https://doi.org/10.15388/vu.thesis.520
https://orcid.org/0000-0001-7296-9521

VILNIUS UNIVERSITY

Marta Karaliutė

Supervised Bayesian classification methods of Gaussian Spatio-temporal data based on generative machine learning models

DOCTORAL DISSERTATION

Natural Sciences,
Informatics (N 009)

VILNIUS 2023

This dissertation was prepared between 2017 and 2022 at Vilnius University.

Academic supervisor – Prof. Dr. Kęstutis Dučinskas (Vilnius University, Natural Sciences, Informatics, N 009).
Academic consultant – Prof. Habil. Dr. Gintautas Dzemyda (Vilnius University, Natural Sciences, Informatics, N 009).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:
Chair – **Prof. Dr. Audronė Jakaitienė** (Vilnius University, Natural Sciences, Informatics, N 009).
Members:
Prof. Dr. Igoris Belovas (Vilnius University, Natural Sciences, Informatics, N 009),
Prof. Dr. Jurgita Markevičiūtė (Vilnius University, Natural Sciences, Mathematics, N 001),
Prof. Dr. Dalius Navakauskas (Vilnius Gediminas Technical University, Natural Sciences, Informatics, N 009),
Assoc. Prof. Dr. Tatiana Tchemisova (University of Aveiro, Portugal, Natural Sciences, Mathematics, N 001).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 1:00 p.m. on the 29th of September, 2023 in Room 203 of the Institute of Data Science and Digital Technologies of Vilnius University. Address: Akademijos str. 4, LT-04812, Vilnius, Lithuania.

The text of this dissertation can be accessed at the Library of Vilnius University and on the website of Vilnius University:
www.vu.lt/lt/naujienos/ivykiu-kalendorius.

https://doi.org/10.15388/vu.thesis.520
https://orcid.org/0000-0001-7296-9521

VILNIAUS UNIVERSITETAS

Marta Karaliutė

Erdvės-laiko Gausinių duomenų prižiūrimo Bajesinio klasifikavimo metodai, pagrįsti generatyviniais mašininio mokymosi modeliais

DAKTARO DISERTACIJA

Gamtos mokslai,
informatika (N 009)

VILNIUS 2023

Disertacija rengta 2017 – 2022 metais Vilniaus universitete.

Mokslinis vadovas:
prof. dr. Kęstutis Dučinskas (Vilniaus universitetas, gamtos mokslai, informatika, N 009).
Mokslinis konsultantas – prof. habil. dr. Gintautas Dzemyda (Vilniaus universitetas, gamtos mokslai, informatika, N 009).

Gynimo taryba:
Pirmininkė – **prof. dr. Audronė Jakaitienė** (Vilniaus universitetas, gamtos mokslai, informatika, N 009).
Nariai:
prof. dr. Igoris Belovas (Vilniaus universitetas, gamtos mokslai, informatika, N 009),
prof. dr. Jurgita Markevičiūtė (Vilniaus universitetas, gamtos mokslai, matematika, N 001),
prof. dr. Dalius Navakauskas (Vilniaus Gedimino technikos universitetas, gamtos mokslai, informatika, N 009),
doc. dr. Tatiana Tchemisova (Aveiro universitetas, Portugalija, gamtos mokslai, matematika, N 001).

Disertacija ginama viešame Gynimo tarybos posėdyje 2023 m. rugsėjo mėn. 29 d. 13 val. Vilniaus universiteto Duomenų mokslo ir skaitmeninių technologijų instituto 203 auditorijoje. Adresas: Akademijos g. 4, LT-04812 Vilnius, Lietuva.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir Vilniaus universiteto interneto svetainėje adresu: www.vu.lt/lt/naujienos/ivykiu-kalendorius.

Contents

Acronyms

ACC Accuracy rate. 37, 50, 51, 71, 73–75, 83, 85

ACE Acute cardiovascular event. 80, 81, 83, 86

AR Autoregressive model. 18, 19, 21, 27, 29, 32, 51, 54, 60, 63, 64, 86, 87

AU Areal unit. 30, 41–43, 47, 49, 54, 58, 59, 71, 73, 83, 85

BAC Balanced accuracy rate. 38, 50, 52, 71, 73–75, 83, 85, 86

BCR Bayes classification rule. 15, 28, 46

BDF Bayes discriminant function. 18, 20, 21, 27, 35, 39, 46–51, 56, 57

CBDF Conditional BDF. 55, 62

CI Conditional Independence. 33, 44

CSD Diseases of the circulatory system. 80, 81, 84, 86

CSW Complete spatial weighting. 48, 49, 51, 71, 73, 83, 84, 86

GAC Geometric accuracy rate. 38, 52, 71, 73, 83, 85

GD Geostatistical Dependence. 33, 61

GGM Gaussian Geostatistical Model. 18, 19, 21, 32, 60, 63, 67, 75, 86, 87

GHMM Gaussian Hidden Markov Model. 18–21, 26, 31, 43, 46, 49, 51, 67, 71, 73, 83, 86, 87

GMRF Gaussian Markov Random Fields. 18–21, 27, 28, 32, 51, 52, 67, 73, 86, 87

GRF Gaussian Random Field. 16, 18, 21, 27–31, 61, 67, 76

HCAR Homogeneous conditionally autoregressive model. 52–55

HMM Hidden Markov model. 15–17, 20, 41–44, 46, 47, 49, 52, 58

MC1 First-order Markov Chain. 18, 26, 42, 43

MD Markov Dependence. 33, 43

ML Maximum Likelihood. 17, 19, 29, 35, 47, 53, 56, 57, 64, 65

MRF Markov Random Field. 25

OBC Observation to be classified. 19, 21, 28, 31, 32, 34, 36, 38, 40, 50, 60

OSA One-step-ahead. 16, 17, 20, 21, 25, 29, 34, 40, 46, 52, 67

PBDF Plug-in BDF. 18, 21, 35, 36, 38, 39, 49–51, 57, 59, 64–66, 71, 74, 75, 83, 85

PDF Probability density function. 34, 43

PSW Partial spatial weighting. 48, 49, 51, 71, 73, 83, 84, 86

RF Random field. 14, 25, 26, 28, 29, 61

ROC Receiver Operator Characteristic. 38

SI Statistical Independence. 33, 52

ST Spatio-temporal. 15–21, 24–35, 39–42, 44, 46, 49–53, 57, 60–63, 65, 67, 76, 80, 84, 86–88

STUA Spatial Temporal Unweighted Average. 65, 66

STWMA Spatial Temporal Weighted Moving Average. 65, 76, 77, 80, 86

SW Spatial weighting. 31, 46

TA Temporal averaging. 32, 52, 56, 64

TWMA Temporal Weighted Moving Average. 65, 76, 77, 80, 86

Notation

A'	transpose of matrix A.
$P(A)$	probability distribution of random quantity A.
$P(A\|B)$	probability distribution of random quantity A conditional on specifying a particular value of random quantity B.
$H(x)$	Heaviside step function.
$Ber(p)$	Bernoulli distribution with parameter p, $0 < p < 1$.
I_n	$n \times n$ identity matrix.
$N(\mu, \sigma^2)$	univariate normal (Gaussian) distribution with mean μ and variance σ^2.
$\Phi(x)$	distribution function for $N(0, 1)$.
$\varphi(x\|\mu, \sigma^2)$	probability density function for $N(\mu, \sigma^2)$.
$N_m(\mu, \Sigma)$	multivariate normal distribution with mean vector μ and covariance matrix Σ.
$N_{n\times p}(M, U, V)$	matrix-variate normal distribution, mean M is $n \times p$, Variance U is $n \times n$ (among-row) and V is $p \times p$ (among-column).
$vec(A)$	vectorization of a matrix A.
$\propto$	the proportionality symbol.
$card(A)$	cardinality of a set A.
$AR(p)$	an autoregressive model of order p.
iid	independent and identically distributed.
$Pr\Big(Y_{T+1}^{(i)} = l, Y\Big)$	the joint distribution for class labels.
$p(z\|y)$	conditional PDF of feature observations given class labels in training sample.

List of Figures

List of Tables

INTRODUCTION

Classification is a task that requires the use of statistical or machine learning algorithms that help to identify the category of new observations based on the information captured from the training data. Machine learning algorithms with the fully labeled data are attributed to a supervised classification category. Two main models of supervised classification are generative and discriminative [4]. Formally we write Z as a vector of feature observations and Y as a vector of class label observations. In general, discriminative models aim to find a decision boundary between different classes during the learning process. Precisely, the goal is to model the posterior distribution $P(Y|Z)$ directly, or identify a direct mapping of the inputs Z to Y. Some of the most popular methods are logistic regression, conditional Random field (RF), Support Vector Machine, Neural Networks, Random Forest and others [89]. It should be noted that deep learning methods, based on a discriminative model, have been intensively studied now [41, 100–102, 106], but they are off scope of this thesis.

The generative models of a supervised classification attempt to include the information on feature conditional distributions, and use a joint distribution $P(Z, Y)$ to generate the new data similar to the existing ones.

Given a new feature observation, it predicts which class (label value) would most likely have generated the given observation.

There are several benefits of using generative models. Some of them are listed below:

- Firstly, they can help us create the new data. For instance, if we need to create a set of data for testing purposes, we can use a generative model to generate the data. This can save us time and effort, and it can also help ensure the accuracy of our tests.

- Secondly, they can be used to generate new statistically possible dependent samples with algorithms that converge faster than discriminative model algorithms. That is highly valuable in cases with small and moderate training data sets.

- Finally, generative models can be used to improve machine learning algorithms in the following sense. By using a generative model to generate data, we can train our machine learning algorithms in a more realistic way. This can lead to a better performance and more accurate results.

These entails motivation to focus attention only on the generative models to a supervised Bayesian classification for parsimoniously specified and clearly detected Spatio-temporal (ST) data model types.

Mathematically, generative models attempt to learn the joint probability distribution expressed by $P(Z, Y)$. Instead of trying to model the complicated joint distribution, it is factored into a product of conditional feature distributions $P(Z|Y)$ and the unconditional class label distribution $P(Y)$.

In terms of this notation, a posterior class label distribution according to Bayes' theorem can be written as

$$P(Y|Z) = P(Z, Y)/P(Z) = P(Z|Y)P(Y)/P(Z)$$

or $P(Y|Z) \propto P(Z|Y)P(Y)$

since $P(Z)$ does dot depend on class label Y.

Therefore, Bayes classification rule (BCR) directly depends only on the product of the conditional feature distribution $P(Z|Y)$ and the unconditional class label distribution $P(Y)$. So the forms of these two distributions completely specify the data model types.

Some of the best-known generative methods are Naïve Bayes, Hidden Markov model (HMM) and Gaussian discriminant analysis [22, 61, 108], that relies on various statistical independence assumptions.

Compared to the general classification problem, the classification of ST data must take account the location specific information of the data and the interaction between feature and label variables at each point in time. It is known that spatial and temporal context is commonly used in image classification, indicating the relationship between a classified pixel and its neighbouring pixels is incorporated into analyses. A detailed comparison of the ST contextual classification methods for remote sensing image classification is provided [49]. Contextual classification models that exploit spatial information by quantifying region spatial relationships can be used for image classification and object detection [83, 86].

In the thesis, the novel methods and techniques for using generative models of supervised Bayesian classification to data with different types of ST data model are developed. Three types ST data models for feature and class label with discrete time index set and discrete or continuous spatio index sets and relied on various statistical dependence forms (from independence to dependence) are explored. One is based on HMM with sequential Gaussian feature observations at fixed number of spatial sites. Other two are based on Gaussian Random Field (GRF) models for feature with separable covariance functions. Neighbourhoods are specified taking into account both spatial and temporal dimensions.

Precisely, it is assumed that the conditional distribution of $P(Z|Y)$ belongs to the family of Gaussian distributions and the unconditional label Y distribution belongs to the family of discrete distributions with Bernoulli marginals. The parameters of both distributions may depend on the spatial and temporal context of the observations.

Considered ST data models can be prescribed to dynamical ones since they are concerned on the solution of problems how current process values have evolved from past process values (see e.g. [15]).

This study focuses on the straightforward procedures of One-step-ahead (OSA) temporal classification (prediction) in the context of high dimensional spatial information expressed by spatial trends and spatial dependence measures (spatial weights, spatial association indices, Euclidean distances, . . .).

However, for the brevity, the word "dynamical" will be skipped in the rest of the thesis.

Within each type the models the ability to correctly classify feature observation in the prescribed set of spatial sites as well as their sensitivity to the amount of incorporated ST information are studied. Numerical analysis is conducted with simulated and real data sets. The accuracy of proposed classifiers are evaluated by using several empirical and probabilistic performance measures. Existing literature suggests various ways to exploit available ST data however, these methods are either good at utilizing spatial information or temporal information but not both. We instead propose a novel algorithms for supervised OSA classification (see e.g. [92]) based on method that combines both spatial as well as temporal analysis.

All explored ST data models rely on the moderate number of parameters that often can be estimated in the closed form by Maximum Likelihood (ML) method. That ensure the low computational costs, competitiveness among previously studied ST data classifiers and widening of the application area.

Statement of the Problem

ST data are frequently occurring in modern science and engineering, but they are often so complicated and high-dimensional that we can hardly extract useful information from them directly and summarize characteristics across time and space. It is known that even the Gaussian models with general parametrical ST trends and covariances (variogram) are intractable, that aligns high computational costs.

In order to overcome the computational complexity, various assumptions on feature and label distributions are introduced. In many applications, spatial and ST model relying on independence or conditional independence assumption for feature observations are explored. Naïve Bayes and HMMs their modifications are the most popular ones but they are characterized on data structures that do not involve spatial context sufficiently.

In the thesis, the novel approach for integration of spatial and temporal context in supervised Bayesian classification of several types of Gaussian data is proposed.

It is focused on three types of discrete time (or sequential) spatial Gaussian data models comprising various assumptions on feature and class label distributions and statistical dependence forms in spatial and temporal senses.

These distributional assumptions ensures simplicity in construction of Bayesian OSA classifiers with moderate number of easily estimable model parameters. That allow to simplify the solution of optimization tasks, to overcome intractability of likelihood functions and widen the application area of proposed methods.

That stipulates the motivation to propose and investigate more such generative models for supervised Bayesian classification of ST data, that reduces computational difficulty and has considerable application area without significant losses in classification accuracy.

Research Object

Novel competitive, computationally efficient methods and algorithms to generative supervised classification of ST lattice and geostatistical data, Gaussian distribution for feature and Bernoulli distribution for class label, Bayes discriminant function (BDF) and Plug-in BDF (PBDF), spatial Gaussian Hidden Markov Model (GHMM), GMRF and Gaussian Geostatistical Model (GGM), time series Autoregressive model (AR).

Research Aim and Objectives

The research aims to develop novel supervised Bayesian classification methods of Gaussian ST data based on generative machine learning models. At the same time this research compliments collection of the existing analogous classifiers of data in the purely spatial context. Focusing on three popular types of ST data models with feature variable following Gaussian distribution and class label variable following Bernoulli distribution the tasks related to parsimony of model parametrization, to simplification in parameter estimation process, to sensitivity analysis of classifiers performance due to changes in the data model structures are emphasized.

Performance of the proposed classification procedures are evaluated for simulated and real ST data.

The first type of data model includes GHMM for spatial data consisting of conditionally independent feature observations given class labels, where the label represents homogeneous First-order Markov Chain (MC1) at each site. Next two types of ST data models for feature comprises spatial lattice (areal) data model specified by GMRF and geostatistical data models specified by GRF and stationary AR model in temporal context. (see e.g. [15, 73]). Hence, these models are imposed by the assumption of separability for ST covariance function with spatial and temporal factors (see e.g. [15, 34, 54]). As it follows, refer to these models as GMRF-AR and GGM-AR. Class label observations are assumed to be independent Bernoulli random variables with parameter being deterministic function of ST context (spatial covariates, spatial weights, time series parameters, . . .).

All three types of data models are used to originally define the relationship and dependence within feature and class label observation in training sample, and relationship between observation to be classified Observation to be classified (OBC) and observations from labeled training data set.

For this aim, the following objectives should be achieved:

1. Develop the novel method and construct an algorithm for classification of spatial GHMM on lattice (areal) data by suggesting the original strategy for deriving synthetic estimators relied on spatial weighting. Detailed comparison of competing classifiers relied on different parameter estimation rules is done in terms of several performance measures.
2. Develop the novel method and construct an algorithm for supervised Bayesian classification of lattice data generated by GMRF-AR and independent Bernoulli variables for class labels with distribution parameters depending on various spatial association indices.
3. Develop the novel method and construct an algorithm for supervised Bayesian classification of GGM-AR and independent Bernoulli variables for labels with distribution parameter depending on Euclidean distances and temporal lags.
4. Evaluate and compare the performance of proposed classifiers within every type of data models specified by different scenarios of incorporated spatial and temporal information.

Research Methods

The proposed classification method is based on the ratio of the logarithms of the univariate conditional likelihood for OBC at each site given a labeled training sample. The performance of classifiers is evaluated by several performance measures using confusion matrices and conditional distributions of OBC.

In empirical research, the proposed classification algorithms are realized with simulated and real data. The data model parameters are estimated using ML estimators from labeled training sample. Within each type of data models, the detailed comparison of the proposed classifiers is performed for various feature and class label distributions.

Scientific Contributions and Practical Value of the Research

This work contributes to the development of novel methods and algorithms to application of a generative models to the supervised Bayesian classification of ST data. Novel scenarios for incorporation the contextual information (statistical and deterministic) into several parametric ST data models are developed.

The novelty is revealed in ST models specification as well as in methods of parameter estimation and classifiers performance evaluation. It should be also noted that explored ST data models rely on the moderate number of parameters that often can be estimated in the closed form by maximum likelihood method. That ensure the low computational costs, competitiveness among previously studied ST data classifiers and considerable application area.

Proposed method for OSA classification is realized on three parametric data model types by simulation study and real data examples. Classifiers based on different levels of incorporated ST information into BDF are compared by several performance (accuracy) measures.

The main contributions of this work can be outlined as follows:

1. Proposed parameter estimation strategies and rules leads to the closed form for the estimators that alleviates the computational difficulties related to considered classification problems solution.
2. The experimental results showed that classifiers based on GHMM with higher spatial weighting level in majority cases have an advantage in ST consistency and classification accuracy over one with lower spatial weighting level. Hence the proposed methodology can be considered as valuable extension to existing spatial HMM with continuous feature and constant transition probabilities.
3. Introduction of Bayesian classifiers for ST lattice data modeled by GMRF with separable covariances for features and by independent Bernoulli variations for class labels with distribution parameters depending on three popular spatial association indices.

4. Proposition the algorithms for the solution of the Bayesian classification problem for ST geostatistical data modeled by GRF with separable covariances for features and by independent Bernoulli variations for class labels with parameters depending on the ST deterministic context of the OBC.

5. Derived closed-form expressions for local error rates and their D-estimators incurred by BDFs and PBDFs allows effective evaluation of the classifiers performance before having the realization of OBC is observed.

Defensive Claims

The following claims are defended in this thesis:

1. For classifiers based on GHMM:
   (a) transition probability estimators based on label persistence rate leads to classifier with higher performance than rival method based on transition rates;
   (b) complete spatial weighting level applied to spatial GHMM parameters estimation ensures the higher OSA classification accuracy than partial spatial weighting level.

2. For classifiers based on GMRF and AR, the influence of the implementation of Moran's I, Geary's C and Getis-Ord G indices does not differ significantly in proposed performance measures.

3. The performance of classifiers using GGM, significantly depends on the form of the spatial covariances function and information incorporated in class labels distribution.

4. Derived analytical expressions for the numerous parameter estimators reduces the intractability for the considered data model types.

Approbation

The research results have been published in 6 papers: three papers in periodic scientific journals indexed by the Web of Science, one article in a periodic scientific journal, and two papers in reviewed scientific conference proceedings. The results were presented at scientific conferences. The following list presents the publications and presentations in conferences:

Papers in periodic scientific journals indexed in Web of Science and Scopus database:

1. Marta Karaliutė, Kęstutis Dučinskas. Classification of Gaussian spatio-temporal data with stationary separable covariances. Nonlinear analysis: modelling and control. Vilnius 26 (2), 2021, p. 363-374. DOI: 10.15388/namc.2021.26.22359.

2. Kęstutis Dučinskas, Marta Karaliutė, Laura Šaltytė-Vaisiauskė. Spatially Weighted Bayesian Classification of Spatio-Temporal Areal Data Based on Gaussian-Hidden Markov Models. Mathematics 11(2), 2023, 347. https://doi.org/10.3390/math11020347

3. Marta Karaliutė, Kęstutis Dučinskas. Performance of the supervised generative classifiers of spatio-temporal areal data using various spatial autocorrelation indexes, Nonlinear Analysis: Modelling and Control, 28(2), 2023, p. 1-14. doi: 10.15388/namc.2023.28.31434.

Paper in a periodic scientific journal:

1. Marta Karaliutė, Kęstutis Dučinskas. Supervised linear classification of Gaussian ST data. Lietuvos matematikos rinkinys, 62(A), 2021, pp. 9-15. DOI: 10.15388/LMR.2021.25214

Papers in peer-reviewed scientific conference proceedings:

1. Marta Karaliutė, Kęstutis Dučinskas; Laura Šaltytė-Vaisiauskė. Expected error regret in linear discrimination of balanced spatial Gaussian time series. DAMSS 2018: 10th international workshop on „Data analysis methods for software systems“: Druskininkai, Lithuania, November 29 - December 1, 2018, p. 42 [abstract book].

2. Marta Karaliutė, Kęstutis Dučinskas; Laura Šaltytė-Vaisiauskė. Expected error rate in linear discrimination of balanced spatial Gaussian time series. Computer data analysis and modeling: stochastics and data science: proceedings of the XII international conference: Minsk, September 18-22, 2019, p. 172-175. http://elib.bsu.by/bitstream/123456789/233358/1/172-175.pdf

Presentations in scientific conferences:

1. DAMSS 2018 (Druskininkai, Lithuania). Expected error regret in linear discrimination of balanced spatial Gaussian time series. (poster presentation). 10th international workshop on „Data analysis methods for software systems“. November 29 - December 1, 2018.

2. Spatial Statistics 2019 (Sitges, Spain). Linear discriminant analysis of spatial temporal unemployment rate data in Lithuania. (poster presentation). Towards Spatial Data Science. July 10-13, 2019.

3. Computer data analysis and modeling 2019 (Minsk, Belarus). Expected Error Rate In Linear Discrimination of Balanced Spatial Gaussian Time Series (oral presentation). Computer data analysis and modeling: stochastics and data science: proceedings of the XII international conference: Minsk, September 18-22, 2019.

4. LMD 2021 (Vilnius, Lithuania). Tiesinė diskriminantinė erdvės-laiko duomenų analizė (oral presentation). Lietuvos Matematikų Draugijos LXII Konferencija. 2021 m. June 16-17 d.

5. KoDi21 (Klaipėda, Lithuania). Prižiūrima vėjo greičio erdvės ir laiko duomenų klasifikacija (oral presentation). „Kompiuterininkų dienos 2021“, September 23-24, 2021.

6. LJMS'10 (online, Vilnius, Lithuania). Erdvės-laiko duomenų klasifikavimas naudojant diskriminantines funkcijas (oral presentation). December 28, 2021.

7. LMD 2022 (Kaunas, Lithuania). Tiesinė diskriminantinė Gauso erdvės-laiko teritorinių duomenų analizė (oral presentation). Lietuvos Matematikų Draugijos LXIII Konferencija. June 16-17, 2022.

8. IBC 2022 (Riga, Latvia). Supervised generative model for classification of Gaussian ST areal data (poster presentation). Participated in the 31st International Biometric Conference, 10-15 July, 2022.

9. Spatial Statistics 2023 (Boulder, USA). Bayesian classification of spatio-temporal lattice data based on Gaussian Hidden Markov models (poster presentation). Climate and the Environment. July 18-21, 2023.

Structure of the thesis

This doctoral thesis consists of the introduction, three chapters, conclusions, appendix and bibliography. The first chapter is designated for description of the considered ST models and generative classification method applied to them. The second chapter includes the issues of modelling spatial data, discusses the estimators for spatial model. Here the main results of the thesis concerned with discriminant analysis of ST data is presented. The third chapter introduces the numerical experiments with simulated and real data.

Finally the general conclusion, future research directions and references are presented. 112 bibliographic references are included at the end of the thesis. The dissertation consists of 144 pages, 17 figures and 24 tables.

Chapter 1

GENERATIVE MODEL TO CLASSIFICATION FOR SPATIO-TEMPORAL DATA

1.1 Review on related works on Spatio-temporal data modeling and classification

Literature review is performed to evaluate the methods used to model feature values with statistical ST contextual information, as well as to evaluate the generative models used to solve the supervised classification problems for models with spatial contextual information. The inclusion of ST contextual information in the distribution of feature and class labels is defined and used for the OSA temporal classification at each location or areal unit.

In the environmental agricultural and other research, data are often collected across space and sequentially over time. Many problems in ecology and the environmental sciences, such as monitoring the presence/absence of a species, involve the observation of spatial binary RFs for the representation of class labels. It is becoming more common for these studies to include a time component as well. Markov Random Field (MRF) models can be modified to incorporate temporal dependence whether the dependence is on a local level or through a global impact. However, it is important when working with MRF models to ensure that the spatial dependence is properly specified. The ST data are usually recorded at regular time intervals (time lags) and at irregular stations (areas) in a compact area (see e.g. [1, 19, 39, 48, 67, 75, 96]).

The method to ST modeling through the stochastic partial differential equation is developed by Cameleti and et al. [12].

Modeling and prediction of this data type has been studied by multiple authors (see e.g. [8, 15, 33, 67, 76, 93, 111]). Some computational aspects in prediction of ST RFs are discussed by De Iaco and Posa [16]. Often before analyzing ST data sets, ST discretization (or aggregation) is applied. The discretization is useful to summarize information and helps to extract features within a ST range rather than measuring a single point [36]. Recently, several R-packages providing tools for modeling of ST data have been developed (see e.g. [7, 68, 97]). ST data mining has broad applications in various fields including environment and climate, health-care (e.g. wind prediction, precipitation forecast, geographical disease pattern study), public safety (e.g. crime prediction) and intelligent transportation (e.g. traffic flow prediction).

There are several ways to evaluate spatial and temporal correlation. One way is to extend spatial models fairly straight-forwardly by taking temporal correlation into account. For example, if we have lattice data and each data point has a set of spatial neighbours, we can define the set of neighbours of a site to be the union of the set of its spatial neighbours and its temporal neighbours. Another way is based on incorporation spatial variability through location-specific covariates and spatial weights, distances and so on. Temporal variability is incorporated by imposing Markovian dependence properties to feature and class label distributions. Spatial and temporal correlations for lattice models depends on specified spatial and temporal neighbourhood based on lags systems. Both cases are considered in this thesis.

ST data classification being important part of data mining is becoming more important in the big data era with the increasing availability and importance of large ST data sets such as maps, virtual globes, remote sensing imagery, the decennial census and GPS trajectories. Recently, deep learning methods via convolutional natural networks have been intensively explored and used in ST data and image analysis (see e.g. [3, 94, 102–105]).

This study discusses three types of ST data models relying on regular time series corresponding to different spatial site. The first type namely the spatial GHMM is suited for applications with conditionally independent feature Gaussian observations at each site and MC1 for label. This method is developed by Dučinskas et al. [28]. The next two types relied on lattice data models generated by GMRFs and the geostatistical model based on GRF with separable covariance. These models can realistically be performed when feature observations at a given site follow the AR model (see e.g. [9]).

In this thesis, we will focus largely on lattice data structures but also somewhat on geostatistical point-level data. Again, to contrast lattice and geostatistical data, we are not concerned with the possibility of a measurement in between adjacent sites.

Next, models that aim to describe relationships between variables with a ST reference and discuss the general class of dynamic space-time models in the framework presented by Xu and Wikle [98], Wikle and Hooten [95].

In geostatistics, Gaussian process is a powerful tool for characterizing and predicting such ST dynamics, for which the specification of a ST covariance function is the key. While the extant literature offers a wide range of choices for flexible stationary ST covariance models, the temporally evolving ST dependence has received scant attention only. To this end, we propose a time-varying ST covariance model for describing the time-evolving ST dependence. Each model class is introduced through a motivating application.

Compared to the general classification problem, the classification of ST data needs to consider the location information of the data and the interaction among feature and label variables at current and previous time moments.

A comprehensive literature survey on state-of-the-art advances in ST data mining is proposed by Hamdi et al. [38]. Systematic review of methods in spatial deep learning is reported by Mishra et al. [59].

There are currently a variety of ways to achieve the classification goal, but one of the most effective is to use the generative classifiers based on BDFs. Spatial supervised classification is a problem of labeling observations based on feature information and information about spatial adjacency relationships with the training sample. Switzer [87] was the first to treat classification of spatial data. Mardia [56] and Atkinson and Lewis [2] reviewed geostatistical techniques for classification of remotely sensed images. De Oliveira [17] proposed spatial classification techniques based clipping of GRF. Spatial contextual classification problems arising in geospatial domain is considered by Shekhar et al. [79]. It is usually assumed that feature observations conditional on labels are independent (conditional independence) and normally distributed and the labels follow the RF model. This method is widely used in the image classification by Nishii and Eguchi [64]. Dučinskas [23, 24] proposed and explored BCRs for spatial Gaussian data by avoiding the assumption of conditional independence.

The generative model to supervised classification of purely spatial data is studied by numerous authors (see e.g. [5, 23, 30, 46, 50, 60, 61, 70, 78, 90, 112]). The method to Bayesian classification of GMRF observation on the lattice has been developed by Dreižienė and Dučinskas [21, 25, 26]. However, in these works only the purely spatial context was considered.

It should be noted that Šaltytė-Benth and Dučinskas [110] were first to treat the supervised classification of ST data modeled by GRF in the very specific case when OBC is uncorrelated with the training sample. Later, the generative models for supervised classification of ST data have been successfully used in land-cover [14, 88] and injury crashes in road [42], but then only considered the cases with discrete features. For continuous feature ST models Spezia et al. [82] took into account spatial weights only for the class labels. Gong et al. [37] proposed the method relied on surplus parametization of spatial weights, so increasing and so large number of estimable parameters.

The generative classification method for geostatistical and lattice ST Gaussian models is introduced by Karaliutė and Dučinskas [43, 44] and Dučinskas et al. [28].

In the thesis, the novel methods and techniques of implementation the generative models to supervised Bayesian classification to different types of ST Gaussian data are developed. It can be named as extension of similar problem from the spatial setting to the ST setting. The novelty is revealed in the following aspects of classifiers construction processes:

1. inclusion of the variety of statistical dependence forms in spatial and/or temporal senses allowing the class label to vary in time for each spatial site;
2. proposition of two original strategies for the estimation of data model parameters;
3. proposition of the original OBC distribution parameter estimation rules for every type of data model;
4. introducing original performance evaluation measures by capturing both global and local ST aspects in modeling.

Considered ST data models rely on the moderate number of parameters that often can be estimated in the closed form by ML method. That ensure the low computational costs, competitiveness among previously studied ST data classifiers and considerable application area.

To incorporate temporal variation, we allow the class labels to change through time and the features to follow AR models. The model is cast in a general framework which allows us to include various types of ST information, and permits fast implementation and full probabilistic inference for the parameters, and OSA classification. To illustrate our methods, we apply it to large data sets from the health sciences: annual mortality rates in Lithuanian regions.

1.2 Spatio-temporal data models

Three types of ST data models for feature and class label with discrete time index set and discrete or continuous spatio index sets, and various statistical dependence forms are explored. In general, ST data consist of two components that generate the quantitative feature RF Z and the qualitative class label RF Y.

Precisely, ST data model are specified by two RFs: GRF representing feature of $\{Z(s;t) : s \in S \subset R^2, t \in D_T\}$ and RF $\{Y(s;t) : s \in S \subset R^2, t \in D_T\}$ that represents class label and takes only value 0 or 1. Usually its marginals follow the Bernoulli distribution $Ber(p)$, with parameter p possibly depending on ST context. Hence, we focus only on the so called binary classification problem.

It should be noted that GRF has a dominant role in spatial statistics and especially in the traditional domain of geostatistics [13, 15].

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 $\mu_l(s;t)$ is deterministic ST mean and $\varepsilon(s;t) \sim N(0, \sigma^2(s,t))$ for all locations $s \in S$ and discrete time index $t \in D_T = \{1, 2, \ldots\}$.

Since the time index is discrete, the ST GRFs $Y(\cdot,\cdot)$, $Z(\cdot,\cdot)$ can be written as $\{Y_t(\cdot) : t = 1, 2, \ldots\}$ and $\{Z_t(\cdot) : t = 1, 2, \ldots\}$ whereas for $t = 1, 2, \ldots$ $Y_t(\cdot) \equiv \{Y(s;t) : s \in S\}$, $Z_t(\cdot) \equiv \{Z(s;t) : s \in S\}$.

This setting of ST process is traditionally called as time series of spatial processes.

It is also assumed that the set of n sites (locations or nonoverlapping Areal unit (AU)s) $S_n = \{s_i \in S; i = 1, \dots, n\}$ are chosen under deterministic spatial sampling design, and all analysis are carried out conditionally on it. In the subsections that follow, the discrete spatial index as well as continuous spatial index will be considered. Notice, that if discrete spatial index means that we are interested only on fixed n spatial sites from S, than in continuous spatial index case n sites can continuously vary in S.

Hence, the ST GRFs $Y(\cdot, \cdot)$, $Z(\cdot, \cdot)$ can be written as n time series, i.e. $Y_{\cdot}^{(1)}, Y_{\cdot}^{(2)}, \dots, Y_{\cdot}^{(n)}$ and $Z_{\cdot}^{(1)}, Z_{\cdot}^{(2)}, \dots, Z_{\cdot}^{(n)}$, where $Y_{\cdot}^{(i)} \equiv \{Y(s_i; t) : t = 1, 2, \dots\}$, $Z_{\cdot}^{(i)} \equiv \{Z(s_i; t) : t = 1, 2, \dots\}$, $i = 1, 2, \dots, n$.

In this thesis method for finding the joint model of n time series comprises the attempts to properly model of separate time series, quantify the dependencies between time series, and to construct classifiers of one series using data from the other series.

It is endowed with a neighborhood system $N = \{N_i : i = 1, \dots, n\}$, where N_i denotes the collection of sites that have "direct dependence" with site s_i. Let $card(N_i)$ denote number of elements in the set N_i. Throughout thesis it is assumed that N_i is fixed, i.e. does not depend on t for $i = 1, \dots, n$ within considered types of ST data models.

For lattice (areal data) as well for geostatistical data, the neighborhood N_i is defined to be those sites with which s_i share features such as administrative boundaries, borders or various proximity measures. This neighborhood system is key in determining spatial weights and spatial dependence forms.

Specify n dimensional quadratic symmetrical matrix of spatial weights $W = (w_{ij} : i, j = 1, \dots, n)$, where w_{ij} represents the degree of spatial relationship between s_i and s_j and takes the positive value if $s_j \in N_i$ and value 0 otherwise. This specification of the spatial dependence form is natural when modeling lattice data, since similarity between sites often depends on some of the sites shared features, such as boundaries, proximity or similarity of explanatory variables. Spatial weight matrix W usually is replaced by matrix of row standardized spatial weights, i.e. $W^* = (w^*_{ij} : i, j = 1, \dots, n)$ where $w^*_{ij} = w_{ij} / \big(\sum\limits_{k=1}^{n} w_{ik} \big)$.

For geostatistical data, spatial weights are usually defined as some parametric functions of distances between locations.

It should be noted that spatial dependencies are represented by graphs as well as covariance functions and variograms. In thesis both methods are considered.

It is assumed that data is recorded for each site at consecutive time moments $t = 1, 2, 3, \dots$ with regular intervals, i.e. $D_T = \{1, 2, \dots\}$. Temporal neighbourhood is specified *by lags*, as for time series and is constant for all sites.

In real data study we follow the popular tradition in medicine and health care research when the data are recorded at regular time intervals (time lags) and at fixed set of sites (see e.g. [3, 39]).

As it follows, we use the following notation $Y(s_i, t) = Y_t^{(i)}$ and $Z(s_i, t) = Z_t^{(i)}$, $Z_t = \left(Z_t^{(1)}, \dots, Z_t^{(n)}\right)'$, $Y_t = \left(Y_t^{(1)}, \dots, Y_t^{(n)}\right)'$, $Z^{(i)} = \left(Z_1^{(i)}, \dots, Z_T^{(i)}\right)'$, $Y^{(i)} = \left(Y_1^{(i)}, \dots, Y_T^{(i)}\right)'$.

Denote by $z_t = \left(z_t^{(1)}, \dots, z_t^{(n)}\right)$ and $y_t = \left(y_t^{(1)}, \dots, y_t^{(n)}\right)$ the realized value of Z_t and Y_t, respectively. Assume that training data consist of T temporal feature and class label observations in S.

They are specified by two $n \times T$ matrices Z and Y.

In the ST setting, there are two ways of training data ordering.

The training data ordering $Z = \left(Z^{(1)}, \dots, Z^{(n)}\right)'$ and $Y = \left(Y^{(1)}, \dots, Y^{(n)}\right)'$ means that we have written its as T-variate GRFs observations, where $Z^{(i)}$ and $Y^{(i)}$ are the T-dimensional vectors of temporal data at the i-th site, $i = 1, \dots, n$.

This case of training data ordering is applied in considering spatial GHMM models. Strategy for OBC distribution parameters estimation is relied on the spatial weighting of local model parameter estimators with ones from neighbouring sites. Call it Spatial weighting (SW) strategy. It will be properly described in the next section.

Alternatively, we could collect all data at time point t obtaining Z_t and Y_t.

The concatenation of such $\{Z_t : t = 1, \ldots, T\}$ and $\{Y_t : t = 1, \ldots, T\}$ amounts to leads to other structure $Z = (Z_1, \ldots, Z_T)$ and $Y = (Y_1, \ldots, Y_T)$.

This training data ordering are used for two types of ST Gaussian models with separable covariance functions.

Other strategy for OBC distribution parameters estimation is relied on the temporal averaging model parameter estimators over all training moments and using AR parameter estimators of particular site. Call it Temporal averaging (TA) strategy. It will be properly described in the next sections.

As it follows, it is assumed that $\{Z_t(\cdot) : t = 1, 2, \ldots\}$ is modeled by GGMs and GMRFs and $Z_{\cdot}^{(i)}$ is modeled by $AR(p)$, $i = 1, 2, \ldots, n$.

The choice of particular class of Gaussian data models for feature depends on the data collection process.

GGMs (e.g. [15]) are usually used in modeling geostatistical data such as measurements on several attributes at point-referenced spatial locations, $s_1, \ldots, s_n$ in a fixed region D where the measurement points vary continuously. These models are generally based on two common assumptions which are second order stationarity and isotropy. Second order stationarity implies that the mean of a process is constant and the covariance function depends on the spatial vector distance between two locations. When the covariance function only depends on the Euclidean distance (no direction) between two locations, the process is called isotropic.

Note, that issue of spatial sampling design of geostatistical data was properly studied by Zhu and Stein [107].

GMRFs (e.g. [6]) are used for modeling lattice (areal) data which are collected over a certain region. Here the focus is taken for the specific case of homogeneous data, where each variable in the random vector has the same distribution type and can be easily characterized by undirected graph models. For graphical data models, it is proved (see e.g. [99], that by imposing the restriction that each variable conditioned on other variables belong to a exponential family distribution, and then performing a Hammersley-Clifford-like analysis it is easy to derive the corresponding joint graphical model distribution, consistent with these node-conditional distributions.

It should be noted, that GGMs specify spatial associations directly through the parametric covariance models while GMRFs through the conditional specification of the precision matrix. For both models, the temporal evolution of feature is introduced by assuming that the spatial process following autoregressive dynamics. In other words, these two structures of training data presentation can be considered as a model of spatial or n-variate time series, or as a model of T spatial samples observed at n sites.

This thesis considers various models of spatial and temporal dependence within observed variables. To be specific, it focus on the following four forms of statistical dependence in space (S) and/or time (t) sense (domains):

1. Statistical Independence (SI);
2. Conditional Independence (CI);
3. Markov Dependence (MD);
4. Geostatistical Dependence (GD).

Scheme of ST model types with indicted indices distributions and statistical dependence forms is depicted in Figure 1.1.

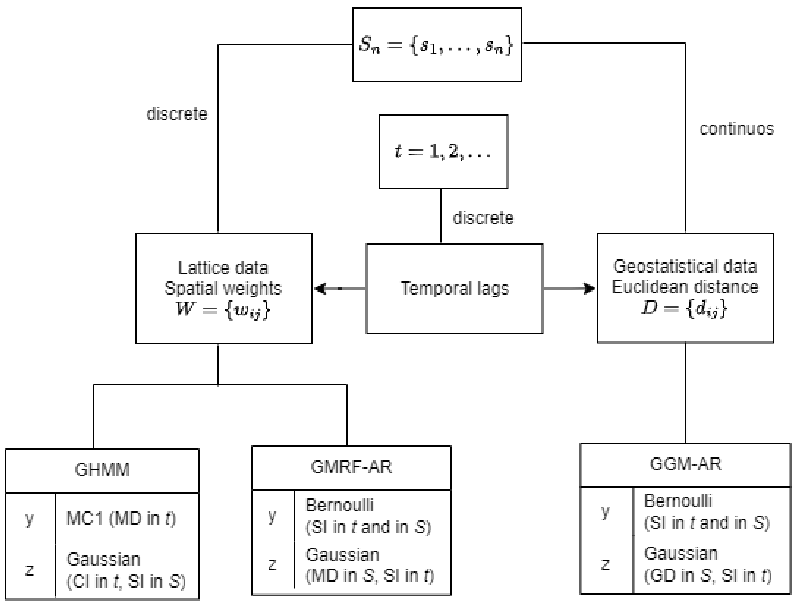


Figure 1.1: Scheme of investigated ST model types for observation at n spatial sites and at regular time moments with statistical dependence forms written in the parenthesis.

1.3 Bayesian one-step-ahead classification of Spatio-temporal observations

General problem is to classify Z_{T+1}, given training sample (Z, Y), i.e. to solve the so called OSA supervised classification problem.

The proposed method is based on the aggregating the classifications of $Z^{(i)}_{T+1}$ for $i = 1, \dots, n$.

The joint distribution of $\{Z_{T+1}, Y_{T+1}, Z, Y\}$ is denoted $P(Z_{T+1}, Y_{T+1}, Z, Y; \Psi)$ where Ψ is a set of all ST model parameters.

As it follows, for simplicity, we will sometimes skip the indication of dependence on Ψ where there is no necessity.

By the rule of recursive decomposition of the joint distribution we have for all sites

$$P\big(Z_{T+1}, Y_{T+1}, Z, Y\big) = P\big(Z_{T+1}\big|Y_{T+1}, Z, Y\big)P\big(Y_{T+1}, Y\big)P\big(Z|Y\big)$$

and for i-th site

$$P\Big(Z^{(i)}_{T+1}, Y^{(i)}_{T+1}, Z, Y\Big) = P\Big(Z^{(i)}_{T+1}\Big|Y^{(i)}_{T+1}, Z, Y\Big)P\Big(Y^{(i)}_{T+1}, Y\Big)P\Big(Z|Y\Big).$$

Denoting by Ψ a set of all ST model parameters the likelihood for observations in i-th site have the form

$$L\Big(z^{(i)}_{T+1}, y^{(i)}_{T+1} = l, z, y, \Psi\Big) = p\Big(z^{(i)}_{T+1}\Big|y^{(i)}_{T+1}, z, y\Big)Pr\Big(y^{(i)}_{T+1}, y\Big)p(z|y),$$

where $p\Big(z^{(i)}_{T+1}\Big|z, y, y^{(i)}_{T+1} = l\Big)$ denotes the conditional Probability density function (PDF) of OBC $Z^{(i)}_{T+1}$ given training sample, $Pr\Big(Y^{(i)}_{T+1} = l, Y\Big)$ denotes the joint distribution (or mass function) for class labels and $p(z|y)$ stands for conditional PDF of feature observations given class labels in training sample.

Hence it is obvious that by the Bayes rule optimal estimator of $Y^{(i)}_{T+1}$ is given by

$$\widehat{Y}^{(i)}_{T+1} = \arg\max_{l=0,1}\Big(L\Big(Z^{(i)}_{T+1}, Y^{(i)}_{T+1} = l, Z, Y\Big)\Big),$$

where $L\Big(Z^{(i)}_{T+1}, Y^{(i)}_{T+1} = l, Z, Y\Big)$ is the likelihood function of distribution

$P\left(Z^{(i)}_{T+1}, Y^{(i)}_{T+1}, Z, Y\right)$.

From the equation

$$\frac{L\left(z^{(i)}_{T+1}, y^{(i)}_{T+1} = 1, z, y, \Psi\right)}{L\left(z^{(i)}_{T+1}, y^{(i)}_{T+1} = 0, z, y, \Psi\right)} = \frac{p\left(z^{(i)}_{T+1}\middle|y^{(i)}_{T+1} = 1, z, y\right) Pr\left(y^{(i)}_{T+1} = 1, y\right)}{p\left(z^{(i)}_{T+1}\middle|y^{(i)}_{T+1} = 0, z, y\right) Pr\left(y^{(i)}_{T+1} = 0, y\right)}$$

log ratio of likelihoods of distributions specified above defining the BDF is

$$W_Z\left(z^{(i)}_{T+1}\right) = ln\frac{p\left(z^{(i)}_{T+1}\middle|z, y, y^{(i)}_{T+1} = 1\right)}{p\left(z^{(i)}_{T+1}\middle|z, y, y^{(i)}_{T+1} = 0\right)} + \gamma$$

where $\gamma = ln\frac{Pr\left(Y^{(i)}_{T+1}=1,y\right)}{Pr\left(Y^{(i)}_{T+1}=0,y\right)}$.

Hence in the thesis we focus on the various models for distributions $P\left(z^{(i)}_{T+1}\middle|z, y, y^{(i)}_{T+1} = l\right)$ *with conditional densities* $p\left(z^{(i)}_{T+1}\middle|z, y, y^{(i)}_{T+1} = l\right)$ *and* $Pr\left(Y^{(i)}_{T+1} = l, Y\right)$ *with joint probabilities* $Pr\left(Y^{(i)}_{T+1} = l, y\right)$ *of class labels.*

BDF classifies the observation $Z^{(i)}_{T+1} = z^{(i)}_{T+1}$ in following way: class $\widehat{Y}^{(i)}_{T+1} = 1$ if $W_Z\left(z^{(i)}_{T+1}\right) \geq 0$, and $\widehat{Y}^{(i)}_{T+1} = 0$ otherwise, i.e. $\widehat{Y}^{(i)}_{T+1} = H\left(W_Z\left(Z^{(i)}_{T+1}\right)\right)$. However, in practical applications all ST model parameters are rarely known.

Then the estimators of unknown parameters derived from training sample by the ML spatial weighting and other *method are plugged into BDF so obtaining PBDF* $\widehat{W}_Z\left(Z^{(i)}_{T+1}\right)$.

Proposed estimation procedures yield various kinds of synthetic estimators.

Label prediction (classification) in site s_i at time moment $t = T + 1$ based on PBDF is given by $\widehat{Y}^{(i)}_{T+1} = H\left(\widehat{W}_Z\left(Z^{(i)}_{T+1}\right)\right)$.

With an insignificant loss of generality, we focus on the linear independent of time mean $\mu_l(s;t) = \left(\beta^{(i)}_l\right)' x(s_i)$ where $x(s_i) = (x_1(s_i), \ldots, x_q(s_i))'$ is the vector of an explanatory variables (covariates) that are the same for all sites, and $\beta_l^{(i)}$ is a q-dimensional vector of regression parameters, $l = 0, 1$. That choice is motivated by considering only the cases when explanatory variables represent only the spatial coordinates or their functions that does not vary in time.

When the sites are locations the explanatory variables are represented only by the intercept term and spatial coordinates or their functions and for areal data only the intercept term is applied. In the cases when sites are areal units, the explanatory variable is usually represented only by the intercept. Note, that non-linear covariate-response relationships can also be included by adding transformations (e.g., squared) or spline basis functions of covariates.

1.4 Performance measures of supervised classifiers

Performance measures in classification are fundamental in assessing the quality of learning (parameter estimation) methods, selected data models and classification rules (see e.g. [29, 31, 91]). This thesis focuses on the following sorts of performance measures:

1. empirical - those based on confusion matrices derived from realized values of training and test samples;
2. probabilistic - those based on conditional distributions of OBC given training sample.

The performance criteria of the generative classifiers based on PBDF is evaluated by the confusion matrix formed for test data. Recording the results of correctly and incorrectly recognized test observations of each class.

This procedure is realized by organizing the observed data into training and testing sets. Then the classifier on the training data and its accuracy are validated on the test data. We focus on using T temporal observations for training. The observations at time moment $t = T + 1$ are using for testing. Let $I(\cdot)$ denote indicator function.

The form of the confusion matrix that will be applied for the assessment of the proposed classifier performances is shown in Table 1.1.

Table 1.1: Confusion matrix for OSA classification at i-th site.

$\boldsymbol{Y_{T+1}^{(i)}}$	$\boldsymbol{\widehat{Y}_{T+1}^{(i)}}$ **0**	**1**
0	$m_{00}^{(i)}$	$m_{01}^{(i)}$
1	$m_{10}^{(i)}$	$m_{11}^{(i)}$

where $m_{kl}^{(i)} = I\Big(Y_{T+1}^{(i)} = k\Big) I\Big(\widehat{Y}_{T+1}^{(i)} = l\Big)$ for $k, l = 0, 1$.

Set $n_{kl} = \sum\limits_{i=1}^{n} m_{kl}^{(i)}$ for $k, l = 0, 1$.

Then performance of particular classifier throughout n sites to be evaluated by aggregated confusion matrix that is obtained by summarizing local confusion matrices and presented in Table 1.2.

Table 1.2: Aggregated confusion matrix.

$\boldsymbol{Y_{T+1}^{(*)}}$	$\boldsymbol{\widehat{Y}_{T+1}^{(*)}}$ **0**	**1**
0	n_{00}	n_{01}
1	n_{10}	n_{11}

where $Y_{T+1}^{(*)}$, $\widehat{Y}_{T+1}^{(*)}$ denotes real and predicted class labels throughout n sites.

Traditionally, the most commonly used empirical measure of classifier performance is called *Accuracy rate* that shows the percentage of correctly classified test data is given by the formula

$$ACC = \frac{n_{00} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}}. \tag{1.1}$$

It should be noted that beside this overall accuracy measure the "partial" accuracy measures such as sensitivity (*true positive rate*) specified by

$$TPR = \frac{n_{00}}{n_{00} + n_{01}}$$

and specificity (*true negative rate*) specified by

$$TNR = \frac{n_{11}}{n_{10} + n_{11}}$$

are frequently used in machine learning.

However, in practice situations with significant disbalance between the majority and minority class examples frequently occur (see e.g. [52, 53, 65, 66]). Then the evaluation of the classifiers' performance must be carried out using specific metrics to take into account the class distribution. In this article we also used other performance evaluation measures based on confusion matrix usually called Balanced accuracy rate (see e.g. [52, 53, 80]) and specified by the formula

$$BAC = \Big(TPR + TNR\Big)/2. \tag{1.2}$$

G-mean is important to measure the avoidance of the overfitting to the minority class and the degree to which the majority class is marginalized (see e.g. [85]) and is specified by the formula

$$GAC = \sqrt{TPR \cdot TNR}. \tag{1.3}$$

This work is applied the Receiver Operator Characteristic (ROC) space [77] that allows to visualize the trade-off between TPR and $FPR = 1 - TNR$ *(false positive rate)*, for any confusion matrix corresponding to selected particular PBDF. Curve in the ROC space visualizes the classifier performances induced by different classifiers. Depicted points represent one considered classifier and a random classifier. It is easy to check that the area under the curve is equal the performance measure BAC in Figure 1.2.

There are several performance measures that is based on misclassifications instead of correct classifications.

The simplest one is the proportion of incorrectly classified observations that is given $PIC = 1 - ACC$.

In order to use the probabilistic performance measures, it is necessary to specify the conditional distribution of OBC.

From distributional assumptions specified above follows that the conditional distribution of $Z_{T+1}^{(i)}$ given $Z = z$, $Y = y$ and $Y_{T+1}^{(i)} = l$, is Gaussian, i. e.

$$\left(Z_{T+1}^{(i)} \middle| z, y, y_{T+1}^{(i)} = l\right) \sim N\left(\mu_{li(z)}^{T+1}, \Sigma_{T+1,i(z)}\right) \tag{1.4}$$

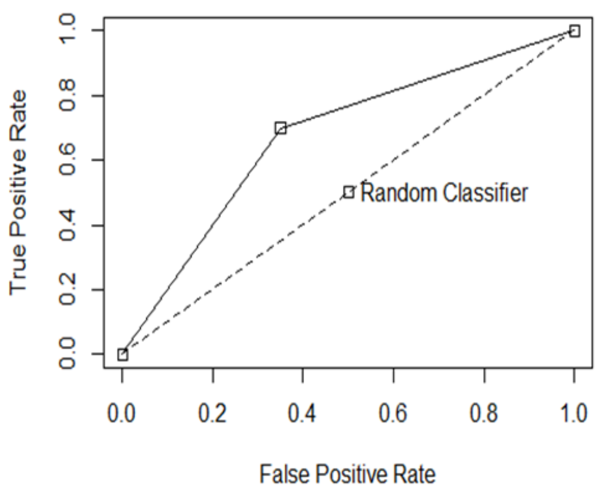


Figure 1.2: ROC plot.

where the form mean $\mu_{li(z)}^{T+1}$ and variance $\Sigma_{T+1,i(z)}$ depends on the type of ST data model, $l = 0, 1$.

The Local Bayes error rate associated with $W_Z\left(Z_{T+1}^{(i)}\right)$ is defined as

$$BE_i = \sum_{l=0}^{1} \pi_l P_l\left((-1)^l W_Z\left(Z_{T+1}^{(i)}\right) > 0\right)$$

where $\pi_l = P\left(y_{T+1}^{(i)} = l, y\right)$ and $P_l()$ stands for probability measure of conditional distribution specified in (1.4).

It is obvious, that probability of correct classification rate incurred by BDF $W_Z\left(Z_{T+1}^{(i)}\right)$ is specified by $PCC_i = 1 - BE_i$.

Plug-in or D-estimator (see e.g. [29, 91]) of Local Bayes error rate for PBDF $\widehat{W}_Z\left(Z_{T+1}^{(i)}\right)$ is by

$$PBE_i = \sum_{l=0}^{1} \widehat{\pi}_l \widehat{P}_l\left((-1)^l \widehat{W}_Z\left(Z_{T+1}^{(i)}\right) > 0\right)$$

where $\widehat{\pi}_l$ and $\widehat{P}_l()$ are the plug-in versions of π_l and $P_l()$ obtained by replacing unknown parameters with their estimates.

The aggregated versions of these error rates are defined as $ABE = \sum\limits_{i=1}^{n} \frac{BE_i}{n}$ and $APBE = \sum\limits_{i=1}^{n} \frac{PBE_i}{n}$.

Then D-estimator of the probability of correct classification is given by $DPCC_i = 1 - PBE_i$.

Define the label ratio for the training sample by $lr^{(i)} = \sum\limits_{t=1}^{T} I\big(Y_t^{(i)} = 1\big)/T$ assuming that it is a constant for all $i = 1, \ldots, n$.

Let for $t = 1, 2, \ldots, T$ denote by $IR_t = \frac{n_1^t}{n_0^t}$ the imbalance ratio where n_l^t designates the number sites with label equal l, i.e.

$$n_l^t = \sum_{i=1}^{n} I\Big(Y_t^{(i)} = l\Big).$$

1.5 Conclusions

Literature review of previous studies on spatial and ST data modeling and supervised classification methods were presented.

The novel framework for using of generative models to supervised Bayesian classification of ST data characterized by Gaussian distribution for feature and Bernoulli distribution for class label was introduced.

Three types of ST Gaussian data models with corresponding four statistical spatial and temporal statistical dependence structures (in the spatial and temporal context) for feature and class label variables are specified.

Also two original strategies for deriving the synthetic estimators of OBC distribution parameters are delivered.

The performance of classifiers was proposed for evaluation by using well-known empirical measures obtained from OSA classification confusion matrices or by using probabilistic measures that rely on conditional distribution of OBC.

Chapter 2

CLASSIFICATION ALGORITHMS FOR PARTICULAR TYPES OF SPATIAL-TEMPORAL DATA MODELS

2.1 Spatial Gaussian Hidden Markov Model

HMM as advanced statistical models originally were introduced as a classifier or predictor for various applications in the late 1960s. They were well suited for pattern recognition, classification and other problems (see. e.g. [55, 63, 109]). The widespread popularity of HMM was due to its strong mathematical structure and statistical foundation, as well as the presence of efficient training procedures in its framework [71, 72]. Although HMM was used for decades, it has been applied explicitly to modeling one dimensional (one site) data and less often used with spatial structure data [51].

This section is concerned with a generative model (see e.g. [61]) to supervised classification of ST data collected at fixed AUs (sites, locations) and specified by HMMs with continuous observations in each class (state). Traditionally this has invoked conditional independence assumption for observation distribution and Markov assumption for label transition probabilities (see e.g. [108]). Extension to the basic HMM when observation may be signified by continuous real value instead of discrete value. We focus on the models that describes the process of randomly generating an unobservable class label (state) sequences from the MC1 and generating a Gaussian observation from each class to produce an observable sequence from single probabilistic distribution [11, 62]. Unlike the conventional conditional density HMM, the proposed HMMs consider the temporal information of each AU which obtain the most possible sequence of label by the incorporating the spatial information through spatial weights.

A spatial HMM was proposed by Spezia et al. [82] and developed by Costes and Perret [14]. However, they ignored the temporal variability and only taken into account the spatial variability for feature and class label observations.

Recently, the discrete feature HMM incorporating the ST variability by introducing spatial and temporal transition probabilities was developed by Madadizadeh et al. [55]. Kadhem, Hewson and Kaimi [42] developed Poisson HMM to model spatial dependence in a network. However, only the cases with geostatistical models of observation based on spatial covariances or semivariograms were considered there.

For continuous feature spatial HMM Gong et al. [37] proposed the method relying on parametization of spatial weights, such increasingly large number of estimable parameters. It should be noted that HMM with spatial weighting have been successfully used in land-cover [14, 37, 88], but only considered in cases with discrete features.

Henceforth, the method introduced in this thesis will involve studies for spatial HMM with continuous (Gaussian) feature observations on lattice and will rely on the original strategies for estimation of model parameters.

In this section, we focus on the lattice data case specified by spatial sampling consisting of n fixed sites. For each site s_i data consist of feature observation $Z_t^{(i)}$ (further simply observation) and class label $Y_t^{(i)} \in \Omega = \{0, 1\}$ observed at consecutive time periods at $t = 1, 2, \ldots$.

It is known that the label sequence Y constitutes a MC1, and the observed sequence Z is only related to the corresponding class label sequence, assuming that the label of the observation is only related to the last year's label. Hence dependence between Y and Z constitutes a first-order HMM. Label variable Y for i-th AU is specified by transition probabilities matrix $A^{(i)} = \left(a_{kl}^{(i)}\right)$ where $a_{kl}^{(i)} = Pr\left(Y_{t+1}^{(i)} = l \middle| Y_t^{(i)} = k\right)$

for $k, l = 0, 1$. An obvious extension to the basic HMM model is to allow continuous observation space instead of a finite number of discrete symbols. In this model the emission probabilities matrix cannot be described as a simple matrix of point probabilities but rather as a complete PDF over the continuous observation space for each state. Therefore the values emission probabilities must be replaced with a conditional PDF of $Z_t^{(i)}$, given class label $Y_t^{(i)} = k$, is denoted by $p_k\left(z_t^{(i)}\right) = p\left(Z_t^{(i)} = z_t^{(i)} \middle| Y_t^{(i)} = k\right)$. The conditional distributions can in principle be arbitrary but usually they are restricted to be simple parametric distributions, like Gaussians. Assume that HMM with feature observation at i-th site given class label equal l follows Gaussian distribution $N\left(\left(\beta_l^{(i)}\right)' x_i, \sigma_i^2\right)$ and denote it by GHMM. Model obtained by comprising GHMM to n sites is called Spatial GHMM (in the following for brevity call it GHMM).

Set of all possible values of parameters of above conditional densities is denoted by $B^{(i)} = \left\{\beta^{(i)}, \sigma_i^2\right\}$ .The last component of GHMM is called an initial label distribution over classes and is denoted by $\pi^{(i)} = \left(\pi_0^{(i)}, \pi_1^{(i)}\right)$, where $\pi_k^{(i)} = Pr\left(Y_t^{(i)} = k\right)$, $k = 0, 1$. Specify the GHMM parameter set for i-th AU by $\Psi^{(i)} = \left(\pi^{(i)}, A^{(i)}, B^{(i)}\right)$, $i = 1, \ldots, n$.

The HMM allows us to talk about both observed events that we meet in the input and hidden events that we think of as causal factors in our probabilistic model. Recall that for HMM, each label realization produces only a single observation. Thus, the sequence of labels and the sequence of observations have the same length. A first-order HMM instantiates two simplifying assumptions.

First, Markov Assumption: Label Markovian-dependence in time (First-order Markov Chain) as with a MC1, the probability of a particular state depends only on the previous state:

$$Pr\left(Y_{t+1}^{(i)} \middle| Y_1^{(i)}, \ldots, Y_t^{(i)}\right) = Pr\left(Y_{t+1}^{(i)} \middle| Y_t^{(i)}\right) \text{ and}$$

$$Pr\left(Y^{(i)} = y^{(i)}\right) = Pr\left(Y_1^{(i)} = y_1^{(i)}\right) \prod_{t=1}^{T-1} Pr\left(Y_{t+1}^{(i)} = y_{t+1}^{(i)} \middle| Y_t^{(i)} = y_t^{(i)}\right).$$

This assumption ensures temporal dependence form for class label observations being of kind MD specified above.

Second, Output Independence: Feature conditional independence in time $p\Big(z^{(i)}\Big|y^{(i)}\Big) = \prod_{t=1}^{T} p\Big(z_t^{(i)}\Big|y_t^{(i)}\Big)$ or $P\Big(Z^{(i)}\Big|Y^{(i)}\Big) = \prod_{t=1}^{T} P\Big(Z_t^{(i)}\Big|Y_t^{(i)}\Big)$ and the probability of an output observation $Z_t^{(i)}$ depends only on the class that produced the observation and not on any other states or any other observations:

$$P\Big(Z_t^{(i)}\Big|Z_1^{(i)},\ldots,Z_T^{(i)},Y_1^{(i)},\ldots,Y_T^{(i)}\Big) = P\Big(Z_t^{(i)}\Big|Y_t^{(i)}\Big).$$

This assumption ensures spatial dependence structure for feature observations being of kind CI specified above.

In general, the goal of HMMs is to infer the most likely Y by giving the observed variables Z.

We assume that for $l = 0, 1$, the model of observation $Z_t^{(i)}$ *conditional on* $Y_t^{(i)} = l$ is

$$Z_t^{(i)} = \mu_l(s_i;t) + \varepsilon_t^{(i)},$$

where $\mu_l(s_i;t) = \big(\beta_l^{(i)}\big)' x(s)$ is deterministic ST mean and random error $\varepsilon_t^{(i)} \sim N(0, \sigma_i^2)$.

Then design matrix $X^{(i)}$ for observation vector $Z^{(i)}$ has the form

$$X^{(i)} = \begin{pmatrix} 1-y_1^{(i)} & y_1^{(i)} \\ 1-y_2^{(i)} & y_2^{(i)} \\ \vdots & \vdots \\ 1-y_T^{(i)} & y_T^{(i)} \end{pmatrix} \otimes x_i' \text{ where } x_i = x(s_i),\ i = 1,\ldots,n.$$

Conditional distribution of $Z^{(i)}$ conditional on $\Big\{Y^{(i)} = y^{(i)}\Big\}$ is T dimensional Gaussian, i.e.

$$\Big(Z^{(i)}\Big|y^{(i)}\Big) \sim N\Big(X^{(i)}\beta^{(i)}, \sigma_i^2 I_T\Big),$$

where $\beta^{(i)} = \begin{pmatrix} \beta_0^{(i)} \\ \beta_1^{(i)} \end{pmatrix}$ and I_T is $T \times T$ identity matrix.

Then under HMM independence assumption, the conditional distribution of $Z_{T+1}^{(i)}$ given $Z = z$ and $Y_{T+1}^{(i)} = l$, is Gaussian, i.e.

$$\Big(Z_{T+1}^{(i)}\big|z^{(i)}, y^{(i)}, y_{T+1}^{(i)} = l\Big) \sim N\Big(\Big(\beta^{(i)}\Big)' x_i, \sigma_i^2\Big), l = 0, 1$$

and $p\left(z_t^{(i)}\middle|y_t^{(i)}\right) = \varphi\left(z_t^{(i)}\middle|\left(\beta_{y_t^{(i)}}^{(i)})\right)'x_i, \sigma_i^2\right)$, where $\varphi\left(z_t^{(i)}\middle|\left(\beta_{y_t^{(i)}}^{(i)}\right)'x_i, \sigma_i^2\right)$ is the conditional density of $Z_t^{(i)}|Y_t^{(i)}$ and traditionally is called emission density.

Given this one-to-one mapping and the Markov assumptions, for a particular hidden state sequence $y^{(i)}$ and an observation sequence $z^{(i)}$, the conditional likelihood of the observation sequence is

$$l_i\left(z^{(i)}\middle|y^{(i)}, \beta^{(i)}, \sigma_i^2\right) = \prod_{t=1}^{T} \varphi\left(z_t^{(i)}\middle|\left(\beta_{y_t^{(i)}}^{(i)}\right)'x_i, \sigma_i^2\right).$$

Diagram of training sample $\left(Z^{(i)}, Y^{(i)}\right)$ at i-th site is depicted in Figure 2.1.

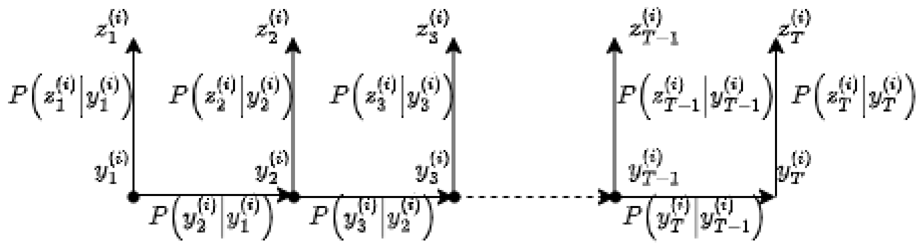


Figure 2.1: Visualization of training sample $\left(Z^{(i)}, Y^{(i)}\right)$ for GHMM in i-th site.

For given an $\Psi^{(i)} = \left(\pi^{(i)}, A^{(i)}, B^{(i)}\right)$ and an observation $z^{(i)}$ and label $y^{(i)}$ sequences (i.e. labeled training sample is known), the likelihood $L_i = L\left(z^{(i)}, y^{(i)}, \Psi^{(i)}\right)$ has the following form $L_i = Pr\left(y^{(i)}\right) l_i\left(z^{(i)}\middle|y^{(i)}\right)$ where $Pr\left(y^{(i)}\right) = Pr\left(Y_1^{(i)} = y_1^{(i)}\right) \prod_{t=1}^{T-1} Pr\left(Y_{t+1}^{(i)} = y_{t+1}^{(i)}\middle|Y_t^{(i)} = y_t^{(i)}\right)$.

For this model
$P\left(Z_{T+1}^{(i)}\middle|Z, Y_{T+1}^{(i)} = l\right) = P\left(Z_{T+1}^{(i)}\middle|y_{T+1}^{(i)} = l\right) = N\left(\left(\beta_l^{(i)}\right)'x_i, \sigma_i^2\right)$ and
$Pr\left(Y_{T+1}^{(i)} = l, Y\right) = Pr\left(Y_{T+1}^{(i)} = l, Y^{(i)}\right) = Pr\left(Y_1^{(i)} = y_1^{(i)}\right) \prod_{t=1}^{T-1} Pr\left(Y_{t+1}^{(i)}\right.$
$\left.= y_{t+1}^{(i)}\middle|Y_t^{(i)} = y_t^{(i)}\right) Pr\left(Y_{T+1}^{(i)} = l\middle|Y_T^{(i)} = y_T^{(i)}\right)$.

Define an augmented likelihood

$$\begin{aligned} L_i^+\left(z_{T+1}^{(i)}, y_{T+1}^{(i)}, \lambda^{(i)}\right) = L\left(z^{(i)}, z_{T+1}^{(i)}, y^{(i)}, y_{T+1}^{(i)}, \Psi^{(i)}\right) = \\ = Pr\left(y^{(i)}, y_{T+1}^{(i)}\right) l_i\left(z^{(i)}, z_{T+1}^{(i)} \middle| y^{(i)}, y_{T+1}^{(i)}\right) \end{aligned} \tag{2.1}$$

where $Pr\left(y^{(i)}, y_{T+1}^{(i)}\right) = Pr\left(Y_1^{(i)} = y_1^{(i)}\right) \prod_{t=1}^{T} Pr\left(Y_{t+1}^{(i)} = y_{t+1}^{(i)} \middle| Y_t^{(i)} = y_t^{(i)}\right)$ and $l_i\left(z^{(i)}, z_{T+1}^{(i)} \middle| y^{(i)}, y_{T+1}^{(i)}\right) = \prod_{t=1}^{T+1} p\left(z_t^{(i)} \middle| y_t^{(i)}\right)$.

Then the criterion for BCR for the OSA temporal prediction of the class label is

$$\widehat{y}_{T+1}^{(i)} = \arg\max_{k=0,1} \left(L_i^+\left(z_{T+1}^{(i)}, y_{T+1}^{(i)} = k, \Psi^{(i)}\right)\right).$$

In the HMM and machine learning context, this problem is called as decoding problem.

Under the assumption that the GHMM is completely specified given labeled training sample $\left(z^{(i)}, y^{(i)}\right)$, it is that known BDF minimizing the total probability of misclassification of $Z_{T+1}^{(i)}$ is formed by the log-ratio of augmented conditional likelihoods of distributions (see [58]) specified in (2.1), that is

$$W_Z\left(z_{T+1}^{(i)} \middle| y_T^{(i)}, \Psi^{(i)}\right) = ln \frac{L_i^+\left(z_{T+1}^{(i)}, y_{T+1}^{(i)} = 1, \Psi^{(i)}\right)}{L_i^+\left(z_{T+1}^{(i)}, y_{T+1}^{(i)} = 0, \Psi^{(i)}\right)}.$$

So BDF classifies the observation $Z_{T+1}^{(i)} = z_{T+1}^{(i)}$ in following way: class label takes value 1 if $W_Z\left(z_{T+1}^{(i)} \middle| y_T^{(i)}, \Psi^{(i)}\right) \geq 0$, and 0 otherwise.

However, in practical applications all statistical parameters $\beta^{(i)}$, σ_i^2, $a_{y_T^{(i)},0}^{(i)}$, $a_{y_T^{(i)},1}^{(i)}$ of model are rarely known.

For this ST model type, the SW strategy for model parameter estimation is implemented. The realization of this strategy is described below.

The estimators of unknown parameters $\widehat{\beta}^{(i)}$, $\widehat{\sigma}_i^2$ are derived by the

ML method by maximizing $l_i\left(z^{(i)}\middle|y^{(i)}\right)$ and $\widehat{a}^{(i)}_{y_T^{(i)},0}$, $\widehat{a}^{(i)}_{y_T^{(i)},1}$ by maximizing $Pr\left(y^{(i)}\right)$. In classical HMM context it is also called a learning problem.

