Course Title (in English)	Advanced Statistical Methods
Course Title (in Russian)	Дополнительные главы математической статистики
Lead Instructor	Spokoiny, Vladimir

Contact Person	Vladimir Spokoiny
Contact Person's E-mail	v.spokoiny@skoltech.ru

Course Description

This course introduces the main notions, approaches, and methods of nonparametric statistics. The main topics include maximum likelihood and penalized maximum likelihood estimation, model selection and parameter tuning. We also discuss Bayesian inference and Bayesian model selection. The study is mainly limited to regression models. The topics of this course form an essential basis for working with complex data structures using modern statistical tools.

Course structure: lectures, seminars, project, exam.

Course Prerequisites / Recommendations	Probability theory, linear algebra, mathematical analysis, basics of parametric statistics.

Аннотация

В рамках курса обсуждаются основные понятия, подходы и методы непараметрической статистики, в том числе оценки максимального правдоподобия и пенализованного максимального правдоподобия, выбор модели, настройка параметров. Также будет уделено внимание байесовскому выводу и байесовскому выбору модели. Большинство методов рассматривается в применении к задаче регрессии. Успешное освоение курса формирует у студента базис для работы со сложными структурами данных с использованием современных инструментов математической статистики.

Структура курса: лекции, семинары, проект, экзамен.

Course Academic Level	Master-level
Number of ECTS credits	3

Topic	Summary of Topic	Lectures (# of hours)	Seminars (# of hours)	Labs (# of hours)
Nonparametric Regression	Regression models: design, errors, and response function. Projection estimation. The case of orthogonal design. Bias, variance, risk of estimation, rate and accuracy, smoothness classes.	4	4	0
Regularization and roughness penalty	Ridge regression, roughness penalty, penalized maximum likelihood estimation, impact of regularization, modeling bias, complexity, bias-variance trade-off.	4	4	0
Model selection by SURE and AIC, Cross validation. Complexity penalization and LASSO.	Problem of model choice. Penalized model selection. Akaike criterion. Stein unbiased risk estimation. Parameter tuning by cross-validation. Complexity penalization, hard thresholding. Sparse penalization, soft-thresholding, LASSO.	4	4	0
Bayesian Model section	Bayes risk and Bayes decision. Full and empirical Bayes approach. Theoretic information inequality.	4	4	0
Inference for stochastically linear models.	Concentration of PMLE, Fisher and Wilks expansion, concentration and Gaussian approximation of the posterior.	4	4	0

Assignment Type	Assignment Summary
Homework Assignments	Nonparametric Regression - Purpose of the task: show the ability to solve mathematical problems on the topic, work with information. Requirements: are given in Section 10. Deadline: the task should be handed over until the end of the first week of the course. The sample of the task structure is given below in Table 1.
Homework Assignments	Regularization and roughness penalty - Purpose of the task: show the ability to solve mathematical problems on the topic, work with information. Requirements: are given in Section 10. Deadline: the task should be handed over until the end of the first week of the course. The sample of the task structure is given below in Table 1.
Homework Assignments	Model selection by SURE and AIC, Cross validation - Purpose of the task: show the ability to solve mathematical problems on the topic, work with information. Requirements: are given in Section 10. Deadline: the task should be handed over until the end of the second week of the course. The sample of the task structure is given below in Table 1.
Homework Assignments	Model section by smallest accepted - Purpose of the task: show the ability to solve mathematical problems on the topic, work with information. Requirements: are given in Section 10. Deadline: the task should be handed over until the end of the second week of the course. The sample of the task structure is given below in Table 1.
Homework Assignments	Wavelet methods: hard and soft thresholding - Purpose of the task: show the ability to solve mathematical problems on the topic, work with information. Requirements: are given in Section 10. Deadline: the task should be handed over until the end of the third week of the course. The sample of the task structure is given below in Table 1.
Homework Assignments	Density model: Kernel and projection methods - Purpose of the task: show the ability to solve mathematical problems on the topic, work with information. Requirements: are given in Section 10. Deadline: the task should be handed over until the end of the third week of the course. The sample of the task structure is given below in Table 1.
Final Project	Purpose of the task: show the ability to implement and apply statistical methods. Requirements: code and Power Point/LaTeX/Jupyter Notebook presentation describing the results. Deadline: the task should be handed over until the end of the fourth week of the course.
Final Exam	Written exam.

Type of Assessment	Graded
Grade Structure	

Activity Type	Activity weight, %
Homework Assignments	30
Final Exam	40
Final Project	30

A:	86
B:	76
C:	66
D:	56
E:	46
F:	0
Attendance Requirements	Mandatory
Course Stream	Science, Technology and Engineering (STE)
Course Term (in context of Academic Year)	Term 3
Course Delivery Frequency	Every year

Students of Which Programs do You Recommend to Consider this Course as an Elective?

Masters Programs	PhD Programs
Data Science	

Course Tags	Math

Required Textbooks	ISBN-13 (or ISBN-10)
Wassermann, L. "All of nonparametric statistics." (2006). Springer	978-1-4419-2044-7
Rohde, Charles A. Introductory statistical inference with the likelihood function. Springer International Publishing, 2014	978-3-319-37481-9

Recommended Textbooks	ISBN-13 (or ISBN-10)
Tsybakov, Alexandre B. "Introduction to nonparametric estimation" (2009)	978-1-4419-2709-5
Spokoiny, V. and Dickhaus, T. “Basics of modern mathematical Statistics” (2015) Springer	978-3642399084

Web-resources (links)	Description
http://ieeexplore.ieee.org/Xplore/home.jsp	IEEE
http://www.scopus.com/	Scopus
www.springer.com	Springer
http://www.sciencedirect.com/	ScienceDirect
http://www.gpntb.ru	State Public Scientific and Technical Library of Russia
https://www.esaim-ps.org/articles/ps/pdf/2005/01/ps0420.pdf	Stéphane Boucheron, Olivier Bousquet and Gábor Lugosi “Theory of Classification: a Survey of Some Recent Advances”, ESAIM: P&S, 2005, Vol. 9, p. 323-375;
https://link.springer.com/content/pdf/10.1023/A:1013999503812.pdf	Bartlett, Peter L., Boucheron, Stephane, Gábor Lugosi (2002) “Model Selection and Error Estimation”, Machine Learning (48), 85-113;

Equipment
Access to the Internet through a computer class and Wi-Fi network of the institute
Audience (till 40 persons)
Syllabus documents and materials on the topics of discipline
The library, including electronic publications

Software
R
Python
Matlab

Knowledge
Main notions and approaches of modern nonparametric statistics.
Main approaches to model choice and model validation.
Modelling methods for complex data.

Skill
Understand the applicability and cross relations between parametric and nonparametric statistical methods.

Experience
Be able to solve practical problems of data modeling and analysis.
Be able to compute statistical estimators using the quasi maximum likelihood method.
Be able to select the best model from the given collection.
Be able to make an oral and written presentation.
Be able to work with research literature on the modern statistical theory.

Select Assignment 1 Type	Homework Assignments
Input Example(s) of Assignment 1 (preferable)	Maximum likelihood in linear regression. You need to answer the question in a written form. Exercise 1. Derive the decomposition for the prediction loss and risk in linear regression Exercise 2. Compute a probabilistic upper and lower bound for the quadratic loss using the bound for quadratic forms
Assessment Criteria for Assignment 1	Each home assignment will have 3-10 problems. Each problem will cost some points according to its difficulty. Problems will be graded in the usual manner: student will receive full or partial credit out of maximum points. Home assignment should be done in a written form. Late home assignