussian Random

Field (GRF) observation by avoiding the assumption of conditional independence is developed by Dučinskas and Dreičienė [5]. However, statistical discriminant analysis of spatio-temporal data has been rarely considered previously. Šaltytė-Benth and Dučinskas [8] considered classification of spatio-temporal data modeled by GRF in particular case when observation of feature at focal location is uncorrelated with the training sample.

In the present paper, avoiding independence restrictions, we focus on the classification of data modeled by random fields with separable spatio-temporal covariance structures specified by geostatistical spatial margins and discrete temporal margins (see e.g. [4]). Separability of covariances was assumed for the sake of reduction of complexity due to interdependencies between features.

The main distinctive feature of the proposed approach is allowing the class label to depend on spatial location as well as on time moment. That essentially widens the application area of presented investigations.

For the performance of classifiers, the values of derived in Local Bayes error rates and empirical error rates are used. For numerical illustrations, the exponential isotropic models for spatial covariance and AR(p) model of temporal covariance are considered. This is the extension of AR(1) case explored in Karaliutė and Dučinskas [7]. Performance of proposed classification rule is compared for various parameters of pure spatial and temporal covariances and prior class probabilities models.

This paper is organized as follows: proposed spatio-temporal data models and conditional distributions are delivered in the next section; in Section 2 conditional Bayes classification rule and its error rates are presented; in Section 3 the numerical illustrations and simulations for various separable stationary spatio-temporal covariance and prior probabilities models are displayed, and finally, the conclusions are in the last section.

1 Spatio-temporal data models and conditional distributions

The main objective of this paper is to classify feature observations of GRF $\{Z(s; t) : s \in D \subset \mathbb{R}^2, t \in D_T = [0, \infty]\}$, where s and t define spatial and temporal coordinates, respectively. Let $\{Y(s; t) : s \in D \subset \mathbb{R}^2, t \in D_T\}$ be a random field that represents class label and takes only the value 0 or 1.

In this study, we assume that for $l = 0, 1$ the model of observation $Z(s; t)$ conditional on $Y(s; t) = l$ is $Z(s; t) = \mu_l(s; t) + \varepsilon(s; t)$, where $\mu_l(s; t)$ – deterministic spatio-temporal drift. The error term is assumed to be generated by the univariate zero-mean GRF $\{\varepsilon(s; t) : s \in D \subset \mathbb{R}^2, t \in T\}$, with the separable spatio-temporal covariance model $cov(\varepsilon(s; t), \varepsilon(u; r)) = C(s, u; t, r)$ for all $s, u \in D$ and $t, r \in T$.

Here $C_S(s, u)$ denotes pure spatial covariance between observations in locations s and u and $C_T(t, r)$, denotes pure temporal covariance between observations at time points t and r . Under this assumption, the spatio-temporal covariance structure factors into a purely spatial and a purely temporal component, which allows for computationally efficient estimation and inference.

In this study we follow the popular tradition in environmental and agricultural research when the data are recorded at regular time intervals (time lags) and at irregular stations (locations) in compact area (see e.g. [2, 6]).

Let $S_n = \{s_i \in D; i = 1, \dots, n\}$ be a set of locations where observations are taken at time $t \in D_m = \{1, 2, \dots, m, m + 1\}$. At every moment of time $t \in D_m$, the set S_n is split into two classes, $S_t^{(0)}$ and $S_t^{(1)}$ (i.e., $S_n = S_t^{(0)} \cup S_t^{(1)}$): $S_t^{(l)} = \{s \in S_n : Y(s, t) = l\}$, $l = 0, 1$. Denote n_{lt} the number of locations (of n) at time t that belong to class l ; thus n_{lt} is the number of points in the set $S_t^{(l)}$, and $n = n_{0t} + n_{1t}$ for every $t \in D_m$. Hence a set of class labels at any time moment can differ in composition.

Joint training sample Z is stratified training sample, specified by $n \times m$ matrix $Z = (Z_1, \dots, Z_m)$, where $Z_t = (Z(s_1, t), \dots, Z(s_n, t))'$. This structure of data presentation is motivated by a model that assumes multivariate (in space) time series. Denote by $z_t = (z_t^1, \dots, z_t^n)$ and $y_t = (y_t^1, \dots, y_t^n)$ the realized value of Z_t and $Y_t = (Y(s_1, t), \dots, Y(s_n, t))'$, respectively.

In this study we consider the linear parametric drift $\mu_l(s; t) = \beta_l' x(s)$, where $x_i = (x_1(s_i), \dots, x_q(s_i))'$ is the vector of a spatial covariates, and β_l is a q dimensional vector of parameters, $i = 1, \dots, n$, $l = 0, 1$ and $\Delta\beta = \beta_1 - \beta_0$.

Denote by X the $n \times 2qm$ matrix $X = (X_{(1)}, X_{(2)}, \dots, X_{(m)})$ where

$$X_{(t)} = \begin{pmatrix} x_1'(1 - y_t^1) & x_1' y_t^1 \\ x_2'(1 - y_t^2) & x_2' y_t^2 \\ \vdots & \vdots \\ x_n'(1 - y_t^n) & x_n' y_t^n \end{pmatrix}.$$

Then the matrix model for Z conditional on $\{Y_t = y_t, t = 1, \dots, m\}$ is $Z = XB + E$, where $B = I_m \otimes \beta$ with $\beta = (\beta_0', \beta_1')'$, and $n \times m$ matrix of Gaussian errors $E = (\varepsilon(s_i; t) : i = 1, \dots, n; t = 1, \dots, m)$. Here I_m is $m \times m$ identity matrix, and β is $2q \times 1$ vector of parameters. Under covariance separability assumption $\text{vec}(E) \sim N_{nm}(C_T \otimes C_s)$ with $C_T = (c_T^{tr} = C_T(t, r); t, r = 1, \dots, m)$ denoting the $m \times m$ matrix of pure temporal covariances and $C_s = (c_s^{ij} = C_s(s_i, s_j); i, j = 1, \dots, n)$ denoting the $n \times n$ matrix of pure spatial covariances.

In present paper we are dealing with the problem of classification of the observations $Z(s_i, m + 1)$, $i = 1, \dots, n$ into one of two classes with given joint training sample Z or, in other words, based on training sample information we want to predict label at an observed location at the time moment $t = m + 1$. Set $c_T^{m+1, r} = C_T(m + 1, r)$; $r = 1, \dots, m$, $c_T^{m+1} = (c_T^{m+1, 1}, \dots, c_T^{m+1, p})'$ and e_i' – the i th row of identity matrix I_n .

Under spatio-temporal data model specification, we can conclude that in $l = 0, 1$, the conditional distribution of $Z(s_i, m + 1)$ given $Z = z$ and $Y(s_i; m + 1) = l$, is Gaussian, i.e.,

$$(Z(s_i, m + 1) | Z = z; Y(s_i; m + 1) = l) \sim N(\mu_{li(z)}^{m+1}, \Sigma_{m+1, i(z)}), \quad (1)$$

where $\Sigma_{m+1, i(z)} = \text{var}(Z(s_i, m + 1)) - c_s^{ii}(c_T^{m+1})' C_T^{-1} c_T^{m+1} = c_s^{ii} \rho_{m+1}$, $\mu_{li(z)}^{m+1} = \beta_l' x_i + ((c_T^{m+1})' C_T^{-1} \otimes e_i') \text{vec}(E)$ with $\rho_{m+1} = c_T^{m+1, m+1} - (c_T^{m+1})' C_T^{-1} c_T^{m+1}$.

In this study, we assume that the conditional distribution of label $Y(s_i, m + 1)$, $i = 1, \dots, n$, given joint training sample Z depends only on class labels values, i.e. conditional distribution of $(Y(s_i, m + 1) = l | Z = z)$ is identical to conditional distribution of $(Y(s_i, m + 1) = l | \{Y_t = y_t, t = 1, \dots, m\})$. Set $P(Y(s_i, m + 1) = l | Z = z) = \pi_l(s_i, m + 1)$, $l = 0, 1$, and, for simplicity, call them prior class probabilities.

2 Conditional Bayes discriminant functions and its error rates

Under the assumption that the classes are completely specified, the conditional Bayes discriminant function (CBDF) minimizing the probability of misclassification is formed by the log-ratio of conditional likelihood of distribution specified in (1), that is

$$W_Z(Z(s_i, m+1)) = \left(Z(s_i, m+1) - \frac{\mu_{1i(z)}^{m+1} + \mu_{0i(z)}^{m+1}}{2} \right) \Sigma_{m+1, i(z)}^{-1} (\mu_{1i(z)}^{m+1} - \mu_{0i(z)}^{m+1}) + \gamma_i(m+1), \quad (2)$$

where $\gamma_i(m+1) = \ln(\pi_1(s_i, m+1)/\pi_2(s_i, m+1))$.

Let us call the probability of misclassification for $W_Z(Z(s_i, m+1))$ as Local Bayes error rate and denote it by P_i . Also let us denote squared *Mahalanobis distance between conditional distributions* by

$$\Delta_{m+1, i(z)}^2 = (\mu_{1i(z)}^{m+1} - \mu_{0i(z)}^{m+1})' \Sigma_{m+1, i(z)}^{-1} (\mu_{1i(z)}^{m+1} - \mu_{0i(z)}^{m+1}). \quad (3)$$

Specific attention is given to the Gaussian spatio-temporal model with pure spatial covariances belonging to the family of powered-exponential isotropic models and with pure temporal covariance of stationary AR(p) model.

Lemma 1. *The Local Bayes error rate is*

$$P_i = \pi_0(s_i, m+1)\Phi(Q_{0i}) + \pi_1(s_i, m+1)\Phi(Q_{1i}),$$

where $\Phi(x)$ is the standard normal cumulative distribution function and

$$Q_{li} = \frac{\Delta_{m+1, i(z)}}{2} + (-1)^l \frac{\gamma_i(m+1)}{\Delta_{m+1, i(z)}}, \quad \Delta_{m+1, i(z)}^2 = ((\Delta\beta)'x_i)^2 / c_s^{ii} \sigma_T^2, \quad l = 0, 1.$$

Proof. It is known that for AR(p) model parameters quantify the temporal dependency and for $t = 1, 2, \dots, m+1$, $c_T^{t,t} = C_T(0) = \sigma_T^2 + \sum_{j=1}^p \alpha_j C_T(j)$, where σ_T^2 is the variance of the temporal white noise.

Then $\mu_{li(z)}^{m+1} = \beta_l' x_i + ((0, \dots, 0, \alpha_p, \dots, \alpha_1)' \otimes e_l') \text{vec}(E)$ and $\Sigma_{m+1, i(z)} = c_s^{ii} \sigma_T^2$.

By using the properties of multivariate Gaussian distribution and inserting the above expressions in formula (3) we complete the proof of Lemma 1. \square

Error estimation is critical to classification because the validity of the resulting classifier model, composed of the classifier and its error estimate, is based on the accuracy of the error estimation procedure. Given a set of sample data, the data can be split between training and test data, with a classifier being designed on the training data and its error being validated on the test data. In this paper our focus is on using p temporal observations for training and the observations at $m+1$ th time moment is using for testing.

3 Numerical illustrations and simulations

For numerical illustrations of the proposed classifier performance, we considered the Gaussian spatio-temporal model with pure spatial exponential covariances and with

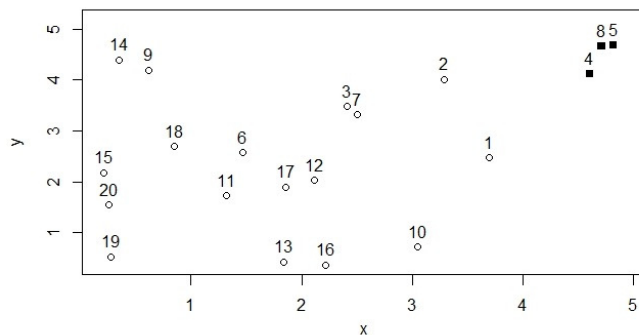


Fig. 1. Spatial sampling set S_{20} at moment $T = 1$.

pure temporal covariance stationary AR(1) model. Then it is easy to derive that $(c_T^{m+1})'C_T^{-1} = (0, \dots, 0, \alpha_1)$.

In the study, the exponential isotropic nuggetless spatial covariance is considered. So $C_s = \sigma_s^2 R$, where $R = (r_{ij})$ denotes the spatial correlation matrix with $r_{ij} = r(|s_i - s_j|) = e^{-|s_i - s_j|/\varphi}$. Here φ is the so called range parameter that represents the spatial dependence.

This choice is based on the smoothness level of sample paths. Sample paths of a GRF with the exponential covariance function are not smooth, when the squared exponential covariance model has smooth sample paths.

Two methods for prior class probabilities is proposed.

First one is based on Temporal Weighted Moving Average (TWMA) method

$$\pi_{1t}(s_i, m+1) = \frac{\sum_{t=1}^m y_i^t t}{(1+m)m/2}.$$

Second one adds spatial correlations for weighting

$$\pi_{1ts}(s_i, m+1) = \frac{\sum_{t=1}^m y_i^t t + \sum_{t=1}^m y_{i_0}^t t r_{ii_0}}{(1+m)m/2 + r_{ii_0}(1+m)m/2},$$

where i_0 denotes the index of the nearest neighbor to s_i . Denote this method by (STWMA). We have compared these four particular cases by calculating the P_i for $i = 1, \dots, n$. Numerical illustrations performed on 20 locations in two dimensional Euclidean area with \circ and \blacksquare for label indicators are depicted in Fig. 1.

The averages $AP = \sum_{i=1}^{20} P_i / 20$ for two models of prior probabilities and for $\phi = 2.5$ and various α_1 at moment $t = 5$ are presented in Fig. 2. As it might be seen from Fig. 2, incorporation of spatial correlation in class prior probabilities (i.e. method STWMA) does not have advantage against method TWMA for almost all values of parameter α_1 .

4 Conclusions

In this paper we propose approach to classification of spatio-temporal data in the framework of Bayes discriminant analysis for separable spatio-temporal covariances.

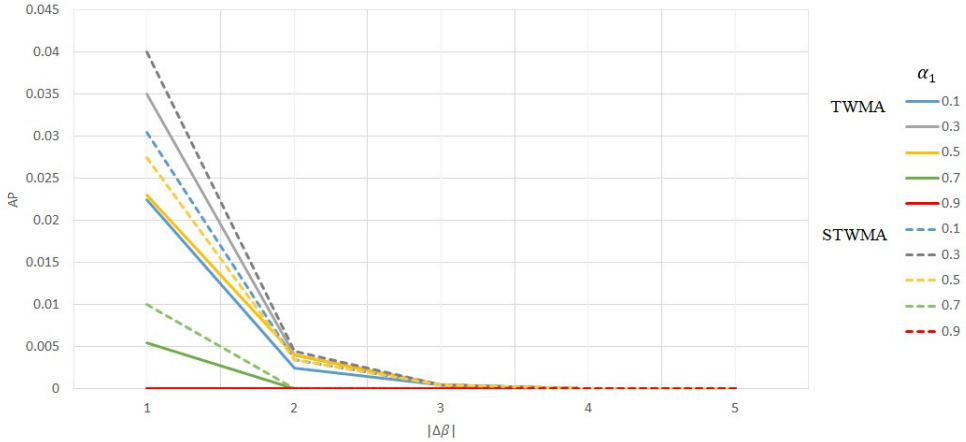


Fig. 2. Average Bayes error rates for $\varphi = 2.5$ and various α_1 at moment $t = 5$.

Several simulation studies were conducted to estimate and compare empirically the classifiers for particular separable stationary spatio-temporal covariance and various prior class probabilities models. Numerical analysis showed that there is no reason to incorporate spatial correlation in the prior probabilities since it does not improve the performance of the proposed classifier.

References

- [1] P.M. Atkinson, P. Lewis. Geostatistical classification for remote sensing: an introduction. *Comput. Geosc.*, **26**(4):361–371, 2000. [https://doi.org/10.1016/S0098-3004\(99\)00117-X](https://doi.org/10.1016/S0098-3004(99)00117-X).
- [2] G. Atluri, A. Karpatne, V. Kumar. Spatio-temporal data mining: a survey of problems and methods. *ACM Comput Surv: CSUR*, **51**(4):83, 2018. <https://doi.org/10.1145/3161602>.
- [3] C. Berrett, C.A. Calder. Bayesian spatial binary classification. *Spatial Stat.*, **16**:72–102, 2016. <https://doi.org/10.1016/j.spasta.2016.01.004>.
- [4] S.S. Demel, J. Du. Spatio-temporal models for some data sets in continuous space and discrete time. *Stat. Sin.*, **25**:81–98, 2015. <https://doi.org/10.5705/ss.2013.223w>.
- [5] K. Dučinskas, L. Dreižienė. Risks of classification of the Gaussian Markov random field observations. *J. Class.*, **35**:422–436, 2018. <https://doi.org/10.1007/s00357-018-9269-7>.
- [6] J. Haslett, A.E. Raftery. Space-time modelling with long memory dependence: Assessing Ireland’s wind power resource. *Appl. Stat.*, **38**:1–50, 1989. <https://doi.org/10.2307/2347679>.
- [7] M. Karaliutė, K. Dučinskas. Classification of gaussian spatio-temporal data with stationary separable covariances. *Nonlinear Anal. Model. Control*, **26**(2):363–374, 2021. <https://doi.org/10.15388/namc.2021.26.22359>.
- [8] J. Šaltytė-Benth, K. Dučinskas. Linear discriminant analysis of multivariate spatial-temporal regressions. *Scand. J. Stat.*, **32**:281–294, 2005.

REZIUOMĖ

Tiesinė diskriminantinė gausinių erdvės-laiko duomenų analizė*M. Karaliūtė, K. Dučinskas*

Nagrinėjama tiesinė diskriminantinė gausinių erdvės-laiko duomenų analizė, naudojant separabilius kovariacijų modelius ir erdvėje bei laike kintančias klasės žymes. Laiko kovariacijai naudojamas AR(p) modelis, o erdvinei kovariacijai naudojamas eksponentinis modelis. Nagrinėjamas klasifikavimo procedūros efektyvumas, kuris vertinamas panaudojant generuotus Gauso atsitiktinių laukų duomenis.

Raktiniai žodžiai: separabili kovariacinė funkcija; AR(p) modelis; Bajeso diskriminantinė funkcija