Hence we obtain the estimators of local HMM parameters estimators $\widehat{\Psi}^{(i)}$ consisted of:

the ML estimator of regression parameters

$$\widehat{\beta}^{(i)} = \left(\left(X^{(i)}\right)' X^{(i)}\right)^{-1}\left(X^{(i)}\right)' Z^{(i)};$$

bias-adjusted ML estimator of variance

$$\widehat{\sigma}_i^2 = \left(Z^{(i)} - X^{(i)}\widehat{\beta}^{(i)}\right)'\left(Z^{(i)} - X^{(i)}\widehat{\beta}^{(i)}\right)/(T-2q);$$

the ML estimators of transition probabilities

$$\widehat{a}_{00}^{(i)} = \frac{\sum_{t=2}^{T} I\left(y_{t-1}^{(i)} = 0\right) I\left(y_t^{(i)} = 0\right)}{\sum_{t=2}^{T} I\left(y_{t-1}^{(i)} = 0\right)}, \widehat{a}_{01}^{(i)} = \frac{\sum_{t=2}^{T} I\left(y_{t-1}^{(i)} = 0\right) I\left(y_t^{(i)} = 1\right)}{\sum_{t=2}^{T} I\left(y_{t-1}^{(i)} = 0\right)},$$

$$\widehat{a}_{10}^{(i)} = \frac{\sum_{t=2}^{T} I\left(y_{t-1}^{(i)} = 1\right) I\left(y_t^{(i)} = 0\right)}{\sum_{t=2}^{T} I\left(y_{t-1}^{(i)} = 1\right)}, \widehat{a}_{11}^{(i)} = \frac{\sum_{t=2}^{T} I\left(y_{t-1}^{(i)} = 1\right) I\left(y_t^{(i)} = 1\right)}{\sum_{t=2}^{T} I\left(y_{t-1}^{(i)} = 1\right)}.$$

It should be noted that we are not interested in the initial label distributions $\pi^{(i)} = \left(\pi_0^{(i)}, \pi_1^{(i)}\right)$ and consider it fixed, since it does not presented in BDF expression.

In this study, we propose two rules of the transition probability estimators for the i-th AU:

(M1) $\widehat{a}^{(i)}_{y_T^{(i)},1} = \widehat{a}_{01}^{(i)} I\left(y_T^{(i)} = 0\right) + \widehat{a}_{11}^{(i)} I\left(y_T^{(i)} = 1\right)$
and $\widehat{a}^{(i)}_{y_T^{(i)},0} = \widehat{a}_{00}^{(i)} I\left(y_T^{(i)} = 0\right) + \widehat{a}_{10}^{(i)} I\left(y_T^{(i)} = 1\right)$;

(M2) $\widehat{a}^{(i)}_{y_T^{(i)},1} = q^{(i)} I\left(y_T^{(i)} = 0\right) + p^{(i)} I\left(y_T^{(i)} = 1\right)$
and $\widehat{a}^{(i)}_{y_T^{(i)},0} = p^{(i)} I\left(y_T^{(i)} = 0\right) + q^{(i)} I\left(y_T^{(i)} = 1\right)$

where $p^{(i)} = \frac{\sum_{t=2}^{T} I\left(y_{t-1}^{(i)}=y_t^{(i)}\right)}{T-1}$ and $q^{(i)} = 1 - p^{(i)} = \frac{\sum_{t=2}^{T} I\left(y_{t-1}^{(i)}\neq y_t^{(i)}\right)}{T-1}$ denote label persistence rate and change rate, respectively.

We propose the following spatial weighting rules for local model parameter estimators

$$\widetilde{\beta}^{(i)} = \Big(\sum_{j=1}^{n} w_{ij}^{*}\widehat{\beta}^{(j)} + \widehat{\beta}^{(i)}\Big)/2, \widetilde{\sigma}_i^2 = \Big(\sum_{j=1}^{n} w_{ij}^{*}\widehat{\sigma}_j^2 + \widehat{\sigma}_i^2\Big)/2,$$

$$\widetilde{a}^{(i)}_{y_T^{(i)},1} = \Big(\sum_{j=1}^{n} w_{ij}^{*}\widehat{a}^{(j)}_{y_T^{(j)},1} + \widehat{a}^{(i)}_{y_T^{(i)},1}\Big)/2, \widetilde{a}^{(i)}_{y_T^{(i)},0} = \Big(\sum_{j=1}^{n} w_{ij}^{*}\widehat{a}^{(j)}_{y_T^{(j)},0} + \widehat{a}^{(i)}_{y_T^{(i)},0}\Big)/2.$$

Denote the four sets of spatially weighted estimators of parameters by $\widehat{\Psi}_1^{(i)} = \Big(\pi^{(i)}, \widehat{A}_1^{(i)}, \widetilde{\beta}^{(i)}, \widetilde{\sigma}_i^2\Big)$, $\widehat{\Psi}_2^{(i)} = \Big(\pi^{(i)}, \widehat{A}_2^{(i)}, \widetilde{\beta}^{(i)}, \widetilde{\sigma}_i^2\Big)$,

$\widetilde{\Psi}_1^{(i)} = \Big(\pi^{(i)}, \widetilde{A}_1^{(i)}, \widetilde{\beta}^{(i)}, \widetilde{\sigma}_i^2\Big)$, $\widetilde{\Psi}_2^{(i)} = \Big(\pi^{(i)}, \widetilde{A}_2^{(i)}, \widetilde{\beta}^{(i)}, \widetilde{\sigma}_i^2\Big)$

where $\widehat{A}_j^{(i)}$ and $\widetilde{A}_j^{(i)}$ denotes the estimators of transition matrices relied on the estimation rules MJ), $J = 1, 2$.

Note, that parameter vector $\pi^{(i)}$ is not estimated due to reasons mentioned before.

The following four discriminant functions obtained by plugging above estimators in BDF are under consideration:

1. $\widehat{W}_{ZP1}\Big(z_{T+1}^{(i)}\Big|\widehat{\Psi}_1^{(i)}\Big) = \Big(z_{T+1}^{(i)} - \frac{\left(\widetilde{\beta}^{(i)}\right)' Fx_i}{2}\Big)\big((\widetilde{\beta}^{(i)})' Gx_i\big)/\widetilde{\sigma}_i^2 + ln\left(\frac{\widehat{a}^{(i)}_{y_T^{(i)},1}}{\widehat{a}^{(i)}_{y_T^{(i)},0}}\right)$
   Partial spatial weighting level with inserted transition probability estimator M1 and denote it by PSW1;

2. $\widehat{W}_{ZP2}\Big(z_{T+1}^{(i)}\Big|\widehat{\Psi}_2^{(i)}\Big) = \Big(z_{T+1}^{(i)} - \frac{\left(\widetilde{\beta}^{(i)}\right)' Fx_i}{2}\Big)\big((\widetilde{\beta}^{(i)})' Gx_i\big)/\widetilde{\sigma}_i^2 + ln\left(\frac{\widehat{a}^{(i)}_{y_T^{(i)},1}}{\widehat{a}^{(i)}_{y_T^{(i)},0}}\right)$
   Partial spatial weighting level with inserted transition probability estimator M2 and denote it by PSW2;

3. $\widehat{W}_{ZP3}\Big(z_{T+1}^{(i)}\Big|\widetilde{\Psi}_1^{(i)}\Big) = \Big(z_{T+1}^{(i)} - \frac{\left(\widetilde{\beta}^{(i)}\right)' Fx_i}{2}\Big)\big((\widetilde{\beta}^{(i)})' Gx_i\big)/\widetilde{\sigma}_i^2 + ln\left(\frac{\widetilde{a}^{(i)}_{y_T^{(i)},1}}{\widetilde{a}^{(i)}_{y_T^{(i)},0}}\right)$
   Complete spatial weighting level with inserted transition probability estimator M1 and denote it by CSW1;

4. $\widehat{W}_{ZP4}\left(z_{T+1}^{(i)}\middle|\widetilde{\Psi}_2^{(i)}\right) = \left(z_{T+1}^{(i)} - \frac{\left(\widetilde{\beta}^{(i)}\right)' F x_i}{2}\right)\left(\left(\widetilde{\beta}^{(i)}\right)' G x_i\right)/\widetilde{\sigma}_i^2 + ln\left(\frac{\widetilde{a}^{(i)}_{y_T^{(i)},1}}{\widetilde{a}^{(i)}_{y_T^{(i)},0}}\right)$

   Complete spatial weighting level with inserted transition probability estimator M2 and denote it by CSW2.

They differ on the exploited estimation techniques and level of the incorporated ST information.

The novelty of our method lies in the inclusion of spatial weights in GHMM parameter estimation from labeled training sample and using PBDF for classification (decoding) of the test observation. This enables us to enrich and generalize the existed generative classification methods of ST data modelled by HMM.

Our method is aimed at situation when BDF is implemented for data that are collected at fixed AUs (or sites as they are also known) over the fixed sequential time periods.

At the beginning, for each local GHMM the maximum likelihood estimators of regression coefficients and variances and two types of transition probabilities estimators are derived from labeled training sample.

Further, GHMM parameter estimators in two spatial weighting levels are plugged in BDF. Two spatial weighting levels are specified as follows:

a) lower level called as PSW indicates that parameter estimators of Gaussian component are spatially weighted when transition probability estimators remain unweighted;

b) higher level called as CSW comprise the cases with spatially weighted estimators of Gaussian parameters and transition probabilities.

Developed classification method via GHMM is realized in the following order:

1. Likelihood for training sample at each AU.
2. Estimators of parameters from local likelihood.
3. Spatial weighting of parameter estimators by rules M1 and M2.

4. Construction of the augmented likelihood including the OBC.
5. Construction of PBDF by replacing the parameters in BDF with spatially weight parameters.
6. PBDF performance evaluation via measures derived from the confusion matrix.

Algorithm of the classifier realization in visualized in Figure 2.2.

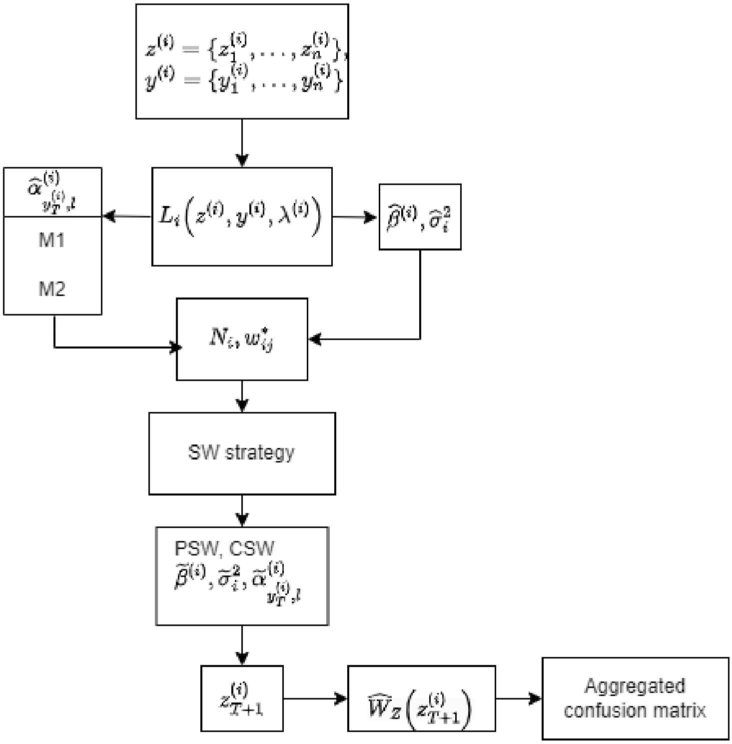


Figure 2.2: Scheme of supervised classification for GHMM.

Performance measures of the classifier are chosen to be the ACC and BAC evaluated from the confusion matrices for a test sample. The proposed methodology is illustrated for simulated data and for real data set, i.e. annual death rate data collected by the Institute of Hygiene of the Republic of Lithuania from the 60 municipalities in the period from 2001 to 2019. Presented comparison of proposed classifier by the introduced criteria with existing ones can aid in the selection of proper classification rules of ST areal data.

The experimental results showed that proposed GHMM in CSW level in a majority of cases have an advantage in classification accuracy over GHMM in PSW level by both performance measures (details in subsections 3.1.1 and 3.2.1).

Comprehensive description of this classifier is presented by Dučinskas et al. [28].

2.2 Gaussian Markov Random Field-Autoregressive model

In this section, lattice data generated by a ST random process with separable covariance functions is considered. These models can realistically be performed when feature observations at every time model is generated by spatial GMRF, and at a given spatial site they follow the time series AR model. The practical use of GMRF for modeling large scale spatial phenomena has significantly increased after recent advances on the efficient simulation of GMRFs (see e.g. [18]).

Some authors have investigated the performance of the BDF when training samples consist of temporally dependent observations (see e.g. [47, 58]) while others focused their attention only on Bayesian supervised classification in the purely spatial context by using GMRF (see e.g. [5, 25]).

However, statistical discriminant analysis of ST data has been rarely considered previously (see e.g. [40, 110]).

As mentioned before, this study comprises the cases wherein at each spatial site the class label can vary over time. That assumption essentially widens the application area of the proposed classification method, especially for the cases with the imbalanced data. Separability of covariances was assumed in order to reduce complexity due to interdependencies between features. The general objective of this section is to extend the previous investigations of ST point referenced geostatistical data to ST areal data. As a novel modeling contribution, we propose using three popular spatial autocorrelation indices, i.e. Moran's I, Geary's C and Getis-Ord G in specification decision threshold values relied on class label distribution.

The performance criterion of the classifier based on PBDF realized on previously described ST data models is chosen to be the ACC, which shows the percentage of correctly classified test data, BAC and GAC evaluated from the confusion matrices. For numerical illustrations, the Homogeneous conditionally autoregressive model (HCAR) (special case of GMRF model) for spatial covariance (see e.g. [81]) and $AR(p)$ model for temporal covariance are considered. This is the extension of the generative classification method for point referenced geostatistical Gaussian ST data developed by Karaliutė and Dučinskas [43, 44] to the particular lattice data model case. Detailed comparison among proposed approach classifiers with various spatial autocorrelation indices (decision threshold values) and classifier based HMM is performed.

The general GMRF model and its particular case HCAR (see e.g. [69, 74]) are explored under the traditional assumption of a feature Markovian dependence for full conditionals in space given by the following equation

$$P\left(Z_t^{(i)}\middle|y_t, z_t^{(j)}, j \neq i\right) = P\left(Z_t^{(i)}\middle|y_t(N_i), z_t(N_i)\right), t = 1, 2, \ldots$$

where $z_t(N_i) = \{z_t(s_j) : s_j \in N_i\}$ and $y_t(N_i) = \{y_t(s_j) : s_j \in N_i\}$.

Another assumption of this ST data model are SI class label observations, i.e. $Pr\left(y^{(i)}, y_{T+1}^{(i)}\right) = \prod_{t=1}^{T+1} Pr\left(Y_t^{(i)} = y_t^{(i)}\right)$, i.e. class labels are independent Bernoulli random variables. However, it is assumed that parameter of Bernoulli distribution depends on ST context of observation, i.e.

$$\left\{Y_t^{(i)}, i = 1, \ldots, n; t = 1, \ldots, T+1\right\} \sim ind.Ber(\pi_1(s_i, t)),$$

where denotes a specific to the ST observation context at location s_i and time t.

Set $\pi_0(s_i, t) = Pr\left(Y_t^{(i)} = 0\right) = 1 - \pi_1(s_i, t)$.

This method generative OSA classification methods is developed in Karaliutė and Dučinskas [45].

For this data model type, TA strategy for parameter estimation from training data is exploited. This estimation strategy is described below.

Denote by X the $n \times 2qT$ matrix $X = \left(X_{(1)}, X_{(2)}, \ldots, X_{(T)}\right)$ where

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

and $x_i = x(s_i)$, $i = 1, \ldots, n$.

Then the matrix model for Z conditional on $\{Y_t = y_t, t = 1, \ldots, T\}$ is

$$Z = XB + E,$$

where $B = I_T \otimes \beta$ with $2q \times 1$ dimensional parameter vector $\beta = \left(\beta_0^{'}, \beta_1^{'}\right)^{'}$ and $n \times T$ matrix of Gaussian errors $E = (\varepsilon(s_i; t) : i = 1, \ldots, n; t = 1, \ldots, T)$ and I_T is $T \times T$ identity matrix.

In the present and next subsections, we restrict our attention to *the separable ST covariance model*

$$C(s, u; t, r) = C_S(s, u) C_T(t, r),$$

where $C_S(s, u)$ denotes pure spatial covariance between observations in areas s and u and $C_T(t, r)$, denotes pure temporal covariance between observations at time points t and r.

Under this assumption, the ST covariance structure factors into a purely spatial and a purely temporal component, allowing for computationally efficient estimation and inference. Therefore, separable covariance models are popular even in situations where they are not physically justifiable. Many statistical tests of separability have recently been proposed and are based on parametric models (see e.g. [10, 34]) or spectral methods [77].

Without an insignificant loss of generality, we restrict our attention to HCAR lattice models for spatial covariance [81] with original parametric structure proposed by De Oliveira and Ferreira [18]. These are well-suited to the case of small samples, and ensures good frequentest properties of ML estimators of regression coefficients, spatial dependence and scale parameters.

In the present work we focus on the Gaussian areal data with Markov models for pure spatial and temporal covariances.

Specific attention is given to the ST Gaussian model with pure spatial covariances belonging to the HCAR models and with pure temporal covariance of stationary AR model.

In the experimental part of this study we focus only on $AR(1)$ model.

So temporal covariance matrix for feature is

$$C_1 = \frac{\sigma_T^2}{1-\alpha^2}\begin{pmatrix} 1 & \alpha & \cdots & \alpha^{T-2} & \alpha^{T-1} \\ \alpha & 1 & \cdots & \alpha^{T-3} & \alpha^{T-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha^{T-2} & \alpha^{T-3} & \cdots & 1 & \alpha \\ \alpha^{T-1} & \alpha^{T-2} & \cdots & \alpha & 1 \end{pmatrix}$$

where α denotes the autoregresion parameter and σ_T^2 denotes the white noise variance. Then the inverse of temporal covariance matrix C_T is obtained by the Yule-Walker equations (see [9]) for $AR(1)$ model, i.e.

$$C_T^{-1} = \frac{1}{\sigma_T^2}\begin{pmatrix} 1 & -\alpha & \cdots & 0 & 0 \\ -\alpha & 1+\alpha^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1+\alpha^2 & -\alpha \\ 0 & 0 & \cdots & -\alpha & 1 \end{pmatrix}.$$

Then $\mu_{li(z)}^{T+1} = \mu_{li}^{T+1} + \big((0,\ldots,0,\alpha_1)' \otimes e_i'\big)vec(E)$

and $\Sigma_{T+1,i(z)} = \sigma_T^2 c_S^{ii}$ where $c_S^{ii} = \sigma_s^2(1+\eta h_i)^{-1}$ and $\mu_{li}^{T+1} = x_i'\beta_l$.

For more general model $AR(p)$ it was derived the following conditional moment functions (see [44])

$\mu_{li(z)}^{T+1} = \mu_{li}^{T+1} + \big((0,\ldots,0,\alpha_p,\ldots,\alpha_2,\alpha_1)' \otimes e_i'\big)vec(E)$

and $\Sigma_{T+1,i(z)} = c_S^{ii}\sigma_T^2$ where $c_S^{ii} = \sigma_s^2(1+\eta h_i)^{-1}$.

Spatial covariance matrix for n AUs $C_s = \sigma_s^2 R$, where $R = (r_{ij})$ denotes the spatial correlation matrix with $R = (I_n + \eta H)^{-1}$. Here $H = \Lambda - W$ the Laplace matrix with Λ the diagonal matrix with diagonal elements $\lambda_{ii} = \sum\limits_{s_j \in N_i} w_{ij}$, $i = 1,\ldots,n$, $\eta \geq 0$ is a spatial dependence parameter and $\sigma_s > 0$ is a scale parameter. Then the spatial precision matrix is defined by $\Omega_S = C_s^{-1} = (I_n + \eta H)/\sigma_s^2$.

Then $P\left(Z_{T+1}^{(i)} \middle| Z, Y, y_{T+1}^{(i)}\right) = P\left(Z_{T+1}^{(i)} \middle| y_{T+1}(N_i), z_{T+1}(N_i), y_{T+1}^{(i)}\right) = N\left(x_i'\beta_{y_{T+1}^{(i)}} + \frac{\eta}{1+\eta h_i}\sum_{j=1}^{n} w_{ij}\left(z_{T+1}^{(j)} - x_j'\beta_{y_{T+1}^{(j)}}\right), \frac{\sigma_s^2}{1+\eta h_i}\right)$.

Hence we can say that spatial variation of feature follows HCAR model with specific parametrization proposed by De Oliveira and Ferreira [18].

Under the assumption that the classes are completely specified, the Conditional BDF (CBDF) minimizing the total probability of misclassification is formed by the log-ratio of conditional likelihood of distributions (see [58]) specified in

$$W_Z\left(z_{T+1}^{(i)}\right) = L_Z\left(z_{T+1}^{(i)}\right) - \gamma_i(T+1), \tag{2.2}$$

where $\gamma_i(T+1) = ln\frac{\pi_0(s_i,T+1)}{\pi_1(s_i,T+1)}$ and

$$L_Z\left(z_{T+1}^{(i)}\right) = \left(z_{T+1}^{(i)} - \frac{\mu_{1i(z)}^{T+1} + \mu_{0i(z)}^{T+1}}{2}\right)\Sigma_{T+1,i(z)}^{-1}\left(\mu_{1i(z)}^{T+1} - \mu_{0i(z)}^{T+1}\right).$$

It is easy to deduce that discriminant function $W_Z\left(z_{T+1}^{(i)}\right)$ is optimal under the criterion of the minimum of misclassification probability (see [58]).

Call the probability of misclassification for $W\left(Z_{T+1}^{(i)}\right)$ as Local Bayes error rate and denote it by P_i. Also denote squared *Mahalanobis distance between conditional distributions by*

$$\Delta_{T+1,i(z)}^2 = \left(\mu_{1i(z)}^{T+1} - \mu_{0i(z)}^{T+1}\right)'\Sigma_{T+1,i(z)}^{-1}\left(\mu_{1i(z)}^{T+1} - \mu_{0i(z)}^{T+1}\right) = \frac{(x_1'(\beta_1 - \beta_0))^2}{c_S^{ii}\sigma_T^2}$$

where $c_S^{ii} = \sigma_s^2(1+\eta h_i)^{-1}$.

Lemma 1. *The Local Bayes error rate is*

$$BE_i = \pi_1(s_i, T+1)\Phi\left(-\frac{\Delta_{T+1,i(z)}}{2} - \frac{\gamma_i(T+1)}{\Delta_{T+1,i(z)}}\right) + \pi_0(s_i, T+1)\Phi\left(-\frac{\Delta_{T+1,i(z)}}{2} + \frac{\gamma_i(T+1)}{\Delta_{T+1,i(z)}}\right).$$

Proof. It is easy to derive that conditional distribution of $W\left(z_{T+1}^{(i)}\right)$ given $Z = Z$, $Y_{T+1}^{(i)} = l$ is univariate Gaussian distribution with mean

$$E\left(W\left(Z_{T+1}^{(i)}\right)\Big|Z = z, Y_{T+1}^{(i)} = l\right) = (-1)^{l+1}\frac{\Delta_{T+1,i(z)}^{2}}{2} + \gamma_i(T+1)$$

and variance

$$Var\left(W\left(Z_{T+1}^{(i)}\right)\Big|Z = z, Y_{T+1}^{(i)} = l\right) = \Delta_{T+1,i(z)}^{2}, l = 0, 1.$$

Using properties of the multivariate Gaussian distribution, we complete the proof of Lemma 1. $\square$

Error estimation is detailed 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.

So BDF classifies the observation $Z_{T+1}^{(i)}$ the following way: class label takes value 1 if $L_Z\left(z_{T+1}^{(i)}\right) \geq \gamma_i(T+1)$, and 0 otherwise.

So $L_Z\left(z_{T+1}^{(i)}\right)$ is a linear term, and $\gamma_i(T+1)$ plays role of a decision threshold.

The probability of misclassification for $W_Z\left(z_{T+1}^{(i)}\right)$ is optimal under the criterion of the minimum of misclassification probability.

However, in practical applications all statistical parameters of populations are rarely known.

In the present section, we will use a TA strategy for parameter estimation to calculate a conditional $Z_{T+1}^{(i)}$ distribution.

Final estimators of β and σ_s^2 are obtained by averaging the ML estimators $\widehat{\beta}_{(t)} = \left(X_{(t)}' R^{-1} X_{(t)}\right)^{-1} X_{(t)}' R^{-1} Z_t$,

$\widehat{\sigma}_{s(t)}^2 = \left(Z_t - X_{(t)}\widehat{\beta}_{(t)}\right)' R^{-1}\left(Z_t - X_{(t)}\widehat{\beta}_{(t)}\right)/(n-2q)$, i.e.

$$\widetilde{\beta} = \sum_{t=1}^{T}\frac{\widehat{\beta}_{(t)}}{T}, \widetilde{\sigma}_s^2 = \sum_{t=1}^{T}\frac{\widehat{\sigma}_{s(t)}^2}{T}.$$

For i-th site, the ML estimators of AR parameters are computed from training sample.

For $AR(1)$ denote its by $\widehat{\alpha}_1^{(i)}$ and $\widehat{\sigma}^2_{T(i)}$.

Then the estimators of unknown parameters are plugged into BDF specified in (2.2).

Replace parameters with their estimators for BDF specified in (2.2). *The PBDF takes the following form*:

$$\widehat{W}_Z\left(z_{T+1}^{(i)}\right) = \left(z_{T+1}^{(i)} - \frac{\widehat{\mu}_{1i(z)}^{T+1} + \widehat{\mu}_{0i(z)}^{T+1}}{2}\right) \times \\ \times \widehat{\Sigma}^{-1}_{T+1,i(z)}\left(\widehat{\mu}_{1i(z)}^{T+1} - \widehat{\mu}_{0i(z)}^{T+1}\right) - \gamma_i(T+1)$$

where $\widehat{\mu}_{li(z)}^{T+1} = \left(x_i'\widetilde{\beta}_l + \frac{\eta}{1+\eta h_i}\sum\limits_{j=1}^{n} w_{ij}\left(z_{T+1}^{(j)} - x_j'\widehat{\beta}_{y_{T+1}^{(j)}}\right)\right)$

and $\widehat{\Sigma}_{T+1,i(z)} = \widehat{\sigma}^2_{T(i)}\frac{\widetilde{\sigma}^2_S}{1+\eta h_i}$.

So PBDF has the same threshold as BDF, but differs in the linear terms.

Then the D-estimator of PBDF performance is given by

$$PBE_i = \pi_1(s_i, T+1)\Phi\left(-\frac{\widehat{\Delta}_{T+1,i(z)}}{2} - \frac{\gamma_i(T+1)}{\widehat{\Delta}_{T+1,i(z)}}\right) \\ + \pi_0(s_i, T+1)\Phi\left(-\frac{\widehat{\Delta}_{T+1,i(z)}}{2} + \frac{\gamma_i(T+1)}{\widehat{\Delta}_{T+1,i(z)}}\right),$$

where $\widehat{\Delta}^2_{T+1,i(z)} = ((\Delta\beta)'x_i)^2/(\widetilde{\sigma}^2_s\widehat{\sigma}^2_{T(i)})$, $l = 0, 1$.

Three models label distribution for observation in s_i at $t = T+1$ are proposed. They differ on the type and level of the incorporated ST information.

It is obvious that each model for label distribution specifies specific decision threshold value $\gamma_i(T+1)$ for considered classifier.

Set $\bar{z}_t = \sum\limits_{i=1}^{n} z_t^i/n$ and $\widetilde{z}_t = z_t - \bar{z}_t 1_n$, $S_0 = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} w_{ij}$, $J = 1_n 1_n'$ where 1_n denotes an n dimensional column vector, whose elements are all equal 1.

Introduce Global Moran's I, Geary's C and Getis-Ord G for n AUs at the time moment t by

$$I(t) = \frac{n}{S_0}\frac{\widetilde{z}_t'W\widetilde{z}_t}{\widetilde{z}_t'\widetilde{z}_t}, C(t) = \frac{(n-1)}{2S_0}\frac{z_t'Lz_t}{\widetilde{z}_t'\widetilde{z}_t}, G(t) = \frac{z_t'Wz_t}{z_t'(J-I)z_t}.$$

It should be noted that the expected value of Moran's I is $-1/(n-1)$. Values of I that exceed $-1/(n-1)$ indicate positive spatial autocorrelation, in which similar values, either high values or low values, are spatially clustered. Values of I below $-1/(n-1)$ indicate negative spatial autocorrelation, in which neighboring values are dissimilar.

The theoretical expected value for Geary's C is 1. A value of Geary's C less than 1 indicates positive spatial autocorrelation, while a value larger than 1 points to negative spatial autocorrelation.

Geary's C is inversely related to Moran's I, but it is not identical. *Moran's I is a measure of global spatial autocorrelation, while Geary's C is more sensitive to local spatial autocorrelation*. Geary's C is also known as Geary's contiguity ratio or simply Geary's ratio.

The Getis-Ord G-Statistic (see e.g. [32, 35]) distinguishes between hot spots and cold spots. It identifies *spatial concentrations*, i.e. G is relatively large if high values cluster together and G is relatively low if low values cluster together.

Label distribution based on Moran's I, Geary's C and Getis-Ord G is denoted by

$$\pi_{M1t}(s_i, T+1) = \frac{1}{1+exp(-I(T)y_i^*(T))},$$
$$\pi_{C1t}(s_i, T+1) = \frac{1}{1+exp(-C(T)y_i^*(T))},$$
$$\pi_{G1t}(s_i, m+1) = \frac{1}{1+exp(-G(m)y_i^*(m))}$$

respectively, where $y_i^*(t) = 2y_t^{(i)} - 1$.

Recall that the HMM method of classification is based on assuming conditional independence for feature observations and first order

Markov property for labels. In this article we restricted our attention on Gaussian observation with regression mean model and constant variance for each AU [42, 62]. The spatially weighted estimators of regression coefficients, variances and transition probabilities are inserted into the PBDF.

The label prediction in AU s_i at time moment $t = T + 1$ is $\widehat{Y}_{T+1}^{(i)} = H\left(\widehat{W}_Z\left(z_{T+1}^{(i)}\right)\right)$, where $H(\cdot)$ is the Heaviside step function.

In the next section, we will explore comparison of performance for a classifier based on $\widehat{W}_Z\left(z_{T+1}^{(i)}\right)$ conducted for several class label distribution parameters for simulated and real data.

Algorithm of the classifier realization in visualized in Figure 2.3.

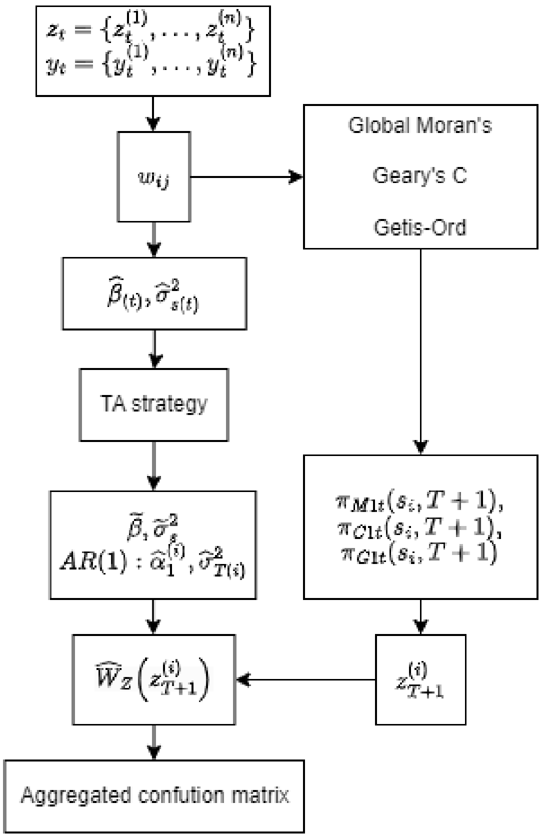


Figure 2.3: Scheme of supervised classification for GMRF-AR.

2.3 Gaussian Geostatistical-Autoregressive model

In this section, we considered the type of data models where feature spatial variability is represented by GGMs and temporal variability is presented by a time series AR model. Hence, spatial covariance of feature is separable, i.e. it is factorized to spatial and temporal covariances (see e.g. [15]). It should be noted that Šaltytė-Benth and Dučinskas [110] considered supervised classification of ST data modeled by GGM in particular cases when an OBC is uncorrelated with the training sample that consists of interdependent feature variables.

As in the previous section, separability and stationarity of covariances was assumed to reduce of complexity due to interdependencies between features.

GGMs are used to model continuous spatial Gaussian processes using a spatial covariances or variograms which are just a function of distance and direction between locations. For this type of data the spatial index can vary continuously in the field domain. GGMs are generally based on two common assumptions, second order stationarity and isotropy. Second order stationarity implies that the mean of a process is constant and that the covariance function depends on the spatial vector distance between two locations. When the covariance function only depends on the Euclidean distance (no direction) between two locations, the process is called isotropic. The spatial covariance of a stationary and isotropic spatial process could be modeled using parametric functions of Euclidean distances.

It is assumed that spatial covariance belongs to the family of directly specified Matern type or other covariance models and that temporal covariance follows the stationary $AR(p)$ model.

For the performance of classifiers, the values derived in Local Bayes error rates and empirical error rates are used. Empirical error rates are validated by a modified leave-one-out method when all but one observation is used to complete the classification rule, and this rule is then used to classify the omitted observation (see e.g. [29]). For numerical illustrations, the two powered-exponential isotropic models for spatial covariance are considered. Temporal covariance is obtained by the Yule-Walker equations for $AR(1)$ models. This is the extension of $AR(1)$ case explored in Karaliutė and Dučinskas [43]. Performance of the proposed classification rule is compared for different parameters pure spatial and temporal covariances and prior class probabilities models.

Independence of labels are also assumed for all sites and temporal moments, i.e. $Pr\left(y^{(i)}, y_{T+1}^{(i)}\right) = \prod_{t=1}^{T+1} Pr\left(Y_t^{(i)} = y_t^{(i)}\right)$, i.e. class labels are independent Bernoulli random variables, with the distribution parameter possibly depending on ST context of observation.

Here we use the structure for training data in the following ordering

$Z = (Z_1, \ldots, Z_T)$ and $Y = (Y_1, \ldots, Y_T)$.

Let $z_t^{(-i)} = \{z^{(j)}, j \neq i\}$, $y_t^{(-i)} = \{y_t^{(j)}, j \neq i\}$, $P\left(Z_{T+1}^{(i)} \middle| z, y, y_{T+1}^{(i)}\right) = P\left(Z_{T+1}^{(i)} \middle| z_t^{(-i)}, y_t^{(-i)}, y_{T+1}^{(i)}\right)$, $i = 1, 2, \ldots, n$.

Denote pure spatial covariance $n \times n$ matrix denote by $C_S = (c_S^{ij} = C_S(s_i, s_j); i, j = 1, \ldots, n)$. In practice, it usually belongs to Matern class ([57], Section 3.2) or powered-exponential class of covariance functions (see e.g. [84], p. 31, [15], section 4.1.1). So, the components of Z_i are statistically dependent, i.e. belongs to Geostatistical Dependence form denoted by GD.

However, in the geostatistical literature, second-order properties are typically not characterized using only the covariance function. Instead, they are more often represented by semivariograms, that are defined directly on the increments. It is known that covariance matrices for RF with finite variances are in a one-to-one correspondence with semivariogram matrices and variances (see e.g. [20]). This was the motivating argument for the consideration of the spatial classification problems via semivariograms. We focused on the particular case when populations are specified by different regression parameters of GRF with second order properties expressed in terms of semivariograms and variances.

Introduce an n-dimensional vector and $n \times n$ matrix for semivariograms $\Gamma_S = (\gamma_{ij}; i, j = 1, \ldots, n)$, where $\gamma_{ij} = Var(\varepsilon(s_i; t) - \varepsilon(s_j; t))/2$.

Then the model of $vec(Z)$ conditional on $Y_t = y_t$, $t = 1, \ldots, T$ is

$$vec(Z) = vec(XB) + vec(E) \tag{2.3}$$

where $vec(E)$ is the $nT \times 1$ vector of random errors that has normal distribution, i.e. $vec(E) \sim N_{nT}(0, \Sigma)$ with $\Sigma = var(vec(E)) = C_T \otimes C_S$ and C_T is a $p \times p$ matrix of pure temporal covariances, $C_T = (c_T^{tr} = C_T(t, r); t, r = 1, \ldots, T)$.

From the covariance separability assumption, it follows that $vec(E) \sim$

$N_{nT}(0, C_T \otimes C_S)$ with $C_T = (c_T^{tr} = C_T(t,r); t,r = 1,\dots,T)$ denoting the $T \times T$ 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.

Then under ST data model specified in (2.3) it follows that

$$Z \sim N_{n \times T}(XB, C_T \otimes C_S).$$

We concern with the problem of classification of the observations $Z_{T+1}^{(i)}$, $i = 1,\dots,n$ 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 $t = T+1$.

Set $c_T^{T+1,r} = C_T(T+1,r); r = 1,\dots,T$, $c_T^{T+1} = \left(c_T^{T+1,1},\dots,c_T^{T+1,T}\right)'$ and $e_i^{'}$ – the i-th row of identity matrix I_n.

For $AR(1)$ model we obtain $c_T^{T+1} = \frac{\sigma_T^2}{1-\alpha^2}(\alpha^T \alpha^{T-1} \dots \alpha)$.

Then we can conclude that in $l = 0,1$ the conditional distribution of $Z_{T+1}^{(i)}$ given $Z = z$ and $Y_{T+1}^{(i)} = l$, is Gaussian, i. e.

$$\left(Z_{T+1}^{(i)} \middle| Z = z; Y_{T+1)}^{(i)} = l\right) \sim N\left(\mu_{li(z)}^{T+1}, \Sigma_{T+1,i(z)}\right), \tag{2.4}$$

where

$$\mu_{li(z)}^{T+1} = \beta_l^{'} x_i + \left(\left(c_T^{T+1}\right)' C_T^{-1} \otimes e_i'\right) vec(E),$$

$$\Sigma_{T+1,i(m)} = var\left(z_{T+1}^{(i)}\right) - c_s^{ii}\left(c_T^{T+1}\right)' C_T^{-1} c_T^{T+1} = c_S^{ii} \rho_{T+1}$$

with $\rho_{T+1} = c_T^{T+1,T+1} - \left(c_T^{T+1}\right)' C_T^{-1} c_T^{T+1}$.

Conditional likelihood $L\left(z_{T+1}^{(i)} | Z, Y, y_{T+1}^{(i)} = 1\right) \pi_1(s_i, T+1)$.

Set $Pr\left(Y_{T+1}^{(i)} = l \middle| Y = y\right) = \pi_l(s_i, T+1)$, $l = 0,1$ and, for simplicity, call them prior class probabilities at location s_i and time moment $T+1$.

Introduce the conditional likelihood $L\left(z_{T+1}^{(i)}, \middle| Z, Y, y_{T+1}^{(i)} = 1\right) \pi_1(s_i, T+1)$ of the distribution $P\left(Z_{T+1}^{(i)} \middle| Z = z; Y_{T+1}^{(i)} = l\right) Pr\left(Y_{T+1}^{(i)} = l \middle| Y\right)$.

Under the assumption that the classes are completely specified, the CBDF minimizing the probability of misclassification is formed by the log-ratio of conditional likelihood of distributions, i.e.

$$W_Z\left(z_{T+1}^{(i)}\right) = ln \frac{L\left(z_{T+1}^{(i)} \middle| z_t^{(-i)}, y_t^{(-i)}, y_{T+1}^{(i)} = 1\right)\pi_1(s_i, T+1)}{L\left(z_{T+1}^{(i)} \middle| z_t^{(-i)}, y_t^{(-i)}, y_{T+1}^{(i)} = 0\right)\pi_0(s_i, T+1)}.$$

For the model specified in (2.3), (2.4) given

$$W_Z\left(z_{T+1}^{(i)}\right) = \left(z_{T+1}^{(i)} - \frac{\mu_{1i(z)}^{T+1} + \mu_{0i(z)}^{T+1}}{2}\right) \times \\ \times \Sigma_{T+1,i(z)}^{-1}\left(\mu_{1i(z)}^{T+1} - \mu_{0i(z)}^{T+1}\right) + \gamma_i(T+1),$$

where $\gamma_i(T+1) = ln \frac{\pi_1(s_i,T+1)}{\pi_2(s_i,T+1)}$.

The probability of misclassification for $W_Z\left(z_{T+1}^{(i)}\right)$ is optimal under the criterion of the minimum of misclassification probability (see [58]).

Like in previous subsection we intendinding to derive Local Bayes error rate incurred by $W_Z\left(z_{T+1}^{(i)}\right)$ in a closed-form. Call the probability of misclassification for as and denote it by BE_i.

Note that for considered ST data model squared Mahalanobis distance between conditional distributions of $Z_{T+1}^{(i)}$ is

$$\Delta_{T+1,i(z)}^2 = \left((\Delta\beta)'x_i\right)^2/\left(c_S^{ii}\sigma_T^2\right).$$

Then GGM model with pure temporal covariance of stationary $AR(p)$ model we prove the following lemma.

Lemma 2. *The Local Bayes error rate for the GGM-AR is*

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

where $Q_{li} = -\frac{\Delta_{T+1,i(z)}}{2} + (-1)^l \frac{\gamma_i(T+1)}{\Delta_{T+1,i(z)}}$, $\Delta_{T+1,i(z)}^2 = ((\Delta\beta)'x_i)^2/\left(c_s^{ii}\sigma_T^2\right)$, $l = 0, 1$.

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

Then $\mu_{li(z)}^{T+1} = \beta_l' x_i + ((0, \ldots, 0, \alpha_p, \ldots, \alpha_1)' \otimes e_i')vec(E)$ and $\Sigma_{T+1,i(z)} = c_S^{ii}\sigma_T^2$.

By using the properties of multivariate Gaussian distribution and inserting the above expressions we complete the proof of Lemma 2. $\square$

When model parameters unknown, we apply the TA strategy by using ML estimators of β and σ_s^2

$$\widetilde{\beta} = \sum_{t=1}^{T} \frac{\widehat{\beta}_{(t)}}{T}, \widetilde{\sigma}_s^2 = \sum_{t=1}^{T} \frac{\widehat{\sigma}_{s(t)}^2}{T}, \tag{2.5}$$

where $\widehat{\beta}_{(t)} = \big(X_{(t)}' C_s^{-1} X_{(t)}\big)^{-1} X_{(t)}' C_s^{-1} Z_t$
and $\widehat{\sigma}_{s(t)}^2 = \big(Z_t - X_{(t)}\widehat{\beta}_{(t)}\big)' C_s^{-1} \big(Z_t - X_{(t)}\widehat{\beta}_{(t)}\big)/(n-2q)$.

For i-th site, the ML estimators of AR parameters are computed from training sample.

For $AR(1)$ denote its by $\widehat{\alpha}_1^{(i)}$ and $\widehat{\sigma}_{T(i)}^2$.

Set $\Delta\widehat{\beta} = \widehat{\beta}_1 - \widehat{\beta}_0$.

We focus on the PBDF:

$$\widehat{W}_Z\Big(z_{T+1}^{(i)}\Big) = \left(z_{T+1}^{(i)} - \frac{\widehat{\mu}_{1i(z)}^{T+1} + \widehat{\mu}_{0i(z)}^{T+1}}{2}\right) \widehat{\Sigma}_{T+1,i(z)}^{-1} \Big(\widehat{\mu}_{1i(z)}^{T+1} - \widehat{\mu}_{0i(z)}^{T+1}\Big) - \gamma_i(T+1).$$

Then PBDF performance evaluated by the D-estimator is given by

$$PBE_i = \pi_0(s_i, T+1)\Phi\Big(\widehat{Q}_{0i}\Big) + \pi_1(s_i, T+1)\Phi\Big(\widehat{Q}_{1i}\Big),$$

where $\widehat{Q}_{li} = -\frac{\widehat{\Delta}_{T+1,i(z)}}{2} + (-1)^l \frac{\gamma_i(T+1)}{\widehat{\Delta}_{T+1,i(z)}}$
and $\widehat{\Delta}_{T+1,i(z)}^2 = ((\Delta\beta)' x_i)^2/(\widetilde{\sigma}_s^2 \widehat{\sigma}_{T(i)}^2)$, $l = 0, 1$.

It should be noted that we focused on pure spatial covariances belonging to the family of powered-exponential isotropic models.

It is known that practitioners sometime prefer to use semivariograms instead of spatial covariances. Under the intercept assumption ($x_1(s) = 1$ for $s \in D$) and error stationarity assumption ($Var(\varepsilon(s_i;t)) = \sigma_0^2$) it is easy to derive the following formulas for the inverse semivariogram matrix Γ_S and covariance matrix C_s (see [27])

$\Gamma_S^{-1} = -C_S^{-1} + C_S^{-1} J C_S^{-1} / (\Sigma^{\cdot\cdot} - 1/\sigma_0^2)$

and $C_S^{-1} = -\Gamma_S^{-1} + \Gamma_S^{-1} J \Gamma_S^{-1} / (\Gamma^{\cdot\cdot} - 1/\sigma_0^2)$,

where $\Sigma^{\cdot\cdot} = 1' C_S^{-1} 1$ and $\Gamma^{\cdot\cdot} = 1' \Gamma_S^{-1} 1$.

Then it is straightforward to derive the closed-form expression of ML regression parameters formula (2.5) through the semivariogram (see [27]) $\widehat{\beta}_{(t)} = \left(X'_{(t)} \Gamma_S^{-1} X_{(t)}\right)^{-1} X'_{(t)} \Gamma_S^{-1} Z_t$.

It is reasonable to use this formula in the construction of PBDF.

Usually the performance of classifier based on PBDF is chosen to be the accuracy, which shows the percentage of correctly classified test data.

Three label distribution models for observation in s_i at $t = T + 1$ are proposed. They differ on the type and level of the incorporated ST information.

The first one is based on a Temporal Weighted Moving Average (TWMA) method for label distribution

$$\pi_{1t}(s_i, T+1) = \frac{\sum\limits_{t=1}^{T} y_t^{(i)} t}{(1+T)T/2}.$$

The second one adds spatial weights

$$\pi_{1ts}(s_i, T+1) = \frac{\sum\limits_{t=1}^{T} y_t^{(i)} t + \sum\limits_{j: s_j \in N_i} \sum\limits_{t=1}^{T} y_t^{(j)} t}{(1+T)T/2(1 + card(N_i))}$$

where j denotes the index of the nearest neighbor to s_i. We will denote this method by using the Spatial Temporal Weighted Moving Average (STWMA).

Third one includes only label context based on Spatial Temporal

Unweighted Average and denoted by STUA

$$\pi_{1s}(s_i, T+1) = \frac{\sum_{t=1}^{T} \left(y_t^{(i)} + \sum_{j:s_j \in N_i} y_t^{(j)} \right)}{T(1 + card(N_i))}.$$

The performance criteria of the generative classifier based on PBDF is evaluated by the confusion matrix formed for test data and that records the results of correctly and incorrectly recognized test observations of each class.

It is obvious that each model for label distribution specifies a particular decision threshold value $\gamma_i(T+1)$ for considered classifier.

Algorithm of the classifier realization in visualized in Figure 2.4.

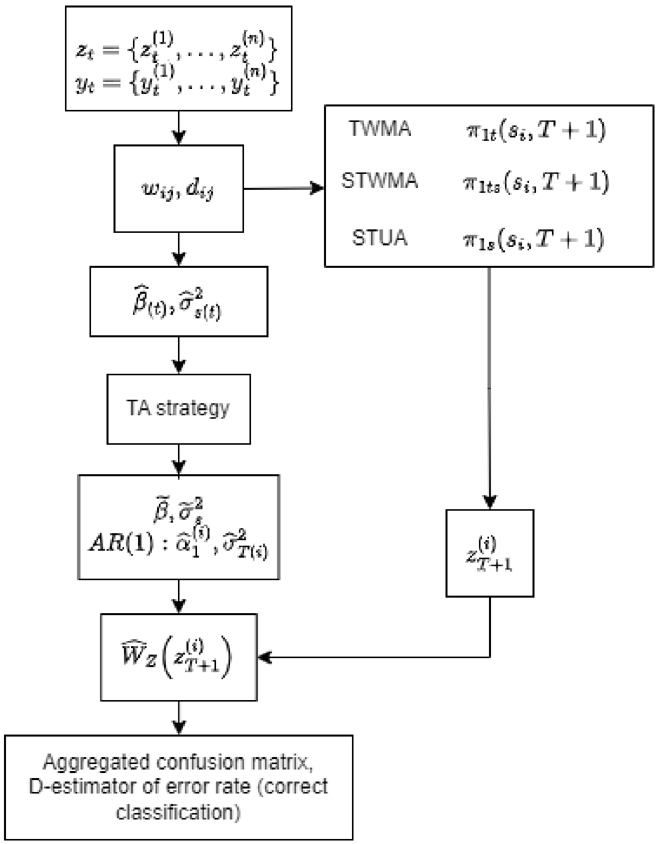


Figure 2.4: Scheme of supervised classification for GGM-AR.

2.4 Conclusions of section

This section we developed a novel supervised generative models to Bayesian OSA classification for the following types of ST data models:

- Areal Gaussian data based on GHMM implementing original strategy based on two spatial weighting levels for the estimators of the model parameters;
- Areal feature generated by GMRF with separable covariances comprising the situations when the independent class labels depends on various spatial association indices;
- Geostatistical data generated by GRF with separable covariances comprising the situations when the independent class labels depends on spatial distances and temporal lags;
- For GMRF and GGM data models Local Bayes error rates and their D-estimators of are derived in the closed form.

Chapter 3

EMPIRICAL INVESTIGATION FOR SIMULATED AND REAL DATA

3.1 Simulated data

To perform the comparison of ten proposed classifiers, we start with some simulation studies based on a fixed spatial array of $n = 20$ sites distributed in $S \in [0, 5] \times [0, 5]$. The data are generated at each site for $t = 1, \cdots, 50$ (i.e. $T = 50$) consecutive time periods and consist of feature and class label observations. It is assumed that feature variable is distributed according to the normal distribution with a mean following the first order trend surface model which consists of three explanatory variables, i.e. the first explanatory variable is a constant equal to 1, and the second and third explanatory variables are x and y coordinates.

Spatial sampling set, illustrated by an undirected graph with 20 vertices (sites) of a neighbourhood system represented by edges, is shown in Figure 3.1.

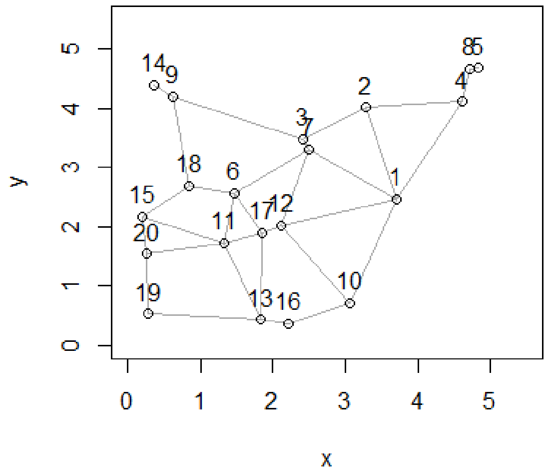


Figure 3.1: Spatial sampling set S_{20} represented by an undirected graph.

Specify the spatial adjacency by n dimensional quadratic matrix $W = (w_{ij} : i, j = 1, \ldots, n)$ with (i, j) element

$$w_{ij} = \begin{cases} 0, & if \quad i = j \\ 1, & if \quad s_j \in N_i \\ 0, & otherwise \end{cases}.$$

Assume, that the neighbourhood system $\{N_i\}$ is specified by graph edges. Then spatial weights matrix, which had been introduced before, has the form

$$
W = \begin{pmatrix}
0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}.
$$

This form of a weight matrix is used for lattice (areal) data models.

Coordinates of the sites for the geostatistical data models are used for the specification of spatial dependencies.

Matrix of Euclidean distances D with elements $d_{ij} = |s_i - s_j|$ between points as depicted in Figure 3.1 has the following form:

$$
D = \begin{pmatrix}
0.0 & 1.6 & 1.6 & 1.9 & 2.5 & 2.2 & 1.5 & 2.4 & 3.5 & 1.9 & 2.5 & 1.6 & 2.8 & 3.9 & 3.5 & 2.6 & 1.9 & 2.9 & 3.9 & 3.6 \\
1.6 & 0.0 & 1.0 & 1.3 & 1.7 & 2.3 & 1.1 & 1.6 & 2.7 & 3.3 & 3.0 & 2.3 & 3.9 & 3.0 & 3.6 & 3.8 & 2.6 & 2.8 & 4.6 & 3.9 \\
1.6 & 1.0 & 0.0 & 2.3 & 2.7 & 1.3 & 0.2 & 2.6 & 1.9 & 2.8 & 2.1 & 1.5 & 3.1 & 2.3 & 2.6 & 3.1 & 1.7 & 1.8 & 3.6 & 2.9 \\
1.9 & 1.3 & 2.3 & 0.0 & 0.6 & 3.5 & 2.3 & 0.6 & 4.0 & 3.7 & 4.1 & 3.3 & 4.6 & 4.3 & 4.8 & 4.5 & 3.5 & 4.0 & 5.6 & 5.1 \\
2.5 & 1.7 & 2.7 & 0.6 & 0.0 & 4.0 & 2.7 & 0.1 & 4.2 & 4.3 & 4.6 & 3.8 & 5.2 & 4.5 & 5.3 & 5.0 & 4.1 & 4.4 & 6.2 & 5.5 \\
2.2 & 2.3 & 1.3 & 3.5 & 4.0 & 0.0 & 1.3 & 3.9 & 1.8 & 2.4 & 0.9 & 0.8 & 2.2 & 2.1 & 1.3 & 2.3 & 0.8 & 0.6 & 2.4 & 1.6 \\
1.5 & 1.1 & 0.2 & 2.3 & 2.7 & 1.3 & 0.0 & 2.6 & 2.1 & 2.7 & 2.0 & 1.3 & 3.0 & 2.4 & 2.6 & 3.0 & 1.6 & 1.8 & 3.6 & 2.9 \\
2.4 & 1.6 & 2.6 & 0.6 & 0.1 & 3.9 & 2.6 & 0.0 & 4.1 & 4.3 & 4.5 & 3.7 & 5.1 & 4.4 & 5.2 & 5.0 & 4.0 & 4.3 & 6.1 & 5.4 \\
3.5 & 2.7 & 1.9 & 4.0 & 4.2 & 1.8 & 2.1 & 4.1 & 0.0 & 4.2 & 2.6 & 2.6 & 4.0 & 0.3 & 2.1 & 4.2 & 2.6 & 1.5 & 3.7 & 2.7 \\
1.9 & 3.3 & 2.8 & 3.7 & 4.3 & 2.4 & 2.7 & 4.3 & 4.2 & 0.0 & 2.0 & 1.6 & 1.3 & 4.6 & 3.2 & 0.9 & 1.7 & 3.0 & 2.8 & 2.9 \\
2.5 & 3.0 & 2.1 & 4.1 & 4.6 & 0.9 & 2.0 & 4.5 & 2.6 & 2.0 & 0.0 & 0.9 & 1.4 & 2.9 & 1.2 & 1.6 & 0.6 & 1.1 & 1.6 & 1.1 \\
1.6 & 2.3 & 1.5 & 3.3 & 3.8 & 0.8 & 1.3 & 3.7 & 2.6 & 1.6 & 0.9 & 0.0 & 1.6 & 3.0 & 1.9 & 1.7 & 0.3 & 1.4 & 2.4 & 1.9 \\
2.8 & 3.9 & 3.1 & 4.6 & 5.2 & 2.2 & 3.0 & 5.1 & 4.0 & 1.3 & 1.4 & 1.6 & 0.0 & 4.2 & 2.4 & 0.4 & 1.5 & 2.5 & 1.6 & 1.9 \\
3.9 & 3.0 & 2.3 & 4.3 & 4.5 & 2.1 & 2.4 & 4.4 & 0.3 & 4.6 & 2.9 & 3.0 & 4.2 & 0.0 & 2.2 & 4.4 & 2.9 & 1.8 & 3.9 & 2.9 \\
3.5 & 3.6 & 2.6 & 4.8 & 5.3 & 1.3 & 2.6 & 5.2 & 2.1 & 3.2 & 1.2 & 1.9 & 2.4 & 2.2 & 0.0 & 2.7 & 1.7 & 0.8 & 1.7 & 0.6 \\
2.6 & 3.8 & 3.1 & 4.5 & 5.0 & 2.3 & 3.0 & 5.0 & 4.2 & 0.9 & 1.6 & 1.7 & 0.4 & 4.4 & 2.7 & 0.0 & 1.6 & 2.7 & 2.0 & 2.3 \\
1.9 & 2.6 & 1.7 & 3.5 & 4.1 & 0.8 & 1.6 & 4.0 & 2.6 & 1.7 & 0.6 & 0.3 & 1.5 & 2.9 & 1.7 & 1.6 & 0.0 & 1.3 & 2.1 & 1.6 \\
2.9 & 2.8 & 1.8 & 4.0 & 4.4 & 0.6 & 1.8 & 4.3 & 1.5 & 3.0 & 1.1 & 1.4 & 2.5 & 1.8 & 0.8 & 2.7 & 1.3 & 0.0 & 2.2 & 1.3 \\
3.9 & 4.6 & 3.6 & 5.6 & 6.2 & 2.4 & 3.6 & 6.1 & 3.7 & 2.8 & 1.6 & 2.4 & 1.6 & 3.9 & 1.7 & 2.0 & 2.1 & 2.2 & 0.0 & 1.0 \\
3.6 & 3.9 & 2.9 & 5.1 & 5.5 & 1.6 & 2.9 & 5.4 & 2.7 & 2.9 & 1.1 & 1.9 & 1.9 & 2.9 & 0.6 & 2.3 & 1.6 & 1.3 & 1.0 & 0.0
\end{pmatrix}.
$$

These distances are used in exponential and squared-exponential spatial correlation models, that will be introduced in the next sections.

At first, we generate $v = 100$ simulation runs (replications) for a test label by the Bernoulli distribution described above and which correspond to a conditional Gaussian observation.

3.1.1 Realization of Gaussian Hidden Markov Model for simulated data

For constructing a classifier for GHMM data, the pseudo algorithm for the proposed supervised generative classifier for the simulation of spatial data sample is as follows:

Algorithm 1 Simulation spatial data sample.

1: **Input**: A network graph with 20 vertices with spatial weights W, number of points time $T+1$, number of time iterations v:
2: **for** $i = 1$ to 20 **do**
3: calculate the estimators of regression parameters $\widehat{\beta}^{(i)}$, variance $\widehat{\sigma}_i^2$, transition probabilities $\widehat{a}_{kl}^{(i)}$, $\widetilde{a}^{(i)}_{y_T^{(i)},l}$
4: calculate the local model parameter estimators $\widetilde{\beta}^{(i)}$, $\widetilde{\sigma}_i^2$, $\widetilde{a}^{(i)}_{y_T^{(i)},l}$
5: calculate PBDF $\widehat{W}_Z\left(z_{T+1}^{(i)}\right)$
6: calculate accuracy rates ACC, BAC, GAC
7: **end for**

Conditional distribution of $Y_{T+1}^{(i)}$ given $Y_T^{(i)} = 1$ is Bernoulli with parameter $\widehat{a}_{11}^{(i)}$, i.e. $\big(Y_{T+1}^{(i)}|Y_T^{(i)} = 1\big) \sim Ber\big(\widehat{a}_{11}^{(i)}\big)$.

Hence, it is obvious that $\big(Y_{T+1}^{(i)}|Y_T^{(i)} = 0\big) \sim Ber\big(\widehat{a}_{00}^{(i)}\big)$.

For all 20 AUs, spatially weighted estimates of Gaussian distribution parameters and transition probabilities (PSW and CSW specified by rules M1 and M2) are presented in Tables 3.1 and 3.2.

Table 3.1: Estimators of Gaussian distribution parameters.

PSW and CSW			
i	$(\widetilde{\beta}_0^{(i)})'$	$(\widetilde{\beta}_1^{(i)})'$	$\widetilde{\sigma}_i^2$
1	−0.043, −0.133, −0.126	0.043, 0.133, 0.126	0.249
2	−0.035, −0.095, −0.128	0.044, 0.125, 0.164	0.323
3	−0.040, −0.103, −0.141	0.044, 0.117, 0.157	0.317
4	−0.015, −0.069, −0.065	0.044, 0.206, 0.194	0.320
5	−0.015, −0.069, −0.068	0.043, 0.205, 0.201	0.314
6	−0.099, −0.114, −0.262	0.008, 0.013, 0.021	0.318
7	−0.044, −0.122, −0.140	0.044, 0.121, 0.138	0.280
8	−0.015, −0.069, −0.066	0.044, 0.206, 0.197	0.317
9	−0.049, −0.024, −0.211	0.024, 0.012, 0.101	0.316
10	−0.142 −0.351, −0.067	−0.028, −0.061, −0.010	0.261
11	−0.165, −0.256, −0.296	−0.013, −0.014, −0.021	0.279
12	−0.119, −0.233, −0.232	0.015, 0.031, 0.030	0.331
13	−0.201, −0.405, −0.080	−0.092, −0.180, −0.037	0.324
14	−0.049, −0.024, −0.211	0.024, 0.012, 0.101	0.316
15	−0.263, −0.064, −0.466	−0.132, −0.033, −0.223	0.263
16	−0.171, −0.378, −0.073	−0.060, −0.120, −0.024	0.293
17	−0.142, −0.244, −0.264	0.001, 0.008, 0.005	0.305
18	−0.099, −0.114, −0.262	0.008, 0.013, 0.021	0.318
19	−0.561, −0.150, −0.466	−0.464, −0.125, −0.348	0.338
20	−0.412, −0.107, −0.466	−0.298, −0.079, −0.285	0.301

Table 3.2: Estimators of transition probabilities.

	PSW				CSW			
	$\widehat{a}^{(i)}_{y_T^{(i)},0}$		$\widehat{a}^{(i)}_{y_T^{(i)},1}$		$\widetilde{a}^{(i)}_{y_T^{(i)},0}$		$\widetilde{a}^{(i)}_{y_T^{(i)},1}$	
i	$M1$	$M2$	$M1$	$M2$	$M1$	$M2$	$M1$	$M2$
1	0.389	0.531	0.387	0.469	0.361	0.561	0.341	0.439
2	0.222	0.694	0.200	0.306	0.167	0.684	0.200	0.316
3	0.889	0.327	0.800	0.673	0.806	0.342	0.777	0.658
4	0.438	0.612	0.303	0.388	0.369	0.663	0.241	0.337
5	0.357	0.612	0.286	0.388	0.329	0.663	0.233	0.337
6	0.579	0.449	0.633	0.551	0.665	0.469	0.635	0.531
7	0.667	0.408	0.706	0.592	0.708	0.403	0.706	0.597
8	0.700	0.286	0.821	0.714	0.651	0.337	0.763	0.663
9	0.733	0.449	0.676	0.551	0.751	0.429	0.699	0.571
10	0.625	0.408	0.697	0.592	0.646	0.459	0.639	0.541
11	0.438	0.633	0.273	0.367	0.413	0.602	0.298	0.398
12	0.733	0.429	0.706	0.571	0.672	0.429	0.692	0.571
13	0.786	0.449	0.686	0.551	0.726	0.480	0.633	0.520
14	0.769	0.408	0.722	0.592	0.751	0.429	0.699	0.571
15	0.571	0.306	0.800	0.694	0.523	0.327	0.767	0.674
16	0.333	0.490	0.419	0.510	0.314	0.531	0.364	0.469
17	0.611	0.429	0.677	0.571	0.630	0.413	0.697	0.587
18	0.750	0.490	0.636	0.510	0.665	0.469	0.635	0.531
19	0.273	0.653	0.237	0.347	0.400	0.653	0.252	0.347
20	0.474	0.347	0.733	0.653	0.562	0.337	0.758	0.663

Furthermore, we compute performance measures ACC, BAC, GAC for each AUs and present their spatially averaged values in Table 3.3.

Table 3.3: Aggregated performance of GHMM model classifiers based on PBDF for simulated data.

lr		PSW1	CSW1	PSW2	CSW2
	ACC	0.8000	0.9500	0.8000	0.8500
0.2	BAC	0.7619	0.9643	0.8095	0.8452
	GAC	0.7559	0.9636	0.8092	0.8452
	ACC	0.8500	0.8500	0.8000	0.9000
0.4	BAC	0.8434	0.8434	0.7980	0.8990
	GAC	0.8409	0.8409	0.7977	0.8989
	ACC	0.8000	0.8000	0.8000	0.8500
0.5	BAC	0.8125	0.8125	0.8125	0.8542
	GAC	0.8101	0.8101	0.8101	0.8539
	ACC	0.8000	0.8500	0.8000	0.8500
0.6	BAC	0.8081	0.8535	0.8081	0.8535
	GAC	0.8040	0.8528	0.8040	0.8528
	ACC	**0.9000**	**0.8500**	0.9500	0.9500
0.8	BAC	**0.8810**	**0.8452**	0.9167	0.9167
	GAC	**0.8797**	**0.8452**	0.9129	0.9129

As we can see from Table 3.3 classifiers based on GHMM with CSW level in majority cases for both estimation rules have an advantage over one with PSW level by ACC as well BAC and GAC performance measures for almost all lr values. Exception to this statement is depicted in the table by bold numbers.

To be specific for the rule M1 CSW shows advantage PSW for all lr values, except $lr = 0.8$ and for the rule M2 CSW shows advantage PSW for all lr values.

3.1.2 Realization of Gaussian Markov Random Field-Autoregressive model for simulated data

For constructing a classifier for spatial GMRF data, the pseudo algorithm for the proposed supervised generative classifier for the simulation of spatial data sample is as follows:

Algorithm 2 Simulation spatial data sample.

1: **Input**: A network graph with 20 vertices with spatial weights W, number of points time $T+1$, number of time iterations v:
2: **for** $i = 1$ to 20 **do**
3: calculate the estimators of regression parameters $\widehat{\beta}^{(i)}_{(t)}$, variance $\widehat{\sigma}^2_{s(i)}$, $AR(1)$ $\widehat{\alpha}^{(i)}_1$, $\widehat{\sigma}^2_{T(i)}$
4: calculate the local model parameter estimators $\widetilde{\beta}^{(i)}$, $\widetilde{\sigma}^2_s$,
5: calculate Global Moran's I $I(t)$, Geary's C $C(t)$, Getis-Ord G $G(t)$
6: calculate PBDF $\widehat{W}_Z\left(z^{(i)}_{T+1}\right)$
7: calculate accuracy rates ACC, BAC
8: **end for**

The values of performance measures for the proposed classifiers specified in (1.1), (1.2), (1.3) for three label distribution models and various class label variables are presented in Table 3.4.

Table 3.4: Performance measures of GMRF-AR model classifiers based on PBDF simulation data.

lr		**GM**	**GO**	**GC**
0.2	ACC	0.8000	0.8000	0.8000
	BAC	0.8333	0.8333	0.8333
0.4	ACC	0.9000	0.9000	**0.8500**
	BAC	0.8889	0.8889	**0.8333**
0.5	ACC	0.8000	0.8000	0.8000
	BAC	0.7500	0.7500	0.7500
0.6	ACC	0.7000	0.7000	0.7000
	BAC	0.7857	0.7857	0.7857
0.8	ACC	0.8000	**0.7500**	0.8000
	BAC	0.8462	**0.8077**	0.8462

Table 3.4 shows the performance of classifiers based on three association indices exceeding 0.75 in more than 90% of cases.

For this table of results it is observed that performance of classifiers based different association indices does not differ little in almost all cases. Exception to this statement is depicted in the table by bold numbers.

3.1.3 Realization of Gaussian Geostatistical-Autoregressive model for simulated data

For constructing a classifier for spatial GGM data, the pseudo algorithm for the proposed supervised generative classifier for the simulation of spatial data sample is as follows:

Algorithm 3 Simulation spatial data sample.

1: **Input**: A network graph with 20 vertices with spatial weights W, number of points time $T+1$, number of time iterations v:
2: **for** $i = 1$ to 20 **do**
3: calculate the estimators of regression parameters $\widehat{\beta}_{(t)}^{(i)}$, variance $\widehat{\sigma}_{s(i)}^2$, $AR(1)$ $\widehat{\alpha}_1^{(i)}$, $\widehat{\sigma}_{T(i)}^2$
4: calculate the local model parameter estimators $\widetilde{\beta}^{(i)}$, $\widetilde{\sigma}_s^2$,
5: calculate probabilities $\pi_{1t}(s_i, T+1)$, $\pi_{1ts}(s_i, T+1)$, $\pi_{1s}(s_i, T+1)$
6: calculate PBDF $\widehat{W}_Z\left(z_{T+1}^{(i)}\right)$
7: calculate accuracy rates ACC, BAC
8: **end for**

The values of performance measures for the proposed classifiers specified in (1.1), (1.2), (1.3) for three label distribution models and various class label variables are presented in Table 3.5.

Table 3.5: Performance measures of GGM-AR model classifiers based on PBDF simulation data.

lr		TWMA	STWMA	STUA
0.2	ACC	**0.8500**	0.6500	0.8000
	BAC	**0.6373**	0.5625	0.5000
0.4	ACC	**0.8500**	**0.8500**	0.7000
	BAC	**0.8542**	**0.8434**	0.7033
0.5	ACC	0.8000	**0.8000**	0.7000
	BAC	0.7879	**0.8000**	0.8000
0.6	ACC	0.8500	**0.9000**	0.7000
	BAC	0.8333	**0.9286**	0.6667
0.8	ACC	0.9000	**1.0000**	0.9000
	BAC	0.5000	**1.0000**	0.8750

As can be seen from Table 3.5, best cases with highest values of performance measures is indicated in bold numbers. The method STWMA has advantage against two rival methods for $lr = 0.5, 0.6, 0.8$. For $lr = 0.2$ method TWMA shows advantage against two rival methods.

3.1.4 Realization of spatial Gaussian Geostatistical-Autoregressive model for simulated data

For numerical illustrations of obtained results we considered the Gaussian ST model with pure spatial covariances belonging to the family of powered-exponential isotropic models and with pure temporal covariance of $AR(1)$ model. It is known that for this model $c_T^{1,1} = c_T^{t,t}$ for $t = 2, \ldots, T+1$ parameter α quantifies temporal dependency by equation $c_T^{1,1} = \frac{\sigma_T^2}{1-\alpha^2}$ where σ_T^2 is the white noise variance.

Then $c_T^{T+1} = \frac{\sigma_t^2}{1-\alpha^2}(\alpha^T \alpha^{T-1} \ldots \alpha)$. It is easy to derive that $(c_T^{T+1})' C_T^{-1} = \alpha e_T^{'}$ and $\rho_{T+1} = \sigma_T^2$, where $e_T^{'}$ denotes the T-th row of identity matrix I_T.

Hence $\mu_{li(z)}^{T+1} = \beta_l^{'} x_i + \alpha\big(e_T^{'} \otimes e_i^{'}\big) vec(E)$ and $\Sigma_{T+1,i(z)} = c_S^{ii}\sigma_T^2$.

Where α is $AR(1)$ model parameter that quantifies temporal dependency and σ_T^2 is the white noise variance for this model.

In the study, the following isotropic nugetless spatial covariance structures belonging to the powered-exponential family are considered. So denoting by $C_s = \sigma_s^2 R$ the spatial covariance matrix, where $R = (r_{ij})$ spatial correlation matrix. The following two particular cases are considered:

1. exponential case with $r_{ij} = r(|s_i - s_j|) = e^{-|s_i - s_j|/\varphi}$;
2. squared-exponential case $r_{ij} = r(|s_i - s_j|) = e^{-(|s_i - s_j|/\varphi)^2}$.

Here φ is the so called range parameter that represents the spatial dependence.

This choice of 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 (TWMA, STWMA) for prior class probabilities is proposed.

We have compared these four particular cases by calculating the PBE_i and PIC_i for $i = 1, \dots, n$, and have presented them in tables.

Class labels for 20 locations and four time points in training sample is presented in Table 3.6.

Table 3.6: Class labels for 20 locations (i) at four time moments (t).

t	i									
	1	2	3	4	5	6	7	8	9	10
1	0	1	0	1	1	0	0	1	0	0
2	0	0	0	0	1	0	1	0	0	0
3	0	0	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	0	0	1	0
t	i									
	11	12	13	14	15	16	17	18	19	20
1	0	0	0	0	0	0	0	1	0	0
2	0	0	1	0	0	0	0	0	0	0
3	0	0	1	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0	1	0	0

Set $\Delta = \left|\mu_{1i(z)}^{T+1} - \mu_{0i(z)}^{T+1}\right|$.

Local Bayes error rates PBE_i and their averages $APBE$ for two cases of spatial covariances and two models for prior probabilities are presented in Table 3.7.

As it might be seen from Table 3.7, for $\alpha = 0.1, 0.3$ classifiers with STWMA priors in majority locations have an advantage against cases with TWMA priors for both spatial covariance models. For large α values, significant difference between these two is not observed.

For $v = 30$ independent replications, local empirical error rates PIC_i and their averages $APIC = \sum_{i=1}^{20} PIC_i/20$, for two cases of spatial covariances and two models for prior class probabilities are presented in Table 3.8.

Table 3.7: Local and average Bayes error rates for various α, $\Delta = 1$, $\varphi = 3$.

	$\pi_{1t}(s_i, T+1)$				$\pi_{1ts}(s_i, T+1)$			
	α=0.1	α=0.3	α=0.5	α=0.7	α=0.1	α=0.3	α=0.5	α=0.7
i	exponential case							
1	0	0	0	0.23	0	0.01	0.1	0.3
2	0	0	0.02	0.02	0	0	0	0.19
3	0	0	0	0.35	0	0.02	0.16	0.35
4	0.02	0.02	0.02	0.32	0.02	0.02	0.15	0.32
5	0	0.23	0.23	0.23	0	0.11	0.28	0.28
6	0	0	0	0.32	0	0	0.14	0.34
7	0	0.11	0.35	0.35	0	0.02	0.18	0.35
8	0.02	0.02	0.32	0.32	0	0.11	0.28	0.28
9	0	0	0.32	0.32	0	0	0.12	0.35
10	0	0	0	0	0	0	0.04	
11	0	0	0	0.02	0	0	0	0
12	0	0	0	0	0	0	0	0
13	0	0.35	0.35	0.35	0	0.19	0.32	0.32
14	0	0	0	0.32	0	0	0.1	0.35
15	0	0	0	0	0	0	0	0
16	0	0	0.23	0.23	0	0.15	0.31	0.31
17	0	0	0	0	0	0	0	0
18	0.02	0.02	0.35	0.35	0	0	0.2	0.35
19	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0
$APBE$	0.003	0.038	0.11	0.187	0.001	0.032	0.119	0.207
	$\pi_{1t}(s_i, T+1)$				$\pi_{1ts}(s_i, T+1)$			
	α=0.1	α=0.3	α=0.5	α=0.7	α=0.1	α=0.3	α=0.5	α=0.7
i	squared-exponential case							
1	0	0	0	0.23	0	0.02	0.17	0.32
2	0	0	0.02	0.02	0	0	0	0.23
3	0	0	0	0.35	0	0.02	0.17	0.35
4	0.02	0.02	0.02	0.32	0.02	0.02	0.17	0.32
5	0	0.23	0.23	0.23	0	0.11	0.28	0.28
6	0	0	0	0.32	0	0	0.17	0.34
7	0	0.11	0.35	0.35	0	0.02	0.17	0.35
8	0.02	0.02	0.32	0.32	0	0.11	0.28	0.28
9	0	0	0.32	0.32	0	0	0.11	0.35
10	0	0	0	0	0	0	0.06	0.06
11	0	0	0.02	0	0	0	0	0
12	0	0	0	0	0	0	0	0
13	0	0.35	0.35	0.35	0	0.17	0.32	0.32
14	0	0	0	0.32	0	0	0.11	0.35
15	0	0	0	0	0	0	0	0
16	0	0	0.23	0.23	0	0.17	0.32	0.32
17	0	0	0	0	0	0	0	0
18	0.02	0.02	0.35	0.35	0	0	0.17	0.34
19	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0
$APBE$	0.003	0.038	0.11	0.187	0.001	0.032	0.125	0.211

Table 3.8: Local and average empirical error rates for various α, $\Delta = 1$, $\varphi = 3$.

	$\pi_{1t}(s_i, T+1)$				$\pi_{1ts}(s_i, T+1)$			
	α=0.1	α=0.3	α=0.5	α=0.7	α=0.1	α=0.3	α=0.5	α=0.7
i	exponential case							
1	0	0	0.03	0	0	0	0.03	0
2	0	0	0.03	0.10	0	0	0.03	0
3	0	0	0.03	0	0	0	0	0
4	0	0	0.03	0	0	0	0.03	0
5	0	0	0	0	0	0	0	0
6	0	0	0.03	0	0	0	0.03	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0
10	0	0	0.07	0.17	0	0	0.07	0.10
11	0	0	0.03	0.33	0	0	0.03	0.33
12	0	0	0	0.23	0	0	0	0.23
13	0	0	0	0	0	0	0	0
14	0	0	0	0.13	0	0	0	0.03
15	0	0	0.13	0.37	0	0	0.13	0.37
16	0	0	0	0	0	0	0	0
17	0	0	0.10	0.23	0	0	0.10	0.23
18	0	0	0	0	0	0	0	0
19	0	0	0.07	0.27	0	0	0.07	0.27
20	0	0	0.13	0.20	0	0	0.13	0.20
ALO	0	0	0.034	0.102	0	0	0.033	0.088
	$\pi_{1t}(s_i, T+1)$				$\pi_{1ts}(s_i, T+1)$			
i	α=0.1	α=0.3	α=0.5	α=0.7	α=0.1	α=0.3	α=0.5	α=0.7
	squared-exponential case							
1	0	0	0.03	0	0	0	0.03	0
2	0	0	0.03	0.10	0	0	0.03	0
3	0	0	0.03	0	0	0	0	0
4	0	0	0.03	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0.03	0	0	0	0.03	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0
10	0	0	0.07	0.17	0	0	0.07	0.10
11	0	0	0.03	0.33	0	0	0.03	0.33
12	0	0	0	0.23	0	0	0	0.23
13	0	0	0	0	0	0	0	0
14	0	0	0	0.13	0	0	0	0
15	0	0	0.13	0.37	0	0	0.13	0.37
16	0	0	0	0	0	0	0	0
17	0	0	0.10	0.23	0	0	0.10	0.23
18	0	0	0	0	0	0	0.03	0
19	0	0	0.07	0.27	0	0	0.07	0.27
20	0	0	0.13	0.20	0	0	0.13	0.20
ALO	0	0	0.034	0.102	0	0	0.033	0.087

As it might be seen from Table 3.8 for all values of α, classifiers with STWMA and TWMA in majority locations have the similar empirical error rates for both spatial covariance models.

The last raw of Tables 3.7, 3.8 (i.e. $APBE$, $APIC$) allow us to compare averages of Bayes and empirical error rates for various combinations of spatial covariance and prior class probability models and to make optimal decisions in construction for the classifiers of ST Gaussian data.

3.2 Real data

The numerical analysis of annual mortality data collected by the Institute of Hygiene of the Republic of Lithuania from the 60 municipalities in the period 2001-2019 is carried out.

The gross death rate for each parish, measured in units per one hundred thousand population is considered as variable Z. We consider six label variables indicated by the threshold for mortality index from an Acute cardiovascular event (ACE) and Diseases of the circulatory system (CSD).

Municipality neighbours number in Table 3.9 and each neighbours in Appendix A.

Cases with index values less than the threshold have a label value equal to 0 and a label value equal to 1. Here we consider the constant mean case, i.e. $\mu(s;t) = \beta_l$, with $x(s) = 1$.

Data in the period from 2001 to 2018 ($t = 1, \ldots, 18$) is used for training and remaining data (period 2019) is used for testing. Hence $T = 18$ and $n = 60$, i.e. we consider $18 \cdot 60$ observations for training and 60 for testing.

Here IR values for testing data are calculated by changing the thresholds in specification of mortality levels.

The graph of IR_t is depicted in Table 3.10.

Table 3.9: List municipalities and number of neighbours $card(N_i)$.

Municipality	$\boldsymbol{card(N_i)}$	**Municipality**	$\boldsymbol{card(N_i)}$	**Municipality**	$\boldsymbol{card(N_i)}$
Akmenės r.	4	Klaipėdos r.	7	Skuodo r.	3
Alytaus m.	1	Kretingos r.	4	Šakių r.	4
Alytaus r.	8	Kupiškio r.	4	Šalčininkų r.	3
Anykščių r.	6	Lazdijų r.	4	Šiaulių m.	1
Birštono	2	Marijampolės	6	Šiaulių r.	7
Biržų r.	4	Mažeikių r.	4	Šilalės r.	6
Druskininkų	3	Molėtų r.	6	Šilutės r.	4
Elektrėnų	4	Neringos	1	Širvintų r.	6
Ignalinos r.	4	Pagėgių	3	Švenčionių r.	4
Jonavos r.	5	Pakruojo r.	5	Tauragės r.	6
Joniškio r.	3	Palangos m.	2	Telšių r.	7
Jurbarko r.	5	Panevėžio m.	1	Trakų r.	8
Kaišiadorių r.	7	Panevėžio r.	9	Ukmergės r.	6
Kalvarijos	3	Pasvalio r.	3	Utenos r.	6
Kauno m.	2	Plungės r.	6	Varėnos r.	4
Kauno r.	9	Prienų r.	7	Vilkaviškio r.	4
Kazlų Rūdos	5	Radviliškio r.	6	Vilniaus m.	2
Kėdainių r.	6	Raseinių r.	6	Vilniaus r.	7
Kelmės r.	6	Rietavo r.	4	Visagino m.	2
Klaipėdos m.	2	Rokiškio r.	5	Zarasų r.	4

Table 3.10: Annual imbalance ratio for various mortality reasons (class label variables).

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
ACE	**0.36**	**0.40**	**0.62**	**0.40**	**0.46**	**0.43**	**0.30**	**0.36**	**0.54**	**0.22**
CSD	**0.94**	**0.67**	**0.76**	**0.62**	**0.62**	**0.46**	**0.46**	**0.62**	**0.71**	1.07

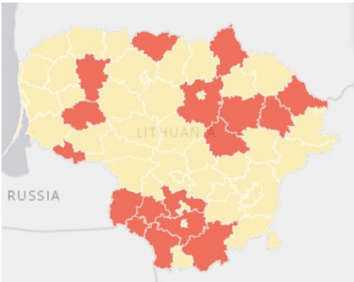
	2011	2012	2013	2014	2015	2016	2017	2018	2019
ACE	**0.28**	**0.25**	**0.22**	**0.17**	**0.09**	**0.07**	**0.09**	**0.05**	**0.01**
CSD	1.07	**0.71**	**0.82**	**0.67**	**0.94**	**0.62**	**0.58**	**0.40**	**0.22**

As we can see in Table 3.10 in majority of time periods CSD has the highest IR and the ACE has the lowest IR.

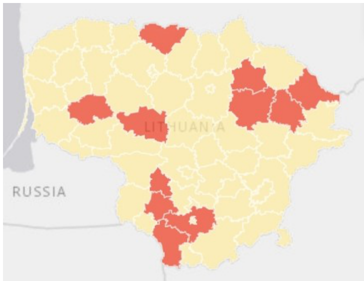
Numerical illustrations are performed on 60 areas in two dimensional areas that are shown in Figure 3.2 and labels changes in the training sample over time shown in Appendix B.

(1) ACE in 2018

(2) ACE in 2019



(3) CSD in 2018



(4) CSD in 2019

Figure 3.2: Classified Lithuanian municipalities. Yellow color areas indicate municipalities with low level of mortality due to disease ACE or CSD (with label value 0) and red areas indicates municipalities with high level of mortality due to disease ACE or CSD (with label value 1).

3.2.1 Realization of Gaussian Hidden Markov model for real data

Then for $i = 1, \ldots, 60$

$$X^{(i)} = \begin{pmatrix} 1 - y_1^{(i)} & y_1^{(i)} \\ 1 - y_2^{(i)} & y_2^{(i)} \\ \vdots & \vdots \\ 1 - y_T^{(i)} & y_T^{(i)} \end{pmatrix}.$$

The pseudo algorithm for the proposed supervised generative classifier for real spatial data sample is as follows:

Algorithm 4 Real spatial data sample.

Input: graph with 60 AUs with spatial weights matrix W, number of points time $T+1$:
for $i = 1$ to 60 **do**
  calculate the estimators of regression parameters $\widehat{\beta}^{(i)}$, variance $\widehat{\sigma}_i^2$, transition probabilities $\widehat{a}_{kl}^{(i)}$, $\widetilde{a}_{y_T^{(i)},l}^{(i)}$
  calculate the local model parameter estimators $\widetilde{\beta}^{(i)}$, $\widetilde{\sigma}_i^2$, $\widetilde{a}_{y_T^{(i)},l}^{(i)}$
  calculate PBDF $\widehat{W}_Z\left(z_{T+1}^{(i)}\right)$
  calculate accuracy rates ACC, BAC, GAC
end for

The values of performance measures for the proposed classifiers specified in (1.1), (1.2), (1.3), TPR and TNR are presented in Table 3.11.

Table 3.11: Performance measures of GHMM models classifiers based on PBDF real data.

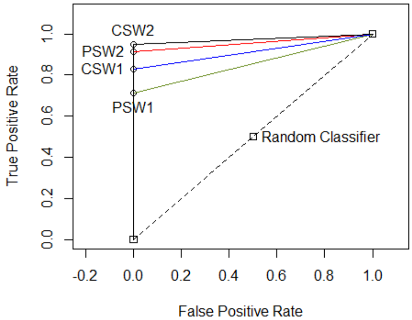
		PSW1	**CSW1**	**PSW2**	**CSW2**
ACE	ACC	0.7167	0.8333	0.9167	0.9500
	BAC	0.8559	0.9153	0.9576	0.9746
	GAC	0.8437	0.9113	0.9567	0.9742
	TPR	0.7119	0.8305	0.9153	0.9492
	TNR	1.0000	1.0000	1.0000	1.0000
CSD	ACC	0.6833	0.8000	**0.8500**	**0.7833**
	BAC	0.6651	0.7718	**0.8377**	**0.7616**
	GAC	0.6645	0.7705	**0.8374**	**0.7608**
	TPR	0.6939	0.8163	**0.9571**	**0.7959**
	TNR	0.6364	0.7273	**0.8182**	**0.7273**

As it might be seen from Table 3.11, classifiers based on GHMM with CSW level in majority cases for both estimation rules have an advantage over one with PSW level by ACC as well BAC and GAC performance measures for all diseases. Exception to this statement is depicted in the table by bold numbers.

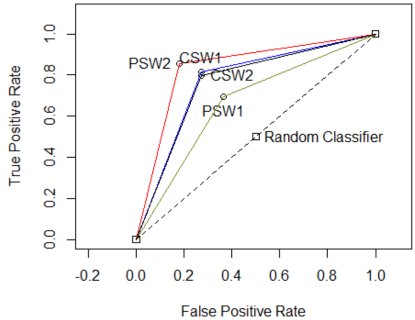
For disease ACE for the rules M1 and M2 CSW shows advantage PSW.

For disease CSD for the rule M1 CSW shows advantage PSW and for the rule M2 PSW shows advantage CSW.

As can be seen from Fig. 3.3, the classifier CSW2 shows advantage against others since has the largest area under the curve that is equal 0.9746 (see in Table 3.11).



(1) ACE



(2) CSD

Figure 3.3: ROC plots for the problems of GHMM models with class label variables.

3.2.2 Realization of Gaussian Markov Random Field-Autoregressive model for real data

For numerical illustrations of obtained results we considered the ST Gaussian model with pure spatial covariances characterizes spatial weights and with pure temporal covariance of stationary $AR(1)$.

Then for $t = 1, \dots, 18$

$$X_{(t)} = \begin{pmatrix} 1 - y_t^{(1)} & y_t^{(1)} \\ 1 - y_t^{(2)} & y_t^{(2)} \\ \vdots & \vdots \\ 1 - y_t^{(n)} & y_t^{(n)} \end{pmatrix}.$$

The pseudo algorithm for the proposed supervised generative classifier for real spatial data sample is as follows:

Algorithm 5 Real spatial data sample.

Input: graph with 60 AUs with spatial weights W, number of points time $T+1$:
for $i = 1$ to 60 **do**
    calculate the estimators of regression parameters $\widehat{\beta}_{(t)}^{(i)}$, variance $\widehat{\sigma}_{s(t)}^{2}$, $AR(1)$ $\widehat{\alpha}_{1}^{(i)}$, $\widehat{\sigma}_{T(i)}^{2}$
    calculate the local model parameter estimators $\widetilde{\beta}^{(i)}$, $\widetilde{\sigma}_{s}^{2}$
    calculate Global Moran's I $I(t)$, Geary's C $C(t)$, Getis-Ord G $G(t)$
    calculate PBDF $\widehat{W}_Z\left(z_{T+1}^{(i)}\right)$
    calculate accuracy rates ACC, BAC, GAC
end for

The values of performance measures for the proposed classifiers specified in (1.1), (1.2), (1.3), TPR and TNR are presented in Table 3.12.

Table 3.12: Performance measures of GMRF-AR model classifiers based on PBDF real data.

		GM	**GO**	**GC**
ACE	ACC	**0.9667**	**0.9667**	**0.9667**
	BAC	**0.9831**	**0.9831**	**0.9831**
	GAC	**0.9829**	**0.9829**	**0.9829**
	TPR	**0.9661**	**0.9661**	**0.9661**
	TNR	**1.0000**	**1.0000**	**1.0000**
CSD	ACC	0.8000	**0.8333**	**0.8333**
	BAC	0.7718	**0.7922**	**0.7922**
	GAC	0.7705	**0.7895**	**0.7895**
	TPR	0.8163	**0.8571**	**0.8571**
	TNR	0.7273	**0.7273**	**0.7273**

Table 3.12 shows the performance of classifiers based on three association indices exceeding 0.75 in more than 90% of cases.

For this table of results, it is observed that the performance of classifiers, based on different association indices, does not differ little in almost all cases. Exception to this statement is depicted in the table by bold numbers.

3.3 Conclusions of section

This section presents the results of numerical experiments with simulated data and the application to the real data for the novel method to generative supervised classification based on three models of Gaussian ST data models corresponding different distributions of feature and class label distributions.

For the GHMM analysis of simulated and real data also show the classifiers based on GHMM with CSW level in majority cases for both estimation rules have an advantage over one with PSW level by all performance measures for almost all lr values.

For the GHMM analysis of real data also show for disease ACE for the rules M1 and M2 CSW shows advantage PSW, the classifier shows advantage against others since has the largest area under the curve that is equal 0.9746 (BAC performance measure). For disease CSD for the rule M1 CSW shows advantage PSW and for the rule M2 PSW shows advantage CSW.

For the GMRF analysis of simulated and real data also show the performance of classifiers based on three association indices exceeding 0.75 in more than 90% of cases. It is observed that the performance of classifiers, based on different association indices, does not differ little in almost all cases.

For the GGM analysis of simulated data also show the method STWMA has advantage against two rival methods for $lr = 0.5, 0.6, 0.8$. For $lr = 0.2$ method TWMA shows advantage against two rival methods.

Numerical analysis of classifiers based on GGM-AR shows that:

- incorporation spatial correlation in class prior probabilities improves the performances of classifiers;
- classifiers with spatial exponential covariance has an advantage against classifiers with squared-exponential covariance.

GENERAL CONCLUSIONS

This thesis proposes the novel supervised Bayesian classification methods of the Gaussian ST data based on generative machine learning models. The research focuses on three types of ST models characterized by different statistical dependence forms and various strategies for the estimation of feature and class label distribution parameters. This is a significant extension of similar studies performed previously in the purely spatial context. Sensitivity analysis and detailed comparison of classifiers within every type of data models, specified by different levels of incorporated spatial and temporal information, were conducted in terms of several performance measures.

The conclusions, obtained during the conducted research, are listed below.

1. For a GHMM type, the calculation results, with the simulated and real data, showed that the spatial weighting level and the rules of transition probabilities' estimation influenced the performance of the proposed classifiers.
2. For a GMRF-AR data model type, the calculations show the performance of classifiers based on three association indices exceeding 0.75 in more than 90% of cases. It is observed that the performance of classifiers, based on different association indices, does not differ little in almost all cases.
3. The proposed method for a supervised Bayesian classification for GGM-AR, with separable covariances for feature variables and independent Bernoulli variables for labels, reveals the priorities of cases with the combined weighted ST context against the ones with a purely temporal and unweighted ST context.
4. Numerical analysis of classifiers, based on GGM-AR, shows that

the classifiers with the spatial exponential covariance has an advantage against those with the squared-exponential covariance.

5. Derived closed forms for Local Bayes error rates and their D-estimators' probability can be directly used in designing of the supervised classifiers for the selected types of ST data models.

6. Detailed comparison the performances of classifiers within every type of data models specified by different scenarios of incorporated spatial and temporal information highlights the advantages of one against others and reveals new application areas and recommendations for users.

Future research directions:

First, there is further scope for exploring techniques supervising classification due to ST Gaussian data models to multiclass case.

Second, future research may also include implementation of the proposed classification technique in the context of other continuous non-Gaussian models of observations.

Third, in future research, it is reasonable to consider second-order or higher order Markov chain models for labels instead of first-order Markov chain. That will enable us to incorporate more information on stochastic temporal dependence.

Bibliography

[1] D. Allcroft and C. Glasbey. A Latent Gaussian Markov Random Field Model for Spatio-Temporal Rainfall Disaggregation. *Journal of the Royal Statistical Society Series C*, 52:487–498, 10 2003. doi: 10.1111/1467-9876.00419.

[2] P. M. Atkinson and P. Lewis. Geostatistical classification for remote sensing: an introduction. *Computers & Geosciences*, 26(4):361–371, 2000. doi: 10.1016/S0098-3004(99)00117-X.

[3] G. Atluri, A. Karpatne, and V. Kumar. Spatio-temporal data mining: a survey of problems and methods. *ACM Comput Surv: CSUR*, 51(4):83, 2018. doi: 10.1145/3161602.

[4] J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, M. West, C. Bishop, and J. Lasserre. Generative or Discriminative? Getting the Best of Both Worlds. *Baysian Statistics*, 8:3–24, 01 2007.

[5] C. Berrett and C. A. Calder. Bayesian spatial binary classification. *Spatial Statistics*, 16:72–102, 2016. doi: 10.1016/j.spasta.2016.01.004.

[6] J. Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974. doi: 10.1111/j.2517-6161.1974.tb00999.x.

[7] M. Blangiardo and M. Cameletti. *Spatial and Spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, 2015.

[8] M. Blangiardo, M. Cameletti, G. Baio, and H. Rue. Spatial and Spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, 7:39–55, 2013.

[9] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer Science & Business Media, New York, 2009. doi: 10.1007/978-1-4419-0320-4.

[10] P. E. Brown, K. F. Karesen, G. O. Roberts, and S. Tonellato. Blur-generated non-separable space-time models. *Journal of the Royal Statistical Society Ser. B*, 62(4):847–860, 2000. doi: 10.1111/1467-9868.00269.

[11] J. D. Bryan and S. E. Levinson. Autoregressive Hidden Markov

Model and the Speech Signal. *Procedia Comput. Sci.*, 61:328–333, 2015.

[12] M. Cameletti, F. Lindgren, D. Simpson, and H. Rue. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, 97, 2011. doi: 10.1007/s10182-012-0196-3.

[13] J. P. Chiles and P. Delfiner. Geostatistics: Modeling spatial uncertainty. *Mathematical Geosciences*, 45:377–380, 2013. doi: 10.1007/s11004-012-9429-y.

[14] B. Costes and J. Perret. A hidden Markov model for matching spatial networks. *J. Spatial Inf. Sci.*, 18:57–89, 2019. doi: 10.5311/JOSIS.2019.18.489.

[15] N. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. Wiley, 2011.

[16] S. De Iaco and D. Posa. Predicting spatio-temporal random fields: Some computational aspects. *Computers Geosciences*, 41:12–24, 2012. doi: 10.1016/j.cageo.2011.11.014.

[17] V. De Oliveira. Bayesian prediction of clipped Gaussian random fields. *Computational Statistics and Data Analysis*, 34:299–314, 2000. doi: 10.1016/S0167-9473(99)00103-6.

[18] V. De Oliveira and M. Ferreira. Maximum likelihood and restricted maximum likelihood estimation for a class of Gaussian Markov random fields. *Metrika*, 74:167–183, 09 2011. doi: 10.1007/s00184-009-0295-7.

[19] S. S. Demel and J. Du. Spatio-temporal models for some data sets in continuous space and discrete time. *Statistica Sinica*, 25:81–98, 2015. doi: 10.5705/ss.2013.223w.

[20] P. J. Diggle, P. J. Ribeiko, and O. F. Christensen. An introduction to model-based geostatistics. *Spat. Stat. Comput. Methods Lect. Notes Stat.*, 173:43–86, 2003.

[21] L. Dreižienė and K. Dučinskas. Comparison of spatial linear mixed models for ecological data based on the correct classification rates. *Spatial statistics*, 35, 2020. doi: 10.1016/j.spasta.2019.100395.

[22] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000.

[23] K. Dučinskas. Approximation of the expected error rate in classification of the Gaussian random field observations. *Statistics and Probability Letters*, 79:138–144, 2009. doi: 10.1016/j.spl.2008.07.042.

[24] K. Dučinskas. Error rates in classification of multivariate Gaussian random field observation. *Lith. Math. J.*, 51(4):477–485, 2011. doi: 10.1007/s10986-011-9142-4.

[25] K. Dučinskas and L. Dreižienė. Risks of classification of the Gaussian Markov random field observations. *Journal of Classification*,

35:422–436, 2018. doi: 10.1007/s00357-018-9269-7.

[26] K. Dučinskas and L. Dreižienė. Performance Evaluations of Gaussian Spatial Data Classifiers Based on Hybrid Actual Error Rate Estimators. *Aust. J. Stat.*, 49:27–34, 2020.

[27] K. Dučinskas and L. Dreižienė. Actual error rates in linear discrimination of spatial Gaussian data in terms of semivariograms. *Communications in Statistics - Theory and Methods*, 52(9):3165–3173, 2023. doi: 10.1080/03610926.2021.1968903.

[28] K. Dučinskas, M. Karaliutė, and L. Šaltytė–Vaisiauskė. Spatially Weighted Bayesian Classification of Spatio-Temporal Areal Data Based on Gaussian-Hidden Markov Models. *Mathematics*, 11(2), 2023. doi: 10.3390/math11020347.

[29] I. Egbo. Evaluation of error rate estimators in discriminant analysis with multivariate binary variables. *American Journal of Theoretical and Applied Statistics*, 5(4):173–179, 2016. doi: 10.11648/j.ajtas.20160504.12.

[30] L. P. Fatti, C. D. Elphinstone, and A. T. Lonergan. Methods for contextual classification of remotely sensed data. *APCOM 87. Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries*, 3:13–29, 1987.

[31] C. Ferri, J. Hernández-Orallo, and R. Modroiu. An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, 30(1):27–38, 2009. doi: https://doi.org/10.1016/j.patrec.2008.08.010.

[32] M. M. Fischer and J. Wang. *Spatial Data Analysis: Models, Methods, and Techniques*. Springer, New York, 2011. doi: 10.1007/978-3-642-21720-3.

[33] L. Fontanella, L. Ippoliti, R. J. Martin, and S. Trivisonno. Interpolation of spatial and spatio-temporal Gaussian fields using Gaussian Markov random fields. *Advances in Data Analysis and Classification*, 2(1):63–79, 2008. doi: 10.1007/s11634-008-0019-2.

[34] M. Fuentes. Testing for separability of spatial–temporal covariance functions. *Journal of Statistical Planning and Inference*, 136:447–466, 2006. doi: 10.1016/j.jspi.2004.07.004.

[35] A. Getis and J. K. Ord. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3):189–206, 1992.

[36] F. Giannotti and D. Pedreschi. *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*. Springer, 2008. doi: 10.1007/978-3-540-75177-9.

[37] W. Gong, S. Fang, and G. Y. M. Ge. Using a Hidden Markov Model for Improving the Spatial-Temporal Consistency of Time Series Land Cover Classification. *Int. J. Geo-Inf.*, 6:292, 2017.

[38] A. Hamdi, K. Shaban, A. Erradi, A. Mohamed, S. K. Rumi, and

F. D. Salim. Spatiotemporal data mining: a survey on challenges and open problems. *Artificial Intelligence Review*, 55:1441–1488, 2022. doi: 10.1007/s10462-021-09994-y.

[39] J. Haslett and A. E. Raftery. Space-time Modelling with Long memory Dependence: Assessing Ireland‘s Wind Power Resourse. *Applied Statistics*, 38:1–50, 1989. doi: 10.2307/2347679.

[40] B. Jeon and D. Landgrebe. Classification with spatio-temporal interpixel class dependency contexts. *IEEE Transactions on Geoscience and Remote Sensing*, 30(4):663–672, 1992. doi: 10.1109/36.158859.

[41] B. Jin, L. Cruz, and N. Gonçalves. Pseudo RGB-D Face Recognition. *IEEE Sensors Journal*, 22(22):21780–21794, 2022. doi: 10.1109/JSEN.2022.3197235.

[42] S. K. Kadhem, P. Hewson, and I. Kaimi. Using hidden Markov models to model spatial dependence in a network. *Aust. N. Z. J. Stat.*, 60(4):423–446, 2018. doi: 10.1111/anzs.12250.

[43] M. Karaliutė and K. Dučinskas. Classification of Gaussian spatio-temporal data with stationary separable covariances. *Nonlinear Analysis: Modelling and Control*, 26(2):363–374, 2021. doi: 10.15388/namc.2021.26.22359.

[44] M. Karaliutė and K. Dučinskas. Supervised linear classification of Gaussian spatio-temporal data. *Lietuvos matematikos rinkinys. Proc. of the Lithuanian Mathematical Society. Ser. A*, 62:9–154, 2021. doi: 10.15388/LMR.2021.25214.

[45] M. Karaliutė and K. Dučinskas. Performance of the supervised generative classifiers of spatio-temporal areal data using various spatial autocorrelation indexes. *Nonlinear Analysis: Modelling and Control*, 28(2):1–14, 2023. doi: 10.15388/namc.2023.28.31434.

[46] A. Lauraitis. *Hybrid classification model for neural impairment detection*. Doctoral dissertation Kauno technologijos universitetas, 2020.

[47] C. R. O. Lawoko and G. L. McLachlan. Discrimination with autocorrelated observations. *Pattern Recognition*, 18:145–149, 1985. doi: doi.org/10.1016/0031-3203(85)90038-X.

[48] N. D. Le and J. V. Zidek. *Statistical Analysis of Environmental Space – Time Processes*. Springer, New York, 2006.

[49] M. Li, S. Zang, B. Zhang, S. Li, and C. Wu. A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information. *European Journal of Remote Sensing*, 47(1): 389–411, 2014. doi: 10.5721/EuJRS20144723.

[50] T. Liogienė. *Hierarchical classification of speech emotions*. Doctoral dissertation Vilnius University, 2017.

[51] R. Liu, C. Men, X. juan Wang, F. Xu, and W. Yu. Application of spatial markov chains to the analysis of the temporal-spatial

evolution of soil erosion. *Water science and technology : a journal of the International Association on Water Pollution Research*, 74 5: 1051–1059, 2016.

[52] V. López, A. Fernández, S. García, V. Palade, and F. Herrera. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250:113–141, 2013. doi: 10.1016/j.ins.2013.07.007.

[53] V. López, A. Fernández, and F. Herrera. On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed. *Information Sciences*, 257:1–13, 2014. doi: 10.1016/j.ins.2013.09.038.

[54] C. Ma. Spatio-temporal stationary covariance models. *Journal of Multivariate Analysis*, 86(1):97–107, 2003. doi: 10.1146/annurev-statistics-042720-115603.

[55] F. Madadizadeh, A. Ghanbarnejad, V. R. Tabar, K. Batmanghelich, A. Bahrampour, and H. Zeraati. Introduction to Spatial First Order Discrete HMMing. *Ann Biom Biostat*, 4(1):1029, 2018.

[56] K. V. Mardia. Spatial discrimination and classification maps. *Communications in Statistics - Theory and Methods*, 13(18):2181–2197, 1984. doi: 10.1080/03610928408828822.

[57] B. Matern. *Spatial variation. Second edition*. Springer–Verlag, New York, 1986. doi: 10.1007/978-1-4615-7892-5.

[58] G. J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, New York, 2004.

[59] B. Mishra, A. Dahal, N. LuinteL, T. B. Shahi, S. Panthi, S. Pariyar, and B. R. Ghimire. Methods in the spatial deep learning: Current status and future direction. *Spat. Inf. Res.*, 30:215–232, 2022.

[60] K. P. Murphy. *Machine learning: a probabilistic perspective*. MA: MIT Press, Cambridge, 2012.

[61] A. Y. Ng and M. I. Jordan. *On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes*. NIPS 14, 2022.

[62] L. Nguyen. Continuous Observation Hidden Markov Model. *Revista Kasmera*, 44(6):65–149, 2016.

[63] L. Nguyen. Tutorial on Hidden Markov Model. *Special Issue "Some Novel Algorithms for Global Optimization and Relevant Subjects", Applied and Computational Mathematics (ACM)*, 6:16–38, 06 2017. doi: 10.11648/j.acm.s.2017060401.12.

[64] R. Nishii and S. Eguchi. Image classification based on Markov random field model with Jeffreys divergence. *Journal of Multivariate Analysis*, 97(9):1997–2008, 2006. doi: 10.1016/j.jmva.2006.01.009.

[65] A. Orriols-Puig and E. Bernadó-Mansilla. Evolutionary rule-based systems for imbalanced datasets. *Soft Computing*, 13(3):213–225,

2009. doi: 10.1016/j.jmva.2006.01.009.

[66] J. Ortigosa-Hernández, I. Inza, and J. A. Lozano. Measuring the class-imbalance extent of multi-class problems. *Pattern Recognition Letters*, 98:32–38, 2017. doi: 10.1016/j.patrec.2017.08.002.

[67] L. Paulionienė. *Statistical modeling of Spatio-temporal data based on spatial interpolation of time series parameters*. Doctoral dissertation Vilnius University, 2013.

[68] E. Pebesma. spacetime: Spatio-Temporal Data in R. *Journal of Statistical Software*, 51(7):1–30, 2012. doi: 10.18637/jss.v051.i07.

[69] A. Pettitt, I. Weir, and A. Hart. A Conditional Autoregressive Gaussian Process for Irregularly Spaced Multivariate Data with Application to Modelling Large Sets of Binary Data. *Statistics and Computing*, 12:353–367, 10 2002. doi: 10.1023/A:1020792130229.

[70] T. Pranckevičius. *Investigation of multi-class classification methods for textual data*. Doctoral dissertation Vilnius University, 2017.

[71] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2): 257–286, 1989. doi: 10.1109/5.18626.

[72] L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3:4–16, 1986.

[73] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications (1st ed.)*. Chapman and Hall, Boca Raton, 2005.

[74] S. Sain and N. Cressie. A Spatial Model for Multivariate Lattice Data. *Journal of Econometrics*, 140:226–259, 09 2007. doi: 10.1016/j.jeconom.2006.09.010.

[75] B. Sanso and L. Guenni. Venezuelan Rainfall Data Analysed by Using a Bayesian Space-time Model. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 48(3):345–362, 2002. doi: 10.1111/1467-9876.00157.

[76] E. Santos-Fernandez, J. M. Ver Hoef, E. E. Peterson, J. McGree, D. J. Isaak, and K. Mengersen. Bayesian spatio-temporal models for stream networks. *Computational Statistics Data Analysis*, 170: 107446, 2022. doi: 10.1016/j.csda.2022.107446.

[77] L. Scaccia and R. J. Martin. Testing axial symmetry and separability of lattice processes. *Journal of Statistical Planning and Inference*, 131:19–39, 2005.

[78] M. Sensoy, L. Kaplan, F. Cerutti, and M. Saleki. Uncertainty-Aware Deep Classifiers Using Generative Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(4):5620–5627, 2020. doi: 10.1609/aaai.v34i04.6015.

[79] S. Shekhar, P. R. Schrater, R. R. Vatsavai, W. Wu, and S. Chawla. Spatial contextual classification and prediction models for mining geospatial data. *IEEE Transactions on Multimedia*, 4(2):174–188,

2002. doi: 10.1109/TMM.2002.1017732.

[80] R. Soleymani, E. Granger, and G. Fumera. *F-measure curves: A tool to visualize classifier performance under imbalance*. Pattern Recognit., 2020.

[81] J. J. Song and V. D. Oliveira. Bayesian model selection in lattice models. *Statistical methodology*, 9:228–238, 2012. doi: 10.1016/j.statmet.2011.01.003.

[82] L. Spezia, N. Friel, and A. Gimona. Spatial hidden Markov models and species distributions. *Journal of Applied Statistics*, 45(9):1595–1615, 2018. doi: 10.1080/02664763.2017.1386771.

[83] G. Stabingis, K. Dučinskas, and L. Stabingienė. Comparison of spatial classification rules with different conditional distributions of class label. *Nonlinear Analysis: Modelling and Control*, 19(1): 109–117, 2014. doi: 10.15388/NA.2014.1.7.

[84] M. L. Stein. *Interpolation of Spatial Data*. Springer, 1999.

[85] C. Su and Y. Hasio. An evaluation of the robustness of MTS for imbalanced data. *IEEE transactions on knowledge and data engineering*, 19(10):1321–1332, 2007.

[86] S. Sun, P. Zhong, H. Xiao, and R. Wang. Spatial contextual classification of remote sensing images using a Gaussian process. *Remote Sensing Letters*, 7(2):131–140, 2016. doi: 10.1080/2150704X.2015.1117152.

[87] P. Switzer. Estimation of spatial distributions from point sources with application to air pollution measurement. *Bull. Internat. Statist. Inst.*, 47:123–137, 1977. doi: 10.1007/BF01029421.

[88] Y. Tang, L. Jing, H. Li, and P. M. Atkinson. A multiple-point spatially weighted k-NN method for object-based classification. *Int. J. Appl. Earth Obs. Geoinf.*, 52:263–274, 2016.

[89] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. 01 2009. doi: 10.1016/B978-1-59749-272-0.X0001-2.

[90] J. Tichonov. *Classification-based compression solutions of digital images*. Doctoral dissertation Vilnius University, Vilnius, 2018.

[91] G. Toussaint. Bibliography on estimation of misclassification. *IEEE Transactions on Information Theory*, 20(4):472–479, 1974. doi: 10.1109/TIT.1974.1055260.

[92] A. Tucker and D. Garway-Heath. The pseudotemporal bootstrap for predicting glaucoma from cross-sectional visual field data. *IEEE Transactions on Information Technology in Biomedicine*, 14(1): 79–85, 2010. doi: 10.1109/TITB.2009.2023319.

[93] M. D. Ugarte, T. Goicoa, and A. F. Militino. Spatio-temporal modeling of mortality risks using penalized splines. *Environmetrics*, 21 (3-4):270–289, 2010. doi: 10.1002/env.1011.

[94] S. Wang, J. Cao, and P. S. Yu. Deep Learning for Spatio-Temporal

Data Mining: A Survey. *arxiv.org*, 2019. doi: 10.1109/TKDE.2020.3025580.

[95] C. K. Wikle and M. B. Hooten. A general science-based framework for dynamical spatio-temporal models. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 19(3): 417–451, November 2010. doi: 10.1007/s11749-010-0209-z.

[96] C. K. Wikle, R. F. Milliff, D. Nychka, and L. M. Berliner. Spatiotemporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds. *Journal of the American Statistical Association*, 96(454):382–397, 2001.

[97] C. K. Wikle, A. Zammit-Mangion, and N. Cressie. *Spatio-Temporal Statistics with R*. Boca Raton, FL: Chapman and Hall/CRC, 2019.

[98] K. Xu and C. K. Wikle. Estimation of parameterized spatio-temporal dynamic models. *Journal of Statistical Planning and Inference*, 137(2):567–588, 2007. doi: https://doi.org/10.1016/j.jspi.2005.12.005.

[99] E. Yang, P. Ravikumar, G. Allen, and Z. Liu. Graphical Models via Generalized Linear Models. *NIPS*, 2012.

[100] L. You, H. Jiang, J. Hu, C. H. Chang, L. Chen, X. Cui, and M. Zhao. GPU-accelerated Faster Mean Shift with euclidean distance metrics. pages 211–216, 2022. doi: 10.1109/COMPSAC54236.2022.00037.

[101] M. Zhao, C. H. Chang, W. Xie, Z. Xie, and J. Hu. Cloud Shape Classification System Based on Multi-Channel CNN and Improved FDM. *IEEE Access*, 8:44111–44124, 2020. doi: 10.1109/ACCESS.2020.2978090.

[102] Q. Zheng, M. Yang, and J. Yang. Improvement of Generalization Ability of Deep CNN via Implicit Regularization in Two-Stage Training Process. *IEEE Access*, 6:15844–15869, 2018. doi: 10.1109/ACCESS.2018.2810849.

[103] Q. Zheng, X. Tian, N. Jiang, and M. Yang. Layer-wise learning based stochastic gradient descent method for the optimization of deep convolutional neural network. *Journal of Intelligent & Fuzzy Systems*, 37(4):5641–5654, 2019. doi: 10.3233/JIFS-190861.

[104] Q. Zheng, X. Tian, M. Yang, Y. Wu, and H. Su. PAC-Bayesian framework based drop-path method for 2D discriminative convolutional network pruning. *Multidimensional Systems and Signal Processing*, 31(3):793–827, 2020. doi: 10.1007/s11045-019-00686-z.

[105] Q. Zheng, M. Yang, X. Tian, N. Jiang, and D. Wang. A full stage data augmentation method in deep convolutional neural network for natural image classification. *Discrete Dynamics in Nature and Society*, 2020:1–11, 2020. doi: 10.1155/2020/4706576.

[106] Q. Zheng, M. Yang, J. Yang, Q. Zhang, and X. Zhang. Spectrum

interference-based two-level data augmentation method in deep learning for automatic modulation classification. *Neural Comput. Appl.*, 33:7723–7745, 2021. doi: 10.1007/s00521-020-05514-1.

[107] Z. Zhu and M. L. Stein. Spatial sampling design for prediction with estimated parameters. *J. Agric. Biol. Environ. Stat.*, 11:24–44, 2006. doi: 10.1007/s00521-020-05514-1.

[108] D. A. Zuanetti and L. A. Milan. Second-order autoregressive Hidden Markov Model. *Braz. J. Probab. Stat.*, 31:653–665, 2017. doi: 10.1214/16-BJPS328.

[109] W. Zucchini and I. L. MacDonald. *Hidden Markov Models for Time Series: An Introduction Using R (1st ed.)*. Chapman and Hall/CRC, 2009.

[110] J. Šaltytė–Benth and K. Dučinskas. Linear discriminant analysis of multivariate spatial–temporal regressions. *Scand. J. Stat.*, 32: 281–294, 2005.

[111] J. Šaltytė-Benth and L. Šaltytė. Spatial–temporal model for wind speed in Lithuania. *Journal of Applied Statistics*, 38(6):1151–1168, 2011. doi: 10.1080/02664763.2010.491857.

[112] T. Žvirblis. *Application of predictive and machine learning models for mechatronic systems parameters prediction and fault classification*. Doctoral dissertation Vilnius Gediminas Technical University, 2022.

APPENDICES

A APPENDIX - Lithuanian municipalities and neighbours list

Table A1: Lithuanian municipalities list and number of neighbours $card(N_i)$.

Municipality	$card(N_i)$	**Neighbours (*nearest neighbor)**
Akmenės r.	4	Joniškio r., Telšių r., Mažeikių r., *Šiaulių r.
Alytaus m.	1	Alytaus r.
Alytaus r.	8	Birštono, Trakų r., Druskininkų, Varėnos r., Lazdijų r., *Alytaus m., Prienų r., Marijampolės
Anykščių r.	6	*Kupiškio r., Molėtų r., Ukmergės r., Panevėžio r., Rokiškio r., Utenos r.
Birštono	2	Alytaus r., *Prienų r.
Biržų r.	4	*Kupiškio r., Panevėžio r., Rokiškio r., Pasvalio r.
Druskininkų	3	Varėnos r., Lazdijų r., *Alytaus r.
Elektrėnų	4	Širvintų r., *Trakų r., Kaišiadorių r., Vilniaus r.
Ignalinos r.	4	*Visagino m., Zarasų r., Utenos r., Švenčionių r.
Jonavos r.	5	Širvintų r., Ukmergės r., *Kaišiadorių r., Kėdainių r., Kauno r.
Joniškio r.	3	Pakruojo r., *Šiaulių r., Akmenės r.
Jurbarko r.	5	Pagėgių, Raseinių r., Tauragės r., *Šakių r., Kauno r.,

Table A1, continuation

Municipality	$card(N_i)$	**Neighbours (*nearest neighbor)**
Kaišiadorių r.	7	Širvintų r., Trakų r., Kauno m., Prienų r., *Jonavos r., Elektrėnų, Kauno r.
Kalvarijos	3	Lazdijų r., Vilkaviškio r., *Marijampolės
Kauno m.	2	Kaišiadorių r., *Kauno r.
Kauno r.	9	Kazlų Rūdos, *Kauno m., Prienų r., Jonavos r., Raseinių r., Kaišiadorių r., Kėdainių r., Jurbarko r., Šakių r.
Kazlų Rūdos	5	Prienų r., Vilkaviškio r., Šakių r., *Marijampolės, Kauno r.
Kėdainių r.	6	Radviliškio r., Ukmergės r., Panevėžio r., Jonavos r., Raseinių r., *Kauno r.
Kelmės r.	6	Telšių r., Radviliškio r., *Šiaulių r., Raseinių r., Tauragės r., Šilalės r.
Klaipėdos m.	2	*Neringos, Klaipėdos r.
Klaipėdos r.	7	*Klaipėdos m., Palangos m., Kretingos r., Rietavo r., Šilutės r., Plungės r., Šilalės r.
Kretingos r.	4	*Palangos m., Skuodo r., Klaipėdos r., Plungės r.
Kupiškio r.	4	*Anykščių r., Panevėžio r., Rokiškio r., Biržų r.
Lazdijų r.	4	*Druskininkų, Alytaus r., Marijampolės, Kalvarijos
Marijampolės	6	*Kazlų Rūdos, Lazdijų r., Alytaus r., Prienų r., Vilkaviškio r., Kalvarijos
Mažeikių r.	4	*Telšių r., Skuodo r., Akmenės r., Plungės r.
Molėtų r.	6	Anykščių r., Širvintų r., Ukmergės r., *Utenos r., Švenčionių r., Vilniaus r.
Neringos	1	*Klaipėdos m.
Pagėgių	3	Šilutės r., *Tauragės r., Jurbarko r.
Pakruojo r.	5	*Joniškio r., Radviliškio r., Panevėžio r., Pasvalio r., Šiaulių r.,
Palangos m.	2	Kretingos r., *Klaipėdos r.
Panevėžio m.	1	*Panevėžio r.

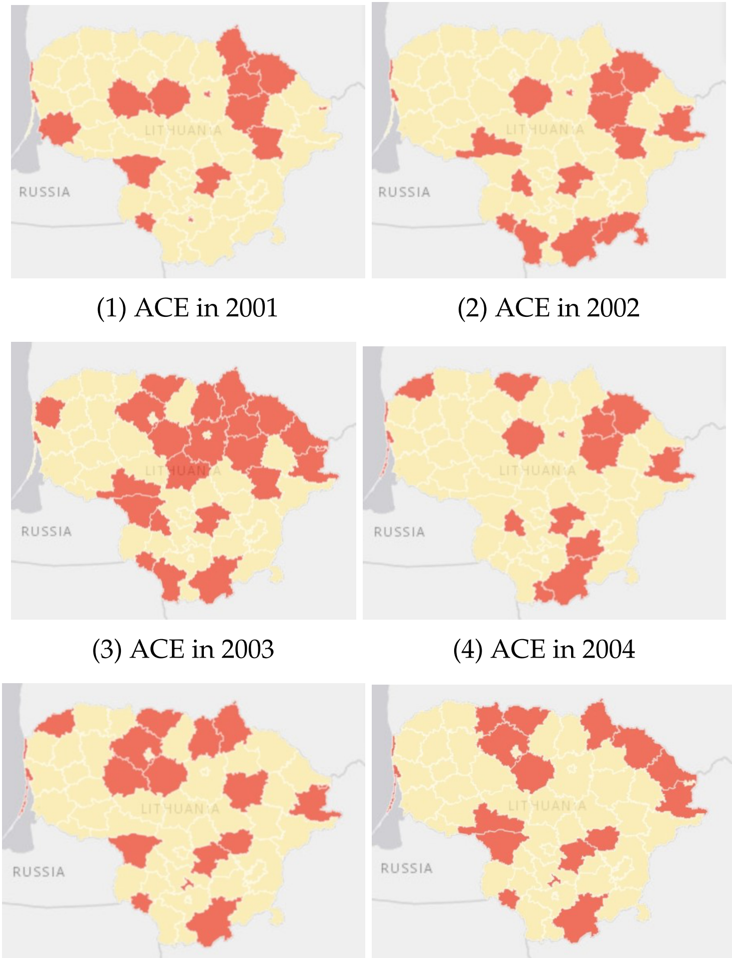
Table A1, continuation

Municipality	$card(N_i)$	**Neighbours (*nearest neighbor)**
Panevėžio r.	9	Pakruojo r., Kupiškio r., Radviliškio r., Anykščių r., Ukmergės r., Biržų r., Pasvalio r., Kėdainių r., *Panevėžio m.
Pasvalio r.	3	Pakruojo r., *Panevėžio r., Biržų r.
Plungės r.	6	Telšių r., Kretingos r., *Rietavo r., Skuodo r., Klaipėdos r., Mažeikių r.
Prienų r.	7	*Birštono, Kazlų Rūdos, Trakų r., Alytaus r., Kaišiadorių r., Marijampolės, Kauno r.
Radviliškio r.	6	Pakruojo r., Panevėžio r., *Šiaulių r., Kelmės r., Raseinių r., Kėdainių r.
Raseinių r.	6	Radviliškio r., Kelmės r., Tauragės r., Kėdainių r., *Jurbarko r., Kauno r.
Rietavo r.	4	Telšių r., Klaipėdos r., *Plungės r., Šilalės r.
Rokiškio r.	5	Kupiškio r., Anykščių r., Biržų r., Zarasų r., *Utenos r.
Skuodo r.	3	Kretingos r., Mažeikių r., *Plungės r.
Šakių r.	4	Kazlų Rūdos, *Jurbarko r., Vilkaviškio r., Kauno r.
Šalčininkų r.	3	Trakų r., Varėnos r., *Vilniaus r.
Šiaulių m.	1	*Šiaulių r.
Šiaulių r.	7	Pakruojo r., Joniškio r., Telšių r., Radviliškio r., Akmenės r., Kelmės r., *Šiaulių m.
Šilalės r.	6	Telšių r., Rietavo r., Šilutės r., Klaipėdos r.,Kelmės r., *Tauragės r.
Šilutės r.	4	Pagėgių, *Klaipėdos r., Tauragės r., Šilalės r.
Širvintų r.	6	Molėtų r., *Ukmergės r., Jonavos r., Kaišiadorių r., Elektrėnų, Vilniaus r.
Švenčionių r.	4	Molėtų r., *Ignalinos r., Utenos r., Vilniaus r.
Tauragės r.	6	Pagėgių, Šilutės r., Kelmės r., Raseinių r., *Šilalės r., Jurbarko r.
Telšių r.	7	Rietavo r., * Mažeikių r., Šiaulių r., Akmenės r., Kelmės r., Plungės r., Šilalės r.

Table A1, continuation

Municipality	$card(N_i)$	Neighbours (*nearest neighbor)
Trakų r.	8	Šalčininkų r., Varėnos r., Alytaus r., Prienų r., Kaišiadorių r., *Elektrėnų, Vilniaus m., Vilniaus r.
Ukmergės r.	6	Molėtų r., Anykščių r., *Širvintų r., Panevėžio r., Jonavos r., Kėdainių r.
Utenos r.	6	*Molėtų r., Anykščių r., Rokiškio r., Ignalinos r., Zarasų r., Švenčionių r.
Varėnos r.	4	Šalčininkų r., *Trakų r., Druskininkų, Alytaus r.
Vilkaviškio r.	4	Kazlų Rūdos, Šakių r., Marijampolės, *Kalvarijos
Vilniaus m.	2	Trakų r., *Vilniaus r.
Vilniaus r.	7	Molėtų r., Šalčininkų r., Širvintų r., Trakų r., Švenčionių r., Elektrėnų, *Vilniaus m.
Visagino m.	2	*Ignalinos r., Zarasų r.
Zarasų r.	4	Lazdijų r., Visagino m., *Ignalinos r., Utenos r.

B APPENDIX - Classified Lithuanian municipalities



(1) ACE in 2001

(2) ACE in 2002

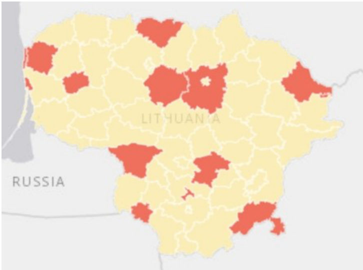
(3) ACE in 2003

(4) ACE in 2004

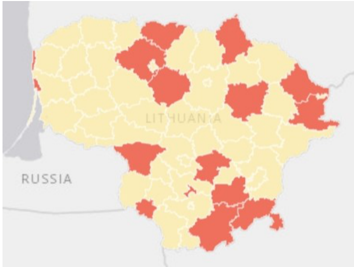
(5) ACE in 2005

(6) ACE in 2006

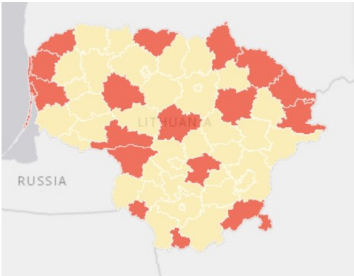
Figure A1: Classified Lithuanian municipalities. Yellow color areas indicate municipalities with low level of mortality due to ACE (with label value 0) and red areas indicates municipalities with high level of mortality due to ACE (with label value 1).



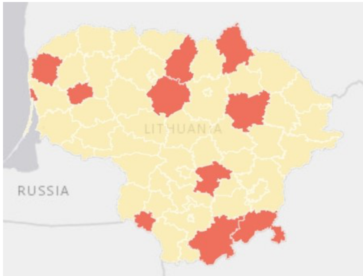
(7) ACE in 2007



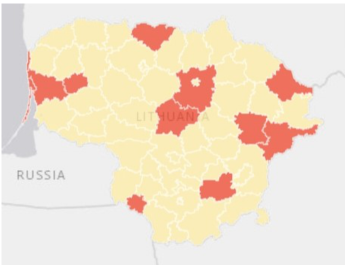
(8) ACE in 2008



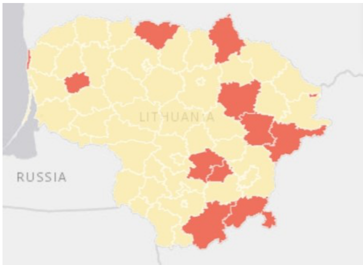
(9) ACE in 2009



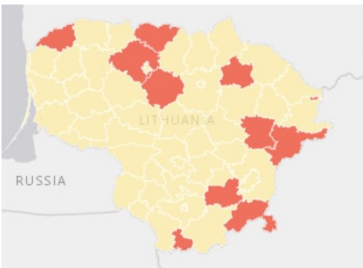
(10) ACE in 2010



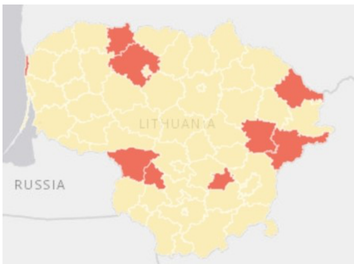
(11) ACE in 2011



(12) ACE in 2012

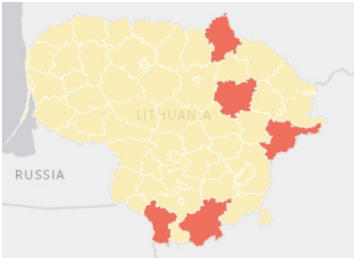


(13) ACE in 2013

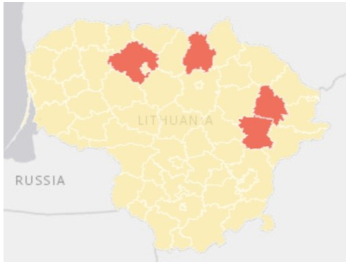


(14) ACE in 2014

Figure A1: Classified Lithuanian municipalities. Yellow color areas indicate municipalities with low level of mortality due to ACE (with label value 0) and red areas indicates municipalities with high level of mortality due to ACE (with label value 1). (continued from previous page)



(15) ACE in 2015



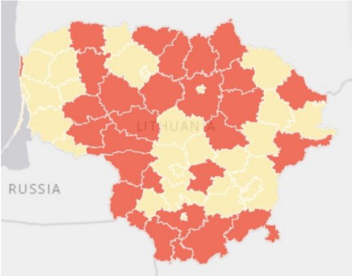
(16) ACE in 2016

(17) ACE in 2017

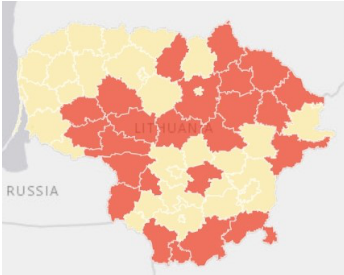
(18) ACE in 2018

(19) ACE in 2019

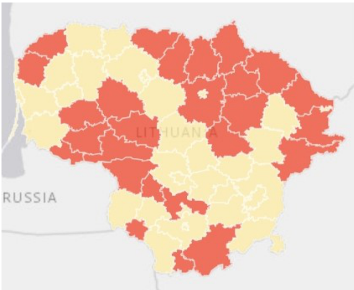
Figure A1: Classified Lithuanian municipalities. Yellow color areas indicate municipalities with low level of mortality due to ACE (with label value 0) and red areas indicates municipalities with high level of mortality due to ACE (with label value 1). (continued from previous page)



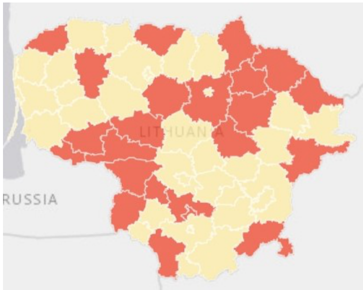
(1) CSD in 2001



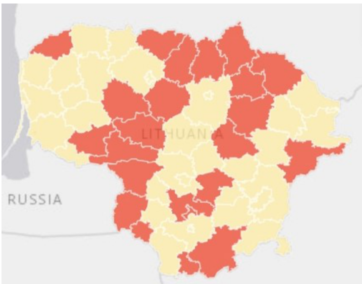
(2) CSD in 2002



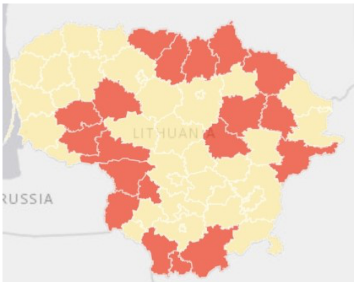
(3) CSD in 2003



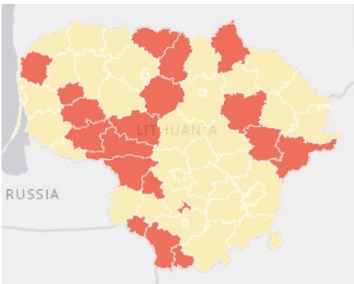
(4) CSD in 2004



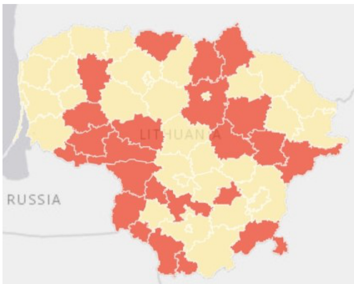
(5) CSD in 2005



(6) CSD in 2006

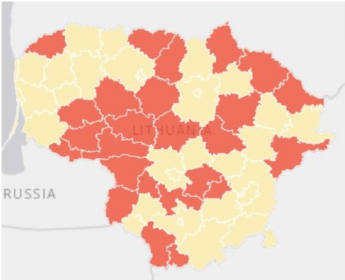


(7) CSD in 2007

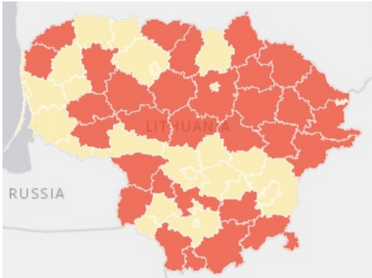


(8) CSD in 2008

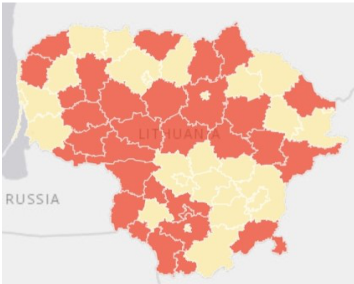
Figure A2: Classified Lithuanian municipalities. Yellow color areas indicate municipalities with low level of mortality due to CSD (with label value 0) and red areas indicates municipalities with high level of mortality due to CSD (with label value 1).



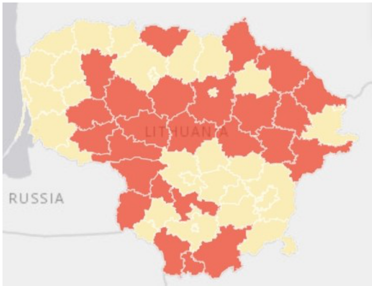
(9) CSD in 2009



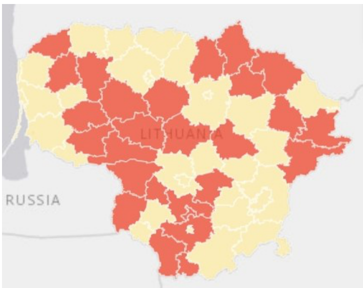
(10) CSD in 2010



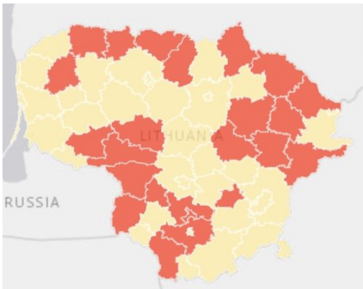
(11) CSD in 2011



(12) CSD in 2012

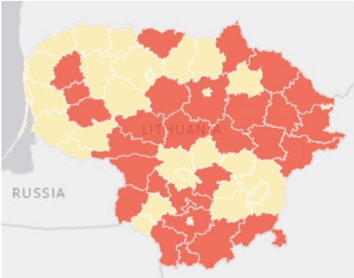


(13) CSD in 2013

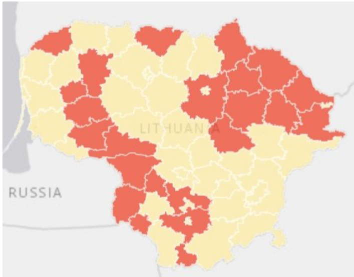


(14) CSD in 2014

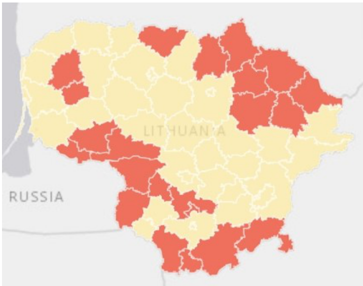
Figure A2: Classified Lithuanian municipalities. Yellow color areas indicate municipalities with low level of mortality due to CSD (with label value 0) and red areas indicates municipalities with high level of mortality due to CSD (with label value 1). (continued from previous page)



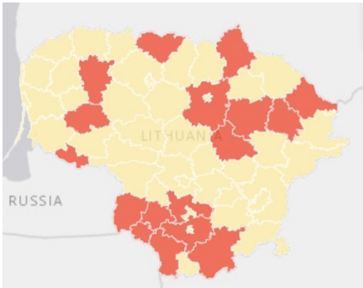
(15) CSD in 2015



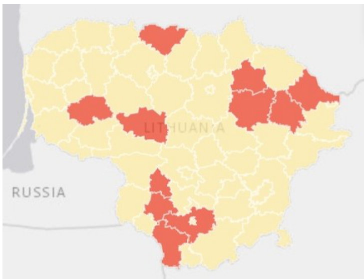
(16) CSD in 2016



(17) CSD in 2017



(18) CSD in 2018



(19) CSD in 2019

Figure A2: Classified Lithuanian municipalities. Yellow color areas indicate municipalities with low level of mortality due to CSD (with label value 0) and red areas indicates municipalities with high level of mortality due to CSD (with label value 1). (continued from previous page)

SANTRAUKA (SUMMARY IN LITHUANIAN)

Mašininio mokymosi (*angl. machine learning*) algoritmai, pagrįsti klasifikuota mokymo imtimi, priskiriami prižiūrimo (angl. *supervised*) klasifikavimo kategorijai. Du pagrindiniai prižiūrimo klasifikavimo modeliai yra generatyviniai (*angl. generative*) ir diskriminatyviniai (*angl. discriminative*) [4].

Šiame darbe pagrindinis dėmesys skiriamas prižiūrimiems generatyviniams modeliams. Matematiškai generatyviniai modeliai skirti požymio vektoriaus Z ir klasės žymės vektoriaus Y jungtiniam tikimybiniam skirstiniui (*angl. joint probability distribution*) $P(Z, Y)$. Įvairios $P(Z|Y)$ ir $P(Y)$ išraiškos leidžia išplėsti prižiūrimų generatyvinių modelių taikymą įvairioms klasifikavimo problemoms spręsti.

Iš Bajeso teoremos seka $P(Y|Z) = P(Z, Y)/P(Z) = P(Z|Y)P(Y)/P(Z)$ arba $P(Y|Z) \propto P(Z|Y)P(Y)$, kadangi $P(Z)$ nepriklauso nuo klasės žymės Y.

Taigi Bajeso klasifikavimo taisyklė (*angl. Bayes classification rule (BCR)*) pilnai apibrėžiama požymio sąlyginiu skirstiniu $P(Z|Y)$ ir žymės besąlyginiu skirstiniu $P(Y)$.

Disertacijoje koncentruojamasi į $P(Z|Y)$ ir $P(Y)$ skirstinius, aprašančius erdvės-laiko (*angl. Spatio-temporal, ST*) duomenis.

Problema ir aktualumas

Yra žinoma, kad net paprasti Gauso modeliai su bendromis parametrinėmis vidurkio ir kovariacijų funkcijomis susiduria su didelėmis parametrų įvertinimų problemomis, kas sąlygoja dideles skaičiavimo laiko sąnaudas.

Darbe sprendžiami parametrų vertinimo uždaviniai trims erdvės-laiko Gauso modelių tipams. Šie modelių tipai skiriasi:

- statistinės (laiko ir erdvės atžvilgiu) priklausomybės formomis požymio ir klasės žymės kintamiesiems;
- modelio parametrų vertinimo taisyklėmis ir strategijomis;
- erdvės ir laiko konteksto įtraukimo scenarijais;
- klasifikatorių tikslumo vertinimo matais.

Norint išspręsti prižiūrimo Bajesinio klasifikavimo uždavinius stebiniams, kurie aprašomi Gauso erdvės-laiko duomenų modeliais panaudojant erdvės-laiko kontekstinę informaciją, yra reikalingi nauji klasifikavimo metodai, kurie praplėstų generatyvinio modelio taikymo galimybes.

Tyrimo objektas

Generatyviniai mašininio mokymosi modeliai: Gauso gardelės ir geostatistiniai duomenys; Bajeso diskriminantinės funkcijos (*angl. Bayes discriminat funkction, BDF*) ir įterptosios Bajeso diskriminantinės funkcijos (*angl. plug-in BDF (PBDF)*); erdvinis Gauso paslėptas Markovo modelis (*angl. Gaussian Hidden Markov Model, GHMM*); erdvinis Gauso Markovo atsitiktinis laukas (*angl. Gaussian Markov Random Field, GMRF*); Gauso atsitiktinis laukas (*angl. Gaussian Random field, GRF*); Gauso geostatistinis modelis (*angl. Gaussian Geostatistical Model, GGM*); laiko eilučių Autoregresinis modelis (*angl. Autoregressive model, AR*); Bernulio skirstinys klasės žymei.

Tyrimo tikslas ir uždaviniai

Tyrimo tikslas – plėtoti prižiūrimo Bajesinio klasifikavimo metodus įvairiems Gauso erdvės-laiko modeliams, panaudojant generatyvinių mašininio mokymosi modelius. Pasiūlytos originalios parametrų vertinimo strategijos ir taisyklės. Išvestos daugelio parametrų įvertinių analitinės išraiškos įgalina efektyviau panaudoti kompiuterius realiuose skaičiavimuose.

Tuo pačiu buvo siekiama apibendrinti analogiškų klasifikavimo procedūrų tyrimus, vykdytus tiktai „grynai“ erdviniam kontekste.

Buvo pasirinkti trijų tipų erdvės-laiko Gauso duomenų modeliai su skirtingomis požymių ir klasės žymių statistinės priklausomybės formomis.

Siekiant numatyto tikslo buvo suformuluoti šie uždaviniai:

1. Pasiūlyti erdvės Gauso paslėpto Markovo modelio klasifikavimo algoritmą gardelės duomenims, naudojant originalią parametrų vertinimo strategiją, išvedant sintetinius (*angl. synthetic*) įverčius, pagrįstais erdviniu svoriu. Konkuruojančių klasifikatorių palyginimas atliekamas pagal parametrų įvertinimo metodus ir tikslumo (*angl. performance*) rodiklius.
2. Plėtoti Gauso Markovo atsitiktinio lauko-Autoregresinio modelio prižiūrimą Bajesinį klasifikavimą su separabiliomis kovariacijomis ir nepriklausomais Bernulio klasių žymių kintamaisiais, kurių parametrai priklauso nuo įvairių erdvės asociacijos indeksų (*angl. spatial association index*).
3. Plėtoti Gauso Geostatistinio-Autoregresinio modelio prižiūrimą Bajesinį klasifikavimą su separabiliomis kovariacijomis ir nepriklausomais Bernulio klasių žymių kintamaisiais, kurių parametrai priklauso nuo Euklido atstumų ir laiko vėlavimų (*angl. lags*).
4. Įvertinti ir palyginti siūlomų klasifikatorių tikslumą kiekvieno duomenų tipo modeliuose, aprašytuose pagal skirtingą erdvės ir laiko informaciją.

Tyrimo metodika

Literatūros apžvalgoje įvertinti metodai, naudojami modeliuojant požymių reikšmes su statistine erdvės-laiko kontekstine informacija, taip pat įvertinti generatyviniai modeliai, naudojami sprendžiant prižiūrimo klasifikavimo uždavinius modeliams su erdvine kontekstine informacija. Apibrėžtas erdvės-laiko kontekstinės informacijos įtraukimas į požymių ir klasių žymių paskirstymą yra naudojamas vieno žingsnio (*angl. one-step-ahead, OSA*) klasifikavimui kiekviename taške arba srityje.

Siūlomas klasifikavimo metodas yra pagrįstas klasifikuojamo stebinio (*angl. Observation to be classified, OBC*) sąlyginio skirstinio logaritmų santykiu kiekviename taške, atsižvelgiant į mokymo imtį. Klasifikatoriai vertinami keliais tikslumo matais, naudojant klasifikavimo matricas ir sąlyginius klasifikuojamo stebinio skirstinius.

Empiriniame tyrime siūlomi klasifikavimo algoritmai realizuojami generuotiems ir realiems duomenimis. Duomenų modelio parametrai įvertinti naudojant maksimalaus tikėtinumo (*angl. Maximum Likelihood, ML*) įverčius iš mokymo imties. Kiekvieno tipo duomenų modeliuose atliekamas detalus siūlomų klasifikatorių palyginimas įvairiems požymių ir klasių žymių skirstiniams.

Mokslinis naujumas ir praktinė reikšmė

Šis darbas skirtas generatyvinių modelių taikymui prižiūrimame erdvės-laiko duomenų Bajesinio klasifikavimo metodų kūrimui ir algoritmų sudarymui. Mokslinis naujumas atskleidžiamas erdvės-laiko modelių specifikacijoje, parametrų vertinimo strategijose ir klasifikatorių tikslumo įvertinimo metoduose. Parinktiems erdvės-laiko duomenų modeliams būdingas nedidelis parametrų skaičius, kurį dažnai galima įvertinti analitinių formulių pagalba, naudojant maksimalaus tikėtinumo metodą. Tai užtikrina mažas skaičiavimo laiko sąnaudas, konkurencingumą tarp anksčiau tirtų erdvės-laiko duomenų klasifikatorių ir dideles taikymo galimybes.

Siūlomas vieno žingsnio klasifikavimo metodas yra realizuotas trijų parametrinių duomenų modelių tipų generuotų ir realių duomenų pavyzdžiais. Klasifikatoriai, pagrįsti skirtingais į Bajeso diskriminantines funkcijas įtrauktos erdvės-laiko informacijos lygiais, lyginami įvairiais tikslumo matais.

Mokslinį naujumą pagrindžia šie esminiai rezultatai:

1. Pasiūlytos parametrų vertinimo strategijos ir taisyklės leidžia sukurti analitines įverčių išraiškas, kurios palengvina skaičiavimo sunkumus, susijusius su klasifikavimo problemų sprendimu.
2. Eksperimentiniai rezultatai parodė, kad klasifikatoriai, pagrįsti Gauso paslėpto Markovo modelio su aukštesniu erdvinio svorio įtraukimu (*angl. spatial weighting level*), daugeliu atvejų turi

didesnį tikslumą lyginant su mažesniu erdvinio svorio įtraukimu. Taigi siūloma metodika gali būti laikoma vertingu esamo erdvinio paslėpto Markovo modelio išplėtimu atvejams, kai naudojami tolygūs požymių ir klasės žymių skirstiniai požymiams ir pastovios perėjimo tikimybės (*angl. transition probabilities*).

3. Bajeso klasifikatorių įvedimas erdvės-laiko gardelės duomenims, modeliuojamiems Gauso Markovo atsitiktinio lauko su separabiliomis kovariacijomis ir nepriklausomais Bernulio klasių žymių kintamaisiais, kurių parametrai priklauso nuo erdvės asociacijos indeksų.

4. Pasiūlyti Bajeso klasifikavimo uždavinio sprendimo algoritmai erdvės-laiko geostatistiniams duomenims, modeliuojamiems Gauso Markovo atsitiktinio lauko su separabiliomis kovariacijomis ir nepriklausomais Bernulio kintamaisiais klasių žymėms, kurių parametrai priklauso nuo klasifikuojamo stebinio erdvės-laiko deterministinio konteksto.

5. Išvestos klasifikavimo klaidų ir jų D-įvertinių analitinės išraiškos, naudojant diskriminantines funkcijas dviem erdvės-laiko duomenų modelių tipams su separabiliomis kovariacijos funkcijomis.

Ginamieji teiginiai

Darbe ginami šie teiginiai:

1. Klasifikatoriams, pagrįstiems Gauso paslėptu Markovo modeliu:
   (a) perėjimo tikimybės įvertinimai, pagrįsti žymės išlaikymo (*angl. persistence*) rodikliu, leidžia sukurti klasifikatorių, kurio tikslumas yra didesnis nei kito metodo, pagrįsto perėjimo dažniais;
   (b) pilnas erdvinio svorio įtraukimas taikomas erdvinio Gauso paslėpto Markovo modelio parametrų įvertinimui, užtikrina didesnį vieno žingsnio laike klasifikavimo tikslumą nei dalinis erdvinio svorio įtraukimas.

2. Skirtingais Moran's I, Geary's C and Getis-Ord G indeksais pagrįstų klasifikatorių tikslumas Gauso Markovo atsitiktinio lauko ir Autoregresiniam modeliui beveik visais atvejais mažai skiriasi.

3. Klasifikatorių, naudojant Gauso Geostatistinį modelį, tikslumas priklauso nuo naudojamos erdvinės kovariacijų funkcijos formos ir informacijos, įtrauktos į klasės žymių paskirstymą.
4. Išvestos parametrų įvertinimo analitinės išraiškos ssumažina naudojamų skaitmeninių procedūrų sudėtingumą.

Disertacijos struktūra ir apimtis

Disertaciją sudaro įvadas, 3 skyriai, išvados, literatūros sąrašas, priedai ir santrauka lietuvių kalba. Pirmame skyriuje aprašomi nagrinėjami erdvės-laiko modeliai ir jiems taikomas generatyvinis klasifikavimo metodas. Antrame skyriuje aptariami erdvinių duomenų modeliavimo klausimai, erdvinio modelio įverčiai. Pateikiami pagrindiniai baigiamojo darbo, susijusio su erdvės-laiko duomenų diskriminantine analize, rezultatai. Trečiame skyriuje pristatomi skaitiniai eksperimentai su generuotais ir realiais duomenimis. Bendrosios išvados pristatomos po trečiojo skyriaus. Disertacijoje įtraukti 112 literatūros šaltiniai, jie yra pateikti darbo pabaigoje. Disertaciją sudaro 144 puslapiai, 17 paveikslų ir 24 lentelių. Disertacija parašyta anglų kalba.

S.1 GENERATYVUS PRIŽIŪRIMAS ERDVĖS-LAIKO DUOMENŲ KLASIFIKAVIMAS

Gamtos ir kituose tyrimuose duomenys dažnai renkami erdvėje ir nuosekliai laike. Daugelis ekologijos ir aplinkos mokslų problemų, tokių kaip populiacijos rūšies buvimo/nebuvimo stebėjimas, apima erdvinių dvejetainių atsitiktinių laukų stebėjimą klasių žymėms vaizduoti. Vis dažniau į šiuos tyrimus įtraukiama ir laiko komponentė. Markovo atsitiktinio lauko (*angl. Markov Random Field, MRF*) modelius galima modifikuoti, kad būtų įtraukta laiko priklausomybė, nesvarbu, ar priklausomybė yra lokaliame ar globaliame lygmenyje. Tačiau dirbant su Markovo atsitiktinio lauko modeliais svarbu užtikrinti, kad erdvinė priklausomybė būtų tinkamai išanalizuota. Erdvės-laiko duomenys paprastai registruojami reguliariais laiko intervalais ir nereguliariose stotyse (srityse) kompaktiškame plote (žr., [19, 39]).

Dažnai prieš analizuojant erdvės-laiko duomenų rinkinius taikomas erdvės-laiko diskretizavimas (arba agregavimas). Diskretizavimas yra naudingas norint apibendrinti informaciją ir išgauti erdvės-laiko diapazono savybes, o ne matuoti vieną tašką [36].

Erdvės-laiko duomenų klasifikacija, kuri yra svarbi duomenų tyrybos dalis, tampa vis svarbesnė didžiųjų duomenų eroje, kai didėja didelių erdvės-laiko duomenų rinkinių, pvz., žemėlapių, virtualių gaublių, nuotolinio stebėjimo vaizdų, GPS trajektorijų, prieinamumas ir svarba.

Darbe aptariami trijų erdvės-laiko duomenų modelių tipai, pagrįsti laiko eilutėmis, kurios atitinka skirtingas erdvines vietas (erdvės laiko eilutės). Erdvinis Gauso paslėptas Markovo modelis tinka taikymams, kuriuose yra sąlygiškai nepriklausomų Gauso požymio stebėjimams kiekviename taške (*angl. sites*) ir pirmosios eilės Markovo grandinės (*angl. First-order Markov Chain, MC1*) žymei. Šį metodą pateikė Dučinskas ir kt. [28].

Erdvės-laiko duomenų modeliai, kurie aprašomi Gauso Markovo atsitiktinio lauko ir geostatistiniu Gauso Markovo atsitiktinio lauko modeliu su separabilia kovariacija, sudaryti, kai objektų stebėjimams tam tikroje vietoje taikomas laiko eilučių autoregresinis modelis.

Išsamią Gauso erdvinių duomenų statistinio klasifikavimo ir diskriminacijos metodų apžvalgą pateikia daugelis autorių (žiūr. [5, 23, 46, 50, 60, 61, 70, 78, 90, 112]). Naują požiūrį į Gauso Markovo atsitiktinio lauko stebėjimo Bajesinį klasifikaciją pateikė Dučinskas ir Dreižienė [25, 26] ir generatyvinį klasifikavimo metodą geostatistiniam erdvės-laiko Gauso modeliui pristatė Karaliutė and Dučinskas [43, 44]. Apskritai erdvės-laiko duomenis sudaro du komponentai, sugeneruoti kiekybinis požymis atsitiktiniam laukui (*angl. Random Field, RF*) Z ir kokybinės klasės žymė atsitiktiniam laukui Y.

Darbe nagrinėjami tik stacionarių erdvės-laiko duomenų modeliai. Norint atsižvelgti į erdvinį kintamumą, modeliuojama funkcija, naudojant erdvinių kovariacijų tiesinę regresiją kiekvienu laiko momentu, apibrėžiant erdvinius svorius ir nurodant erdvinės kovariacijos parametrus.

Darbe nagrinėjami erdvės-laiko duomenys dviem atsitiktiniams laukams: Gauso atsitiktinis laukas $\{Z(s;t) : s \in S \subset R^2, t \in D_T\}$ ir atsitiktinis laukas $\{Y(s;t) : s \in S \subset R^2, t \in D_T\}$, vaizduojantis klasės žymę su reikšmėmis 0 arba 1, t.y. pasiskirstęs pagal Bernulio skirstinį $Ber(p)$ su parametru p, kuris gali priklausti nuo erdvės-laiko konteksto.

Todėl sutelkiamas dėmesys tik į dviejų klasių atvejį.

Tariame, kad $l = 0, 1$ stebėjimo $Z(s;t)$ modelis su sąlyga $Y(s;t) = l$ yra

$$Z(s;t) = \mu_l(s;t) + \varepsilon(s;t),$$

kur $\mu_l(s;t)$ – determinuotas erdvės-laiko vidurkis ir $\varepsilon(s;t) \sim N(0, \sigma^2(s,t))$ visiems erdvės taškams $s \in S$ ir diskretaus laiko indeksui $t \in D_T = 1, 2, \ldots$.

Erdvinio indekso galimų reikšmių rinkiniui $S_n = \{s_i \in S; i = 1, \ldots, n\}$ yra apibrėžiama kaimynystės sistema $N = \{N_i : i = 1, \ldots, n\}$, kur N_i žymi vietų rinkinį, kurios priklauso nuo vietos s_i. Tegu $card(N_i)$ žymi elementų skaičių aibėje N_i.

Tiek gardelės (sričių duomenims), tiek geostatistiniams duomenims kaimynystė N_i apibrėžiama kaip tos vietos, su kuriomis s_i turi „artimas" savybes, tokias kaip ribos, sienos ar įvairūs artumo rodikliai.

Apibrėžiama kvadratinė simetrinė erdvinių svorių matrica $W = (w_{ij} : i, j = 1, \ldots, n)$, kur w_{ij} parodo tiesioginio erdvinio ryšio tarp s_i ir s_j buvimą ir įgauna teigiamą reikšmę, jei $s_j \in N_i$, o 0 kitu atveju.

Pažymime $Y(s_i, t) = Y_t^{(i)}$ ir $Z(s_i, t) = Z_t^{(i)}$, $Z_t = \left(Z_t^{(1)}, \ldots, Z_t^{(n)}\right)'$, $Y_t = \left(Y_t^{(1)}, \ldots, Y_t^{(n)}\right)'$, $Z^{(i)} = \left(Z_1^{(i)}, \ldots, Z_T^{(i)}\right)'$, $Y^{(i)} = \left(Y_1^{(i)}, \ldots, Y_T^{(i)}\right)'$.

Pažymime $z_t = \left(z_t^{(1)}, \ldots, z_t^{(n)}\right)$ ir $y_t = \left(y_t^{(1)}, \ldots, y_t^{(n)}\right)$, kurie realizuoja Z_t ir Y_t.

Pagrindinis uždavinys yra klasifikuoti Z_{T+1} su mokymo imtimi (Z, Y), kuris dažnai vadinamas vienu žingsnio laike uždaviniu.

Pažymime bendrą $\{Z_{T+1}, Y_{T+1}, Z, Y\}$ skirstinį $P(Z_{T+1}, Y_{T+1}, Z, Y; \Psi)$, kur Ψ yra visų erdvės-laiko modelio parametrų rinkinys. Dėl paprastumo kartais praleisime priklausomybės nuo Ψ, kur nebūtina.

Pagal bendro skirstinio skaidymo taisyklę turime

$$P\big(Z_{T+1}, Y_{T+1}, Z, Y\big) = P\big(Z_{T+1} \big| Y_{T+1}, Z, Y\big) P\big(Y_{T+1}, Y\big) P\big(Z|Y\big)$$

ir i-ajai vietai

$$P\Big(Z_{T+1}^{(i)}, Y_{T+1}^{(i)}, Z, Y\Big) = P\Big(Z_{T+1}^{(i)} \Big| Y_{T+1}^{(i)}, Z, Y\Big) P\Big(Y_{T+1}^{(i)}, Y\Big) P\Big(Z|Y\Big).$$

Tai pagal Bajeso taisyklę optimalus $Y_{T+1}^{(i)}$ įvertis yra gaunamas

$$\widehat{Y}_{T+1}^{(i)} = \arg\max_{l=0,1}\Big(L\Big(Z_{T+1}^{(i)}, Y_{T+1}^{(i)} = l, Z, Y\Big)\Big),$$

kur $L\Big(Z_{T+1}^{(i)}, Y_{T+1}^{(i)} = l, Z, Y\Big)$ yra skirstinio $P\Big(Z_{T+1}^{(i)}, Y_{T+1}^{(i)}, Z, Y\Big)$ tikėtinumo funkcija.

Iš lygties

$$\frac{L\Big(z_{T+1}^{(i)}, y_{T+1}^{(i)} = 1, z, y, \Psi\Big)}{L\Big(z_{T+1}^{(i)}, y_{T+1}^{(i)} = 0, z, y, \Psi\Big)} = \frac{p\Big(z_{T+1}^{(i)}\Big|y_{T+1}^{(i)} = 1, z, y\Big)Pr\Big(y_{T+1}^{(i)} = 1, y\Big)}{p\Big(z_{T+1}^{(i)}\Big|y_{T+1}^{(i)} = 0, z, y\Big)Pr\Big(y_{T+1}^{(i)} = 0, y\Big)}$$

log tikimybių santykis yra

$$W_Z\Big(z_{T+1}^{(i)}\Big) = ln\frac{p\Big(z_{T+1}^{(i)}\Big|z, y, y_{T+1}^{(i)} = 1\Big)}{p\Big(z_{T+1}^{(i)}\Big|z, y, y_{T+1}^{(i)} = 0\Big)},$$

kur $\gamma = ln\frac{Pr\left(y_{T+1}^{(i)}=1,y\right)}{Pr\left(y_{T+1}^{(i)}=0,y\right)}$.

Todėl darbe sutelkiamas dėmesys į įvairius skirstinių modelius $P\Big(z_{T+1}^{(i)}\Big|z, y, y_{T+1}^{(i)} = l\Big)$ su sąlyginiais tankiais $p\Big(z_{T+1}^{(i)}\Big|z, y, y_{T+1}^{(i)} = l\Big)$ ir $P\Big(Y_{T+1}^{(i)} = l, Y\Big)$ su bendrom klasių žymių tikimybėm $P\Big(Y_{T+1}^{(i)} = l, y\Big)$.

Bajeso diskriminantinė funkcija klasifikuoja stebėjimą $Z_{T+1}^{(i)} = z_{T+1}^{(i)}$: klasė $\widehat{Y}_{T+1}^{(i)} = 1$, jeigu $W_Z\Big(z_{T+1}^{(i)}\Big) \geq 0$, o $\widehat{Y}_{T+1}^{(i)} = 0$ kitu atveju, t.y. $\widehat{Y}_{T+1}^{(i)} = H\Big(W_Z\Big(Z_{T+1}^{(i)}\Big)\Big)$. Tačiau praktikoje visi erdvės-laiko modelio parametrai retai žinomi.

Tada nežinomų parametrų įverčiai, gauti iš mokymo rinkinio maksimalaus tikėtinumo metodu, yra įterpiami į Bajeso diskriminantinę funkciją, kuri vadinama įterptąja Bajeso diskriminantine funkcija $\widehat{W}_Z\Big(Z_{T+1}^{(i)}\Big)$.

Klasifikavimo matrica (*angl. confusion matrix*), kuri bus taikoma vertinant siūlomus klasifikatorius, pateikta S.1 lentelėje.

S.1 lentelė: Klasifikavimo matrica i sričiai.

$\boldsymbol{Y_{T+1}^{(i)}}$	$\boldsymbol{\widehat{Y}_{T+1}^{(i)}}$ **0**	**1**
0	$m_{00}^{(i)}$	$m_{01}^{(i)}$
1	$m_{10}^{(i)}$	$m_{11}^{(i)}$

kur $m_{kl}^{(i)} = I\Big(Y_{T+1}^{(i)} = k\Big) I\Big(\widehat{Y}_{T+1}^{(i)} = l\Big)$, $k, l = 0, 1$, $n_{kl} = \sum\limits_{i=1}^{n} m_{kl}^{(i)}$, $k, l = 0, 1$.

Tradiciškai, dažniausiai naudojamas empirinis klasifikatoriaus vertinimo matas vadinamas tikslumo rodikliu (*angl. Accuracy rate, ACC*), kuris parodo teisingai klasifikuotų testo duomenų dalį, apskaičiuojamas pagal formulę

$$ACC = \frac{n_{00} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}}. \tag{S.1}$$

Reikia pažymėti, kad šalia bendro tikslumo yra "dalinio" tikslumo matai: jautrumas (*teisingai teigiamas rodiklis*) (*angl. sensitivity (true positive rate, TPR)*)

$$TPR = \frac{n_{00}}{n_{00} + n_{01}}$$

ir specifiškumas (*teisingai neigiamas rodiklis*) (*angl. specificity (true negative rate, TNR)*)

$$TNR = \frac{n_{11}}{n_{10} + n_{11}}$$

dažnai naudojami mašininiame mokyme.

Tačiau praktikoje dažnai pasitaiko situacijų, kai daugumos ir mažumos klasių aibės smarkiai nesutampa (žiūr., [52, 53, 65, 66]). Tada klasifikatorių įvertinimas turi būti atliktas naudojant tam tikrus rodiklius, kad būtų atsižvelgta į klasių skirstinius. Šiame darbe taip pat naudojamos kitos vertinimo priemonės, pagrįstomis klasifikavimo matrica, vadinama subalansuotu tikslumu (*angl. Balanced accuracy rate, BAC*) (žiūr., [52, 53, 80]) ir apskaičiuojama formule

$$BAC = (TPR + TNR)/2. \tag{S.2}$$

Geometrinis tikslumas (*angl. G-mean*) yra svarbus norint išmatuoti, kaip išvengti perteklinės mažumos klasės, (žiūr., [85]) ir apskaičiuojama formule

$$GAC = \sqrt{TPR \cdot TNR}. \tag{S.3}$$

Taikoma vizualizavimo (*angl. Receiver Operator Characteristic, ROC*) kreivė [77], kuri leidžia vizualizuoti kompromisą (*angl. trade-off*) tarp TPR ir $FPR = 1 - TNR$ (*klaidingai teigiamas rodiklis*) (*angl. false positive rate, FPR*) bet kokiai klasifikavimo matricai, atitinkančiai pasirinktą įterptąją Bajeso diskriminantinę funkciją.

Yra keletas tikslumo rodiklių, kurie yra pagrįsti klaidingu klasifikavimu.

Paprastai yra skaičiuojama neteisingai klasifikuotų stebėjimų dalis, kuri gaunama $PIC = 1 - ACC$.

Norint naudoti tikimybėmis pagrįstus tikslumo rodiklius, būtina nurodyti sąlyginį klasifikuojamo stebinio skirstinį.

Mokymo imties žymių santykis apibrėžiamas kaip $lr^{(i)} = \sum\limits_{t=1}^{T} I\big(Y_t^{(i)} = 1\big)/T$ visiems $i = 1, \dots, n$.

$t = 1, 2, \dots, T$ pažymime n_l^t sričių vienetų skaičių su žyme l ir $IR_t = \frac{n_1^t}{n_0^t}$ disbalanso santykiu, kur $n_l^t = \sum\limits_{i=1}^{n} I\Big(Y_t^{(i)} = l\Big)$.

Išsamesnis aprašymas ir skyriaus išvados pateikiami disertacijos tekste.

S.2 KLASIFIKAVIMO ALGORITMAI ERDVĖS-LAIKO DUOMENŲ MODELIAMS

S.2.1 Erdvės Gauso paslėpto Markovo modelis

Gauso paslėpto Markovo modelio parametrai i-jam gardelės taške $\Psi^{(i)} = \Big(\pi^{(i)}, A^{(i)}, B^{(i)}\Big)$, $i = 1, \dots, n$. Parametrų vertinimams pagal mokymo imtį naudojama erdvinio svorio (*angl. Spatial weighting, SW*) strategija.

Maksimalaus tikėtinumo įverčiai Gauso parametrams

$\widehat{\beta}^{(i)} = \Big(\Big(X^{(i)}\Big)' X^{(i)}\Big)^{-1} \Big(X^{(i)}\Big)' Z^{(i)}$;

$\widehat{\sigma}_i^2 = \Big(Z^{(i)} - X^{(i)}\widehat{\beta}^{(i)}\Big)' \Big(Z^{(i)} - X^{(i)}\widehat{\beta}^{(i)}\Big)/(T - 2q)$.

Perėjimo tikimybės įvertinamos pagal dvi taisykles:

(M1) $\widehat{a}^{(i)}_{y^{(i)}_T,1} = \widehat{a}^{(i)}_{01} I\Big(y^{(i)}_T = 0\Big) + \widehat{a}^{(i)}_{11} I\Big(y^{(i)}_T = 1\Big)$

ir $\widehat{a}^{(i)}_{y^{(i)}_T,0} = \widehat{a}^{(i)}_{00} I\Big(y^{(i)}_T = 0\Big) + \widehat{a}^{(i)}_{10} I\Big(y^{(i)}_T = 1\Big)$,

kur $\widehat{a}^{(i)}_{00} = \frac{\sum_{t=2}^{T} I\left(y^{(i)}_{t-1}=0\right) I\left(y^{(i)}_{t}=0\right)}{\sum_{t=2}^{T} I\left(y^{(i)}_{t-1}=0\right)}, \widehat{a}^{(i)}_{01} = \frac{\sum_{t=2}^{T} I\left(y^{(i)}_{t-1}=0\right) I\left(y^{(i)}_{t}=1\right)}{\sum_{t=2}^{T} I\left(y^{(i)}_{t-1}=0\right)}$,

$\widehat{a}^{(i)}_{10} = \frac{\sum_{t=2}^{T} I\left(y^{(i)}_{t-1}=1\right) I\left(y^{(i)}_{t}=0\right)}{\sum_{t=2}^{T} I\left(y^{(i)}_{t-1}=1\right)}, \widehat{a}^{(i)}_{11} = \frac{\sum_{t=2}^{T} I\left(y^{(i)}_{t-1}=1\right) I\left(y^{(i)}_{t}=1\right)}{\sum_{t=2}^{T} I\left(y^{(i)}_{t-1}=1\right)}$;

(M2) $\widehat{a}^{(i)}_{y^{(i)}_T,1} = q^{(i)} I\Big(y^{(i)}_T = 0\Big) + p^{(i)} I\Big(y^{(i)}_T = 1\Big)$

ir $\widehat{a}^{(i)}_{y^{(i)}_T,0} = p^{(i)} I\Big(y^{(i)}_T = 0\Big) + q^{(i)} I\Big(y^{(i)}_T = 1\Big)$,

kur $p^{(i)} = \frac{\sum_{t=2}^{T} I\left(y^{(i)}_{t-1}=y^{(i)}_{t}\right)}{T-1}$ ir $q^{(i)} = 1 - p^{(i)} = \frac{\sum_{t=2}^{T} I\left(y^{(i)}_{t-1}\neq y^{(i)}_{t}\right)}{T-1}$.

Parametrų įvertinimai naudojant erdvinius svorius turi pavidalą

$$\widetilde{\beta}^{(i)} = \Big(\sum_{j=1}^{n} w^*_{ij}\widehat{\beta}^{(j)} + \widehat{\beta}^{(i)}\Big)/2, \widetilde{\sigma}^2_i = \Big(\sum_{j=1}^{n} w^*_{ij}\widehat{\sigma}^2_j + \widehat{\sigma}^2_i\Big)/2,$$

$$\widetilde{a}^{(i)}_{y^{(i)}_T,1} = \Big(\sum_{j=1}^{n} w^*_{ij}\widehat{a}^{(j)}_{y^{(j)}_T,1} + \widehat{a}^{(i)}_{y^{(i)}_T,1}\Big)/2, \widetilde{a}^{(i)}_{y^{(i)}_T,0} = \Big(\sum_{j=1}^{n} w^*_{ij}\widehat{a}^{(j)}_{y^{(j)}_T,0} + \widehat{a}^{(i)}_{y^{(i)}_T,0}\Big)/2.$$

Erdvinio svorio vertinimo strategija apibrėžiama 4 parametrų įverčių aibėmis $\widehat{\Psi}^{(i)}_1 = \Big(\pi^{(i)}, \widehat{A}^{(i)}_1, \widetilde{\beta}^{(i)}, \widetilde{\sigma}^2_i\Big)$, $\widehat{\Psi}^{(i)}_2 = \Big(\pi^{(i)}, \widehat{A}^{(i)}_2, \widetilde{\beta}^{(i)}, \widetilde{\sigma}^2_i\Big)$,

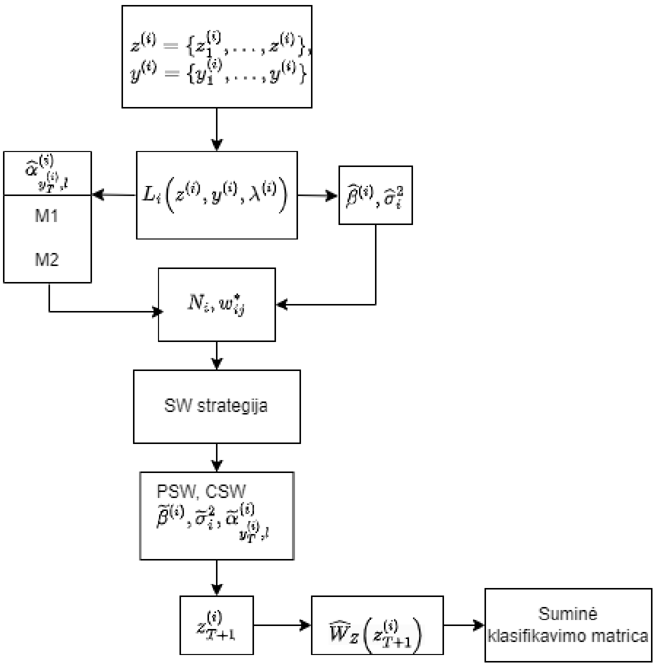
$\widetilde{\Psi}^{(i)}_1 = \Big(\pi^{(i)}, \widetilde{A}^{(i)}_1, \widetilde{\beta}^{(i)}, \widetilde{\sigma}^2_i\Big)$, $\widetilde{\Psi}^{(i)}_2 = \Big(\pi^{(i)}, \widetilde{A}^{(i)}_2, \widetilde{\beta}^{(i)}, \widetilde{\sigma}^2_i\Big)$.

Įterptosios Bajeso diskriminantinės funkcijos, atitinkančios šias aibes yra apibrėžiamos:

1. $\widehat{W}_{ZP1}\Big(z^{(i)}_{T+1}\Big|\widehat{\Psi}^{(i)}_1\Big) = \Big(z^{(i)}_{T+1} - \frac{\left(\widetilde{\beta}^{(i)}\right)' F x_i}{2}\Big)\big((\widetilde{\beta}^{(i)})' G x_i\big)/\widetilde{\sigma}^2_i + ln\left(\frac{\widehat{a}^{(i)}_{y^{(i)}_T,1}}{\widehat{a}^{(i)}_{y^{(i)}_T,0}}\right)$ dalinis erdvinio svorio (*angl. Partial spatial weighting*) įtraukimas su įterptomis perėjimo tikimybių įverčiais M1 ir žymimas PSW1;

2. $\widehat{W}_{ZP2}\left(z_{T+1}^{(i)}\middle|\widehat{\Psi}_2^{(i)}\right) = \left(z_{T+1}^{(i)} - \frac{\left(\widetilde{\beta}^{(i)}\right)' F x_i}{2}\right)((\widetilde{\beta}^{(i)})'Gx_i)/\widetilde{\sigma}_i^2 + ln\left(\frac{\widehat{a}^{(i)}_{y_T^{(i)},1}}{\widehat{a}^{(i)}_{y_T^{(i)},0}}\right)$ dalinis erdvinio svorio įtraukimas su įterptomis perėjimo tikimybių įverčiais M2 ir žymimas PSW2;
3. $\widehat{W}_{ZP3}\left(z_{T+1}^{(i)}\middle|\widetilde{\Psi}_1^{(i)}\right) = \left(z_{T+1}^{(i)} - \frac{\left(\widetilde{\beta}^{(i)}\right)' F x_i}{2}\right)((\widetilde{\beta}^{(i)})'Gx_i)/\widetilde{\sigma}_i^2 + ln\left(\frac{\widetilde{a}^{(i)}_{y_T^{(i)},1}}{\widetilde{a}^{(i)}_{y_T^{(i)},0}}\right)$ pilnas erdvinio svorio (*angl. Complete spatial weighting*) įtraukimas su įterptomis perėjimo tikimybių įverčiais M1 ir žymimas CSW1;
4. $\widehat{W}_{ZP4}\left(z_{T+1}^{(i)}\middle|\widetilde{\Psi}_2^{(i)}\right) = \left(z_{T+1}^{(i)} - \frac{\left(\widetilde{\beta}^{(i)}\right)' F x_i}{2}\right)((\widetilde{\beta}^{(i)})'Gx_i)/\widetilde{\sigma}_i^2 + ln\left(\frac{\widetilde{a}^{(i)}_{y_T^{(i)},1}}{\widetilde{a}^{(i)}_{y_T^{(i)},0}}\right)$ pilnas erdvinio svorio įtraukimas su įterptomis perėjimo tikimybių įverčiais M2 ir žymimas CSW2.

Klasifikatoriaus algoritmas pateiktas S.1 paveiksle.



S.1 pav.: Prižiūrimo Gauso paslėpto Markovo modelio klasifikavimo schema.

S.2.2 Gauso Markovo atsitiktinio lauko-Autoregresinis modelis

Gauso Markovo atsitiktinio lauko-Autoregresinio modelio parametrai i-jam gardelės taške $\Psi^{(i)} = \left(\beta, \sigma_s^2, \alpha_1^{(i)}, \sigma_{T(i)}^2\right)$, $i = 1, \ldots, n$. Parametrų vertinimams pagal mokymo imtį naudojama laiko vidurkinimo (*angl. Temporal averaging, TA*) strategija.

Įverčiai visiems požymio skirstinio parametrams

$$\widetilde{\beta} = \sum_{t=1}^{T} \frac{\widehat{\beta}_{(t)}}{T}, \widetilde{\sigma}_s^2 = \sum_{t=1}^{T} \frac{\widehat{\sigma}_{s(t)}^2}{T},$$

kur maksimalaus tikėtinumo įverčiai $\widehat{\beta}_{(t)} = \left(X'_{(t)} R^{-1} X_{(t)}\right)^{-1} X'_{(t)} R^{-1} Z_t$,

$\widehat{\sigma}_{s(t)}^2 = \left(Z_t - X_{(t)}\widehat{\beta}_{(t)}\right)' R^{-1} \left(Z_t - X_{(t)}\widehat{\beta}_{(t)}\right)/(n - 2q)$.

Maksimalaus tikėtinumo įverčiai $AR(1)$ žymimi $\widehat{\alpha}_1^{(i)}$ ir $\widehat{\sigma}_{T(i)}^2$.

Įterptoji Bajeso diskriminantinė funkcija yra apibrėžiama:

$$\widehat{W}_Z\left(z_{T+1}^{(i)}\right) = \left(z_{T+1}^{(i)} - \frac{\widehat{\mu}_{1i(z)}^{T+1} + \widehat{\mu}_{0i(z)}^{T+1}}{2}\right) \times \\ \times \widehat{\Sigma}_{T+1,i(z)}^{-1} \left(\widehat{\mu}_{1i(z)}^{T+1} - \widehat{\mu}_{0i(z)}^{T+1}\right) - \gamma_i(T+1),$$

kur $\widehat{\mu}_{li(z)}^{T+1} = \left(x_i'\widetilde{\beta}_l + \frac{\eta}{1+\eta h_i} \sum_{j=1}^{n} w_{ij} \left(z_{T+1}^{(j)} - x_j'\widehat{\beta}_{y_{T+1}^{(j)}}\right)\right)$

ir $\widehat{\Sigma}_{T+1,i(z)} = \widehat{\sigma}_{T(i)}^2 \frac{\widetilde{\sigma}_S^2}{1+\eta h_i}$.

Klasifikavimo klaidos įvertinimas, naudojant įterptąją Bajeso diskriminantinę funkciją, yra

$$PBE_i = \pi_1(s_i, T+1)\Phi\left(- \frac{\widehat{\Delta}_{T+1,i(z)}}{2} - \frac{\gamma_i(T+1)}{\widehat{\Delta}_{T+1,i(z)}}\right) \\ + \pi_0(s_i, T+1)\Phi\left(- \frac{\widehat{\Delta}_{T+1,i(z)}}{2} + \frac{\gamma_i(T+1)}{\widehat{\Delta}_{T+1,i(z)}}\right),$$

kur $\widehat{\Delta}_{T+1,i(z)}^2 = ((\Delta\beta)'x_i)^2/(\widetilde{\sigma}_s^2\widehat{\sigma}_{T(i)}^2)$, $l = 0, 1$.

Skaičiuojant $\gamma_i(T+1)$ stebėjimams s_i laike $t = T+1$ naudojami trys modeliai: Globalaus Moran's I, Geary's C ir Getis-Ord G indeksai n sričių laiko momente t yra

$$I(t) = \frac{n}{S_0}\frac{\widetilde{z}_t' W \widetilde{z}_t}{\widetilde{z}_t' \widetilde{z}_t}, C(t) = \frac{(n-1)}{2S_0}\frac{z_t' L z_t}{\widetilde{z}_t' \widetilde{z}_t}, G(t) = \frac{z_t' W z_t}{z_t'(J-I)z_t},$$

kur $\bar{z}_t = \sum\limits_{i=1}^{n} z_t^i/n$ ir $\widetilde{z}_t = z_t - \bar{z}_t 1_n$, $S_0 = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} w_{ij}$, $J = 1_n 1_n'$

Žymių pasiskirstymas pagal Moran's I, Geary's C and Getis-Ord G yra žymimas

$$\pi_{M1t}(s_i, T+1) = \frac{1}{1+exp(-I(T)y_i^*(T))},$$

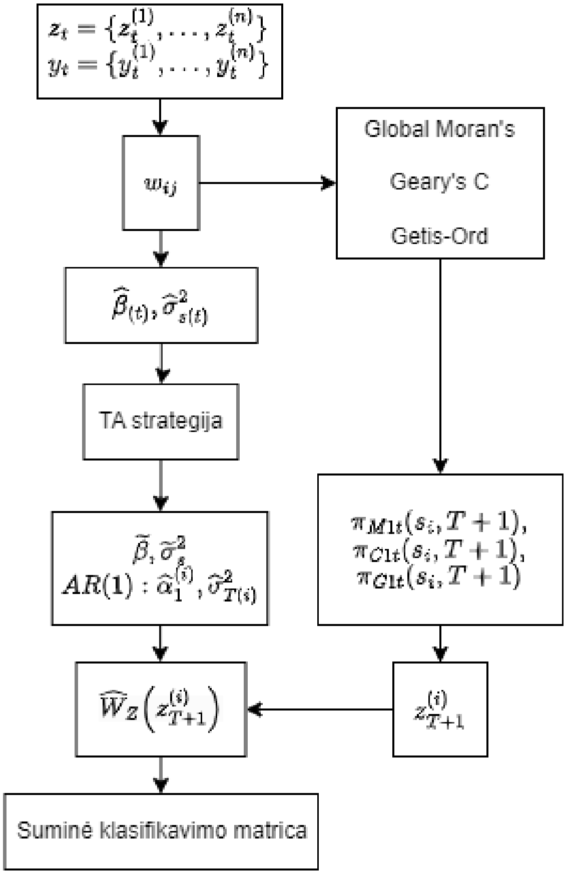
$$\pi_{C1t}(s_i, T+1) = \frac{1}{1+exp(-C(T)y_i^*(T))},$$

$$\pi_{G1t}(s_i, m+1) = \frac{1}{1+exp(-G(m)y_i^*(m))}$$

atitinkamai, kur $y_i^*(t) = 2y_t^{(i)} - 1$.

Žymės prognozavimas taške s_i laike $t = T+1$ yra $\widehat{Y}_{T+1}^{(i)} = H\left(\widehat{W}_Z\left(z_{T+1}^{(i)}\right)\right)$, kur $H(\cdot)$ yra Hevisaido (*angl. Heaviside*) funkcija.

Klasifikatoriaus algoritmas pateiktas S.2 paveiksle.



S.2 pav.: Prižiūrimo Gauso Markovo atsitiktinio lauko-Autoregresinio modelio klasifikavimo schema.

S.2.3 Gauso Geostatistinis-Autoregresinis modelis

Gauso Geostatistinio-Autoregresinio modelio parametrai i-jam gardelės taške $\Psi^{(i)} = \left(\beta, \sigma_s^2, \alpha_1^{(i)}, \sigma_{T(i)}^2\right)$, $i = 1, \ldots, n$. Parametrų vertinimams pagal mokymo imtį naudojama laiko vidurkinimo strategija.

Įverčiai visiems požymio skirstinio parametrams

$$\widetilde{\beta} = \sum_{t=1}^{T} \frac{\widehat{\beta}_{(t)}}{T}, \widetilde{\sigma}_s^2 = \sum_{t=1}^{T} \frac{\widehat{\sigma}_{s(t)}^2}{T},$$

kur maksimalaus tikėtinumo įverčiai $\widehat{\beta}_{(t)} = \left(X'_{(t)}C_s^{-1}X_{(t)}\right)^{-1}X'_{(t)}C_s^{-1}Z_t$ ir $\widehat{\sigma}^2_{s(t)} = \left(Z_t - X_{(t)}\widehat{\beta}_{(t)}\right)'C_s^{-1}\left(Z_t - X_{(t)}\widehat{\beta}_{(t)}\right)/(n-2q)$.

Maksimalaus tikėtinumo įverčiai $AR(1)$ žymimi $\widehat{\alpha}_1^{(i)}$ ir $\widehat{\sigma}^2_{T(i)}$.

Įterptoji Bajeso diskriminantinė funkcija yra apibrėžiama:

$$\widehat{W}_Z\left(z_{T+1}^{(i)}\right) = \left(z_{T+1}^{(i)} - \frac{\widehat{\mu}_{1i(z)}^{T+1} + \widehat{\mu}_{0i(z)}^{T+1}}{2}\right)\times$$
$$\times\, \widehat{\Sigma}^{-1}_{T+1,i(z)}\left(\widehat{\mu}_{1i(z)}^{T+1} - \widehat{\mu}_{0i(z)}^{T+1}\right) - \gamma_i(T+1)$$

Klasifikavimo klaidos įvertinimas, naudojant įterptąją Bajeso diskriminantinę funkciją, yra

$$PBE_i = \pi_0(s_i, T+1)\Phi\left(\widehat{Q}_{0i}\right) + \pi_1(s_i, T+1)\Phi\left(\widehat{Q}_{1i}\right),$$

kur $\widehat{Q}_{li} = -\frac{\widehat{\Delta}_{T+1,i(z)}}{2} + (-1)^l\frac{\gamma_i(T+1)}{\widehat{\Delta}_{T+1,i(z)}}$ ir $\widehat{\Delta}^2_{T+1,i(z)} = ((\Delta\beta)'x_i)^2/(\widetilde{\sigma}_s^2\widehat{\sigma}^2_{T(i)})$, $l = 0, 1$.

Nagrinėjame metodus stebėjimams s_i laike $t = T+1$, kurie skiriasi pagal įtrauktos erdvės-laiko informacijos išraišką, ir apskaičiuojame apriorines tikimybes:

1. Laiko svertinis slenkamasis vidurkis (*angl. Temporal Weighted Moving Average, TWMA*)

$$\pi_{1t}(s_i, T+1) = \frac{\sum\limits_{t=1}^{T} y_t^{(i)}t}{(1+T)T/2};$$

2. Laiko ir erdvės svertinis slenkamasis vidurkis (*angl. Spatial Temporal Weighted Moving Average, STWMA*):

$$\pi_{1ts}(s_i, T+1) = \frac{\sum\limits_{t=1}^{T} y_t^{(i)}t + \sum\limits_{j:s_j\in N_i}\sum\limits_{t=1}^{T} y_t^{(j)}t}{(1+T)T/2(1+card(N_i))},$$

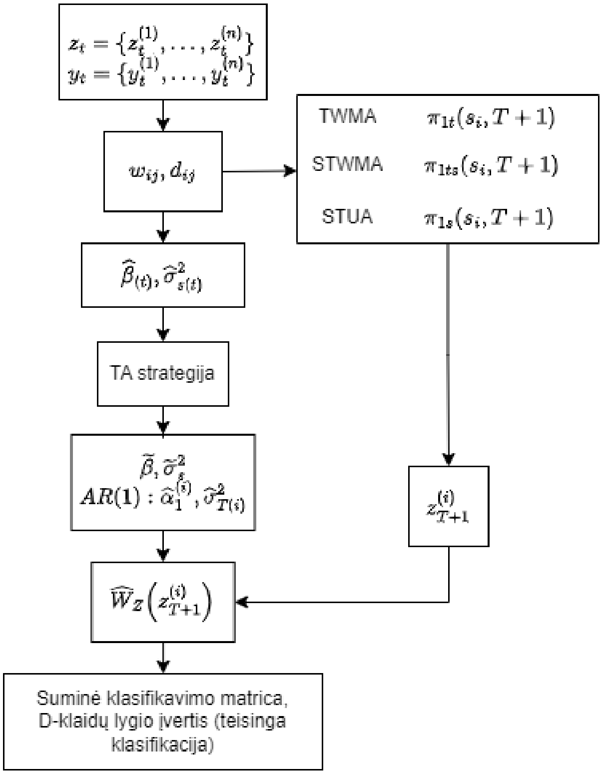
   kur j žymi s_i artimiausio kaimyno indeksą;

3. Laiko ir erdvės nesvertinis vidurkis (*angl. Spatial Temporal Un-*

weighted Average, STUA):

$$\pi_{1s}(s_i, T+1) = \frac{\sum\limits_{t=1}^{T} \left(y_t^{(i)} + \sum\limits_{j:s_j \in N_i} y_t^{(j)} \right)}{T(1 + card(N_i))}.$$

Klasifikatoriaus algoritmas pateiktas S.3 paveiksle.



S.3 pav.: Prižiūrimo Gauso Geostatistinio-Autoregresinio modelio klasifikavimo schema.

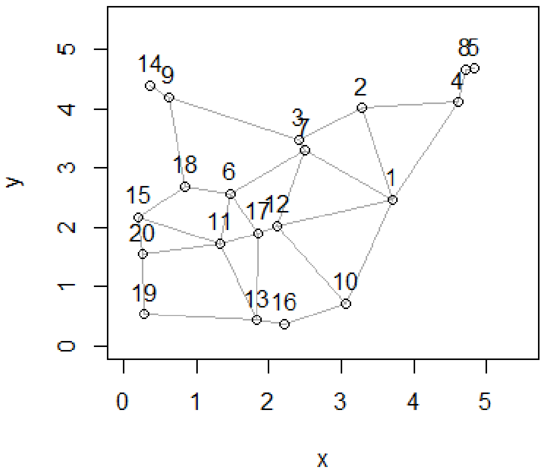
Išsamesni modelių aprašymai ir skyriaus išvados pateikiamos disertacijos tekste ir publikacijose [28, 43, 44].

S.3 EMPIRINIS TYRIMAS GENERUOTIEMS IR REALIEMS DUOMENIMS

S.3.1 Generuoti duomenys

Norint palyginti 10 siūlomų klasifikatorių, atliktas tyrimas su generuotais duomenimis, kurių $n = 20$ taškai, pasiskirstę $S \in [0,5] \times [0,5]$. Duomenys yra sugeneruoti kiekviename taške laiko momentais $t = 1, \dots, 50$ (t.y. $T = 50$) ir sudaryti iš požymio ir klasės žymių stebėjimų. Požymio kintamasis yra pasiskirstęs pagal normalųjį skirstinį su vidurkiu pagal pirmos eilės trendo paviršiaus modelį, susidedantį iš trijų kintamųjų, t. y. pirmasis kintamasis yra lygus 1, o kiti x ir y koordinatės.

Tyrimas atliktas 20 taškų (generuotų) dvimatėje erdvėje, pavaizduotame S.4 paveiksle. Kaimynystės sistemą nurodo grafo briaunos.



S.4 pav.: Erdvinis duomenų rinkinys S_{20}.

S.2 lentelėje pateikta kaimynų skaičius bei trumpiausias ir ilgiausias atstumas tarp kaimynų.

S.2 lentelė: Kaimynų skaičius bei trumpiausias ir ilgiausias atstumas tarp kaimynų.

i	$card(N_i)$	$max_{ij}d_{ij}$	$min_{ij}d_{ij}$
1	5	3,9422	1,4609
2	3	4,6089	1,0276
3	3	3,6442	0,1853
4	2	5,6251	0,5598
5	1	6,1610	0,1102
6	4	3,9578	0,6295
7	4	3,5736	0,1853
8	2	6,0702	0,1102
9	2	4,2413	0,3379
10	3	4,5645	0,9079
11	5	4,5812	0,5571
12	4	3,7873	0,2966
13	4	5,1973	0,3851
14	1	4,5645	0,3379
15	3	5,2501	0,6355
16	2	5,0436	0,3851
17	4	4,0715	0,2966
18	2	4,4408	0,6295
19	2	6,1610	1,0151
20	3	5,5363	0,6355

Siūlomo prižiūrimo generatyvinio klasifikavimo modelio generuotiems duomenims algoritmas pateiktas 2.1 skyrelyje.

Iš pradžių sugeneruojame $v = 100$ modeliavimo paleidimų (pakartojimų) testavimo žymei pagal aukščiau aprašytą Bernulio skirstinį ir atitinkamą sąlyginį Gauso stebėjimą.

S.3.1.1 Erdvinio Gauso paslėpto Markovo modelio realizacija generuotiems duomenims

Sąlyginis $Y_{T+1}^{(i)}$ skirstinys, kai $Y_T^{(i)} = 1$, yra Bernulio su parametru $\widehat{a}_{11}^{(i)}$, t. y. $\big(Y_{T+1}^{(i)}|Y_T^{(i)} = 1\big) \sim Ber\big(\widehat{a}_{11}^{(i)}\big)$. Akivaizdu, kad $\big(Y_{T+1}^{(i)}|Y_T^{(i)} = 0\big) \sim Ber\big(\widehat{a}_{00}^{(i)}\big)$.

Visiems 20 taškams erdviniai svoriai Gauso pasiskirstymo parametrų ir perėjimo tikimybių įverčiai pateikti disertacijos darbe, 3.1 ir 3.2 lentelėse.

Apskaičiuojame kiekvieno taško tikslumo matus ACC, BAC, GAC ir pateikiame jų erdvines vertes S.3 lentelėje.

S.3 lentelė: Klasifikatorių, pagrįstų įterptosiomis Bajeso diskriminantinėmis funkcijomis, tikslumo matai generuotiems duomenims.

lr		**PSW1**	**CSW1**	**PSW2**	**CSW2**
0,2	ACC	0,8000	0,9500	0,8000	0,8500
	BAC	0,7619	0,9643	0,8095	0,8452
	GAC	0,7559	0,9636	0,8092	0,8452
0,4	ACC	0,8500	0,8500	0,8000	0,9000
	BAC	0,8434	0,8434	0,7980	0,8990
	GAC	0,8409	0,8409	0,7977	0,8989
0,5	ACC	0,8000	0,8000	0,8000	0,8500
	BAC	0,8125	0,8125	0,8125	0,8542
	GAC	0,8101	0,8101	0,8101	0,8539
0,6	ACC	0,8000	0,8500	0,8000	0,8500
	BAC	0,8081	0,8535	0,8081	0,8535
	GAC	0,8040	0,8528	0,8040	0,8528
0,8	ACC	**0,9000**	**0,8500**	0,9500	0,9500
	BAC	**0,8810**	**0,8452**	0,9167	0,9167
	GAC	**0,8797**	**0,8452**	0,9129	0,9129

Kaip matome iš S.3 lentelės, klasifikatoriai, pagrįsti paslėptu Markovo modeliu su pilnu erdvinio svorio įtraukimu, daugeliu atvejų turi didesnį tikslumą, palyginti su daliniu erdvinio svorio įtraukimu pagal tikslumo matą ACC, taip pat BAC ir GAC įvairioms lr reikšmėms. Šios taisyklės išimtys lentelėje paryškinti skaičiai.

M1 taisyklei pilnas erdvinio svorio įtraukimas rodo didesnį tikslumą prieš dalinį erdvinio svorio įtraukimą visoms lr reikšmėms, išskyrus $lr = 0{,}8$, o taisyklei M2 pilno erdvinio svorio įtraukimas rodo didesnį tikslumą prieš dalinį erdvinio svorio įtraukimą visoms lr reikšmėms.

S.3.1.2 Gauso Markovo atstiktinio lauko-Autoregresinio modelio realizacija generuotiems duomenims

Siūlomo klasifikatoriaus reikšmių tikslumo matai, nurodyti lygtyse (S.1), (S.2), (S.3) pateikti S.4 lentelėje.

S.4 lentelė: Klasifikatorių, pagrįstų įterptosiomis Bajeso diskriminantinėmis funkcijomis, tikslumo matai su apriorinėmis tikimybėmis.

lr		**GM**	**GO**	**GC**
0,2	ACC	0,8000	0,8000	0,8000
	BAC	0,8333	0,8333	0,8333
0,4	ACC	0,9000	0,9000	**0,8500**
	BAC	0,8889	0,8889	**0,8333**
0,5	ACC	0,8000	0,8000	0,8000
	BAC	0,7500	0,7500	0,7500
0,6	ACC	0,7000	0,7000	0,7000
	BAC	0,7857	0,7857	0,7857
0,8	ACC	0,8000	**0,7500**	0,8000
	BAC	0,8462	**0,8077**	0,8462

S.4 lentelėje parodytas klasifikatorių tikslumas pagal tris asociacijos indeksus, viršijančius $0{,}75$ daugiau nei 90% atvejų. Pastebėta, kad klasifikatorių, pagrįstų skirtingais asociacijos indeksais, tikslumas beveik visais atvejais mažai skiriasi. Šio teiginio išimtis lentelėje pavaizduota paryškintais skaičiais.

S.3.1.3 Gauso Geostatistinio-Autoregresinio modelio realizacija generuotiems duomenims

Siūlomo klasifikatoriaus reikšmių tikslumo matai, nurodyti lygtyse (S.1), (S.2), (S.3) pateikti S.5 lentelėje.

Kaip matyti iš S.5 lentelės, geriausi atvejai su didžiausiomis klasifikavimo tikslumo matų reikšmėmis yra paryškinti skaičiai. STWMA metodas turi didesnį tikslumą prieš kitus du metodus, kai $lr = 0{,}5$, $0{,}6$, $0{,}8$. Kai $lr = 0{,}2$, TWMA metodas tikslesnis prieš kitus du metodus.

S.5 lentelė: Klasifikatorių, pagrįstų įterptosiomis Bajeso diskriminantinėmis funkcijomis, tikslumo matai su apriorinėmis tikimybėmis.

lr		**TWMA**	**STWMA**	**STUA**
0,2	ACC	**0,8500**	0,6500	0,8000
	BAC	**0,6373**	0,5625	0,5000
0,4	ACC	**0,8500**	**0,8500**	0,7000
	BAC	**0,8542**	**0,8434**	0,7033
0,5	ACC	0,8000	**0,8000**	0,7000
	BAC	0,7879	**0,8000**	0,8000
0,6	ACC	0,8500	**0,9000**	0,7000
	BAC	0,8333	**0,9286**	0,6667
0,8	ACC	0,9000	**1,0000**	0,9000
	BAC	0,5000	**1,0000**	0,8750

S.3.1.4 Erdvinio Gauso Geostatistinio-Autoregresinio modelio realizacija generuotiems duomenims

Nagrinėjamas generuotų Gauso erdvės-laiko duomenų klasifikavimo uždavinys su eksponentiniu modeliu ir stacionaria $AR(1)$ laiko kovariacija. Yra žinoma, kad šiam modeliui $c_T^{1,1} = c_T^{t,t}$ $t = 2, \ldots, T+1$ parametras α kiekybiškai nustato laiko priklausomybę pagal $c_T^{1,1} = \frac{\sigma_T^2}{1-\alpha^2}$, kur σ_T^2 yra balto triukšmo dispersija.

Tyrimo metu atsižvelgiama į dvi izotropines erdvės kovariacijų struktūras, priklausiančias eksponentinei šeimai. Tariame, kad $C_s = \sigma_s^2 R$ – erdvės kovariacijų matrica, kur $R = (r_{ij})$ erdvės koreliacijų matrica. Nagrinėjame du atvejus:

1. Eksponentinis $r_{ij} = r(|s_i - s_j|) = e^{-|s_i - s_j|/\varphi}$;
2. Kvadratinis ekponentinis $r_{ij} = r(|s_i - s_j|) = e^{-(|s_i - s_j|/\varphi)^2}$.

Čia φ yra rango parametras, kuris atspindi erdvinę priklausomybę.

Nagrinėjame metodus (TWMA, STWMA), kurie apskaičiuoja apriorines tikimybes.

Palyginome šiuos 4 konkrečius atvejus, apskaičiuodami PBE_i ir PIC_i, $i = 1, \ldots, n$ ir pateikėme rezultatus lentelėse.

Klasių suskirstymas 20 taškų ir 4 laiko momentams pateikiamas S.6 lentelėje.

S.6 lentelė: Klasių suskirstymas 20 taškų 4 laiko momentuose.

t	i									
	1	2	3	4	5	6	7	8	9	10
1	0	1	0	1	1	0	0	1	0	0
2	0	0	0	0	1	0	1	0	0	0
3	0	0	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	0	0	1	0
t	i									
	11	12	13	14	15	16	17	18	19	20
1	0	0	0	0	0	0	0	1	0	0
2	0	0	1	0	0	0	0	0	0	0
3	0	0	1	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0	1	0	0

Tegu $\Delta = \left|\mu_{1i(z)}^{T+1} - \mu_{0i(z)}^{T+1}\right|$.

Bajeso klaidos tikymybės PBE_i ir jų vidurkiai $APBE = \sum_{i=1}^{20} \frac{PBE_i}{20}$ dviem erdvės kovariacijų atvejais ir dviem apriorinių tikimybių modeliais pateikiamos 3.7 lentelėje 2 skyriuje.

Kaip matyti iš 3.7 lentelės 2 skyriuje, $\alpha = 0{,}1$, $0{,}3$, klasifikavimas su STWMA daugelyje taškų turi didesnį tikslumą prieš TWMA atvejus. Esant didelėms α reikšmėms, skirtumo tarp šių metodų nepastebėta.

$v = 30$ nepriklausomų pakartojimų, empirinės klaidos PIC_i ir jų vidurkiai $APIC = \sum\limits_{i=1}^{20} \frac{PIC_i}{20}$, dviem erdvės kovariacijų atvejais ir dviem apriorinių tikimybių modeliais pateikiamos 3.8 lentelėje 2 skyriuje.

Kaip matyti iš 3.8 lentelės 2 skyriuje, visoms α, klasifikavimas su STWMA, TWMA turi panašias empirines klaidas.

3.7 ir 3.8 lentelių paskutinių eilučių (t.y., $APBE$ ir $APIC$) leidžia palyginti ir analizuoti Bajeso ir empirinių klaidų vidurkius nagrinėjamuose atvejuose.

S.3.2 Realūs duomenys

Atliekama Lietuvos Respublikos Higienos instituto surinktų 60 savivaldybių metinių mirtingumo duomenų skaitinė analizė 2001–2019 m. laikotarpiu.

Apytikslis kiekvienos savivaldybės mirtingumas, matuojamas vienetais šimtui tūkstančių gyventojų, yra laikomas kintamojo Z reikšme. Trijų klasių žymių kintamasis Y nurodo mirtingumą dėl ūminio kardiovaskulinio įvykio (*angl. Acute cardiovascular event, ACE*) ir kraujotakos sistemos ligų (*angl. Diseases of circulatory system, CSD*).

Savivaldybių kaimynų skaičius yra nuo 1 iki 9, A priede pateikiamas savivaldybių kaimynų sąrašas.

Atvejai, kurių indekso reikšmės mažesnės už slenkstį, reikšmė lygi 0, ir priešingu atveju reikšmė lygi 1. Čia nagrinėjamas atvejis su pastoviu vidurkiu, t.y. $\mu(s;t) = \beta_l$ su $x(s) = 1$.

Mokymui naudojami 2001-2018 metų laikotarpio duomenys ($t = 1, \ldots, 18$), o likę (2019 metai) yra naudojami testavimui. Tai $T = 18$ ir $n = 60$, t. y. yra $18 \cdot 60$ mokymo stebėjimai ir 60 testavimui.

Pažymėkime $lr = \sum\limits_{i=1}^{n} \Big(\sum\limits_{t=1}^{T} I(Y_t^{(i)} = 1)/T \Big)/n$.

IR_t reikšmės pateiktos S.7 lentelėje.

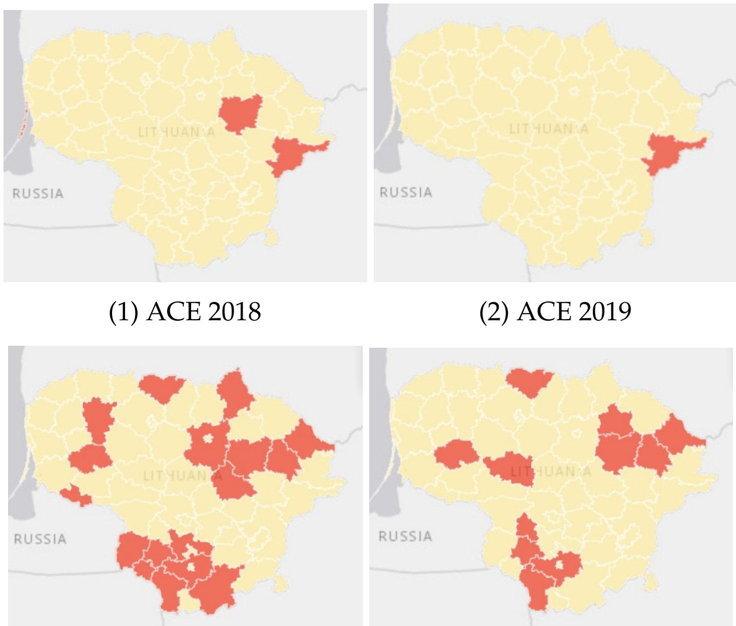
S.7 lentelė: Metinis disbalanso koeficientas dėl įvairių mirtingumo priežasčių (klasės žymių kintamieji).

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
ACE	0,36	0,40	0,62	0,40	0,46	0,43	0,30	0,36	0,54	0,22
CSD	0,94	0,67	0,76	0,62	0,62	0,46	0,46	0,62	0,71	1,07

	2011	2012	2013	2014	2015	2016	2017	2018	2019
ACE	0,28	0,25	0,22	0,17	0,09	0,07	0,09	0,05	0,01
CSD	1,07	0,71	0,82	0,67	0,94	0,62	0,58	0,40	0,22

Kaip matome iš S.7 lentelės, daugeliu laikotarpiu CSD turi aukščiausius IR, o ACE atvejais turi mažiausius IR.

Grafinis 60 savivaldybių dviejose klasėse suskirstymas pateikiamas S.5 paveiksle ir mokymo imties žymių kitimas laike pateiktas priede B.



(1) ACE 2018 (2) ACE 2019

(3) CSD 2018 (4) CSD 2019

S.5 pav.: Suskirstytos Lietuvos savivaldybės. Geltonos spalvos sritys nurodo savivaldybes, kuriose mirtingumas nuo ligos yra mažas (su reikšme 0), o raudonos sritys nurodo savivaldybes, kuriose mirtingumas nuo ligos yra aukštas (su reikšme 1).

Siūlomo prižiūrimo generatyvinio klasifikavimo modelio realiems duomenims algoritmas pateiktas 2.2. skyrelyje.

S.3.2.1 Gauso paslėpto Markovo modelio realizacija realiems duomenims

Tada $i = 1, \ldots, 60$

$$X^{(i)} = \begin{pmatrix} 1-y_1^{(i)} & y_1^{(i)} \\ 1-y_2^{(i)} & y_2^{(i)} \\ \vdots & \vdots \\ 1-y_T^{(i)} & y_T^{(i)} \end{pmatrix}.$$

Siūlomų klasifikatorių, aprašytų lygtimis (S.1), (S.2), (S.3), TPR ir TNR rezultatai pateikti S.8 lentelėje.

S.8 lentelė: Klasifikatorių, pagrįstų įterptosiomis Bajeso diskriminantinėmis funkcijomis, tikslumo matai realiems duomenims.

		PSW1	**CSW1**	**PSW2**	**CSW2**
	ACC	0,7167	0,8333	0,9167	0,9500
	BAC	0,8559	0,9153	0,9576	0,9746
ACE	GAC	0,8437	0,9113	0,9567	0,9742
	TPR	0,7119	0,8305	0,9153	0,9492
	TNR	1,0000	1,0000	1,0000	1,0000
	ACC	0,6833	0,8000	**0,8500**	**0,7833**
	BAC	0,6651	0,7718	**0,8377**	**0,7616**
CSD	GAC	0,6645	0,7705	**0,8374**	**0,7608**
	TPR	0,6939	0,8163	**0,9571**	**0,7959**
	TNR	0,6364	0,7273	**0,8182**	**0,7273**

Kaip galima matyti iš S.8 lentelės, metodai, pagrįsti paslėptu Markovo modeliu su pilnu erdvinio svorio įtraukimu, daugeliu atvejų turi didesnį tikslumą, palyginti su daliniu erdvinio svorio įtraukimu pagal tikslumo matą ACC, taip pat BAC ir GAC įvairioms lr reikšmėms. Šios taisyklės išimtys lentelėje paryškinti skaičiai.

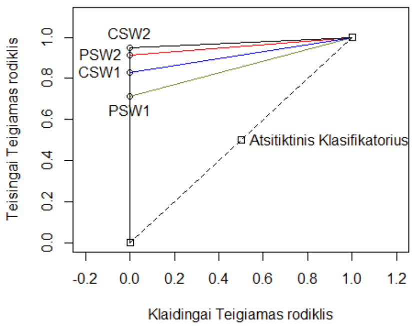
Dėl ligos ACE pagal M1 ir M2 taisykles pilnas erdvinio svorio įtraukimas rodo didesnį tikslumą prieš dalinį erdvinio svorio įtraukimą.

Ligai CSD taisyklei M1 pilnas erdvinio svorio įtraukimas rodo didesnį tikslumą prieš dalinį erdvinio svorio įtraukimą, o taisyklei M2 dalinis erdvinio svorio įtraukimas rodo didesnį tikslumą prieš pilną erdvinio svorio įtraukimą.

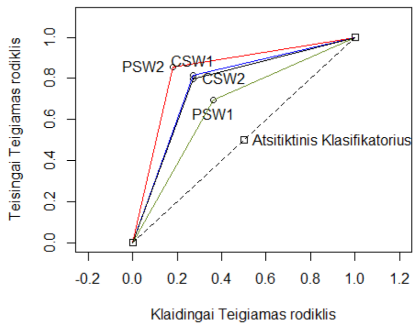
S.6 paveiksle pateiktas vizualizavimo grafikas, kuriame vaizduojamas klasifikatoriaus tikslumai, nagrinėjant skirtingus klasifikatorius.

Pavaizduoti taškai reiškia 4 nagrinėjamus klasifikatorius ir atsitiktinį klasifikatorių, pavaizduotą punktyrine linija. Nesunku patikrinti, ar plotas po kreive yra lygus tikslumo matui BAC.

Kaip matyti iš S.6 paveikslo, klasifikatorius CSW2 rodo didesnį tikslumą prieš kitus, nes jo plotas po kreive yra didžiausias ir lygus $0{,}9576$ (S.8 lentelė).



(1) ACE



(2) CSD

S.6 pav.: Vizualizavimo grafikas paslėptam Markovo modeliui su klasių žymių kintamaisiais.

S.3.2.2 Gauso Markovo atstiktinio lauko-Autoregresinio modelio realizacija realiems duomenims

Gautų rezultatų skaitmeninei iliustracijai nagrinėjome Gauso erdvės-laiko modelį su grynomis erdvinėmis eksponentinėmis kovariacijomis ir gryna stacionarios $AR(1)$ laiko kovariacija.

Tada $t = 1, \ldots, 18$

$$X_{(t)} = \begin{pmatrix} 1 - y_t^{(1)} & y_t^{(1)} \\ 1 - y_t^{(2)} & y_t^{(2)} \\ \vdots & \vdots \\ 1 - y_t^{(n)} & y_t^{(n)} \end{pmatrix}.$$

Siūlomo klasifikatoriaus reikšmių tikslumo matai, nurodyti lygtyse (S.1), (S.2), (S.3) pateikti S.9 lentelėje.

S.9 lentelė: Klasifikatorių, pagrįstų įterptosiomis Bajeso diskriminantinėmis funkcijomis, tikslumo matai su apriorinėmis tikimybėmis.

		GM	**GO**	**GC**
	ACC	**0,9667**	**0,9667**	**0,9667**
	BAC	**0,9831**	**0,9831**	**0,9831**
ACE	GAC	**0,9829**	**0,9829**	**0,9829**
	TPR	**0,9661**	**0,9661**	**0,9661**
	TNR	**1,0000**	**1,0000**	**1,0000**
	ACC	0,8000	**0,8333**	**0,8333**
	BAC	0,7718	**0,7922**	**0,7922**
CSD	GAC	0,7705	**0,7895**	**0,7895**
	TPR	0,8163	**0,8571**	**0,8571**
	TNR	0,7273	**0,7273**	**0,7273**

S.9 lentelėje parodytas klasifikatorių tikslumas pagal tris asociacijos indeksus, viršijančius $0{,}75$ daugiau nei 90% atvejų. Pastebėta, kad klasifikatorių, pagrįstų skirtingų asociacijos indeksų, tikslumas beveik visais atvejais mažai skiriasi. Šio teiginio išimtis lentelėje pavaizduota paryškintais skaičiais.

IŠVADOS

Darbe siūlomas naujas metodas generatyvinių modelių pritaikymui prižiūrimam Bajesiniam klasifikavimui, pagrįstas trijų tipų Gauso erdvėslaiko duomenų modeliais, pasižyminčiais skirtingomis statistinės priklausomybės formomis ir įvairiomis strategijomis, skirtomis požymių ir klasių žymių pasiskirstymo parametrams įvertinti. Tai panašių tyrimų, anksčiau atliktų grynai erdviniame kontekste, išplėtimas. Jautrumo analizė ir detalus klasifikatorių palyginimas kiekvienam duomenų modelių tipui, atliekamas pagal keletą tikslumo matų.

Toliau pateikiamos atlikto darbo išvados:

1. Gauso paslėpto Markovo modelio skaičiavimo rezultatai su generuotais ir realiais duomenimis parodė, kad erdvinio svorio įtraukimas paveikė siūlomų klasifikatorių kokybę.

2. Gauso Markovo atsitiktinio lauko-Autoregresinio duomenų modelio skaičiavimai parodė klasifikatorių tikslumą pagal tris asociacijos indeksus, viršijančius $0{,}75$ daugiau nei 90% atvejų. Pastebėta, kad klasifikatorių, pagrįstų skirtingų asociacijos indeksų, tikslumas beveik visais atvejais mažai skiriasi.

3. Siūlomas prižiūrimo Bajesinio klasifikavimo metodas, pritaikytas Gauso Geostatistiniam-Autoregresiniam modeliui, atskleidžia atskirą klasifikatorių, pagrįstą klasių žymių skirstiniais kompleksiškai įtraukiančiais erdvės-laiko kontekstą, kuris rodo didesnį tikslumą prieš atvejus, orientuotus tik į laiko kontekstą.

4. Skaitinė klasifikatorių analizė, pagrįsta Gauso Geostatistiniu-Autoregresiniu modeliu, rodo, kad klasifikatoriai su erdvine eksponentine kovariacija turi didesnį tikslumą prieš klasifikatorių su eksponentine kvadratine kovariacija.

5. Išvestos analitinės lokalių Bajeso klaidų tikimybių ir jų D-įvertinių išraiškos gali būti tiesiogiai naudojamos pasirenkant klasifikuojamo stebinio padėtį erdvėje.

6. Detalus klasifikatorių palyginimas kiekvieno tipo duomenų modeliuose, nurodytuose pagal skirtingus erdvinės ir laiko informacijos įtraukimo būdus, parodo vieno pranašumus prieš kitus ir atskleidžia naujas taikymo galimybes ir rekomendacijas vartotojams.

About the author

Marta Karaliutė was born in Klaipėda district, Lithuania in 1984. In 2003 she graduated from "Vaivorykštės" high school in Gargždai. She obtained BSc degrees of mathematics in 2007 and MSc degrees of statistics in 2009 at Klaipėda University. In 2017-2022 she studied in PhD study program of informatics in Vilnius University. Since 2009 she teaches at Klaipėda University.

Trumpos žinios apie autorę

Marta Karaliutė gimė 1984 m. birželio 30 d. Klaipėdos rajone. 2003 m. baigė Gargždų "Vaivorykštės" gimnaziją. 2003-2007 m. studijavo matematiką Klaipėdos universitete ir įgijo matematikos bakalauro laipsnį. 2007-2009 m. ten pat studijavo statistiką ir operacijos tyrimą ir įgijo statistikos magistro laipsnį. 2017-2022 m. studijavo informatikos doktorantūroje Vilniaus universitete. Nuo 2009 m. dėsto Klaipėdos universitete.

Acknowledgements

First of all, I would like to sincerely thank my academic supervisor Prof. Dr. Kęstutis Dučinskas for comprehensive support and confidence, for shared knowledge and scientific guidance, for showing direction, and for all the help provided while preparing this dissertation.

Thank you to the academic consultant Prof. Habil. Dr. Gintautas Dzemyda for valuable suggestions and comments.

I am also thankful to dissertation reviewers Prof. Dr. Audronė Jakaitienė and Prof. Dr. Igoris Belovas for accurate insights, critical remarks and beneficial suggestions for improving this work.

I sincerely thank my family members and friends for their support, even in the most challenging moments, which motivates me to move forward and not give up.

PUBLICATIONS BY THE AUTHOR

Papers in periodic scientific journals indexed in Web of Science and Scopus database:

1. Marta Karaliutė, Kęstutis Dučinskas. Classification of Gaussian spatio-temporal data with stationary separable covariances. Nonlinear analysis: modelling and control. Vilnius 26 (2), 2021, p. 363-374. doi: 10.15388/namc.2021.26.22359.
2. Kęstutis Dučinskas, Marta Karaliutė, Laura Šaltytė-Vaisiauskė. Spatially Weighted Bayesian Classification of Spatio-Temporal Areal Data Based on Gaussian-Hidden Markov Models. Mathematics 11(2), 2023, 347. https://doi.org/10.3390/math11020347
3. Marta Karaliutė, Kęstutis Dučinskas. Performance of the supervised generative classifiers of spatio-temporal areal data using various spatial autocorrelation indexes, Nonlinear Analysis: Modelling and Control, 28(2), 2023, p. 1-14. doi: 10.15388/namc.2023.28.31434.

Paper in a periodic scientific journal:

1. Marta Karaliutė, Kęstutis Dučinskas. Supervised linear classification of Gaussian ST data. Lietuvos matematikos rinkinys, 62(A), 2021, pp. 9-15. DOI: 10.15388/LMR.2021.25214

Papers in peer-reviewed scientific conference proceedings:

1. Marta Karaliutė, Kęstutis Dučinskas, Laura Šaltytė-Vaisiauskė. Expected error regret in linear discrimination of balanced spatial

Gaussian time series. DAMSS 2018: 10th international workshop on „Data analysis methods for software systems“: Druskininkai, Lithuania, November 29 - December 1, 2018, p. 42 [abstract book].

2. Marta Karaliutė, Kęstutis Dučinskas; Laura Šaltytė-Vaisiauskė. Expected error rate in linear discrimination of balanced spatial Gaussian time series. Computer data analysis and modeling: stochastics and data science : proceedings of the XII international conference: Minsk, September 18-22, 2019, p. 172-175. http://elib.bsu.by/bitstream/123456789/233358/1/172-175.pdf

Marta Karaliutė

Supervised Bayesian classification methods of Gaussian
Spatio-temporal data based on generative machine learning models

Doctoral Dissertation
Natural Sciences
Informatics (N 009)
Thesis Editors: Sandra Burokaitė and Kevin Samuel Martin

Erdvės-laiko Gausinių duomenų prižiūrimo Bajesinio klasifikavimo
metodai, pagrįsti generatyviniais mašininio mokymosi modeliais

Daktaro disertacija
Gamtos mokslai
Informatika (N 009)
Santraukos redaktorė: Laura Šaltytė-Vaisiauskė

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius
El. p. info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Tiražas 20 egz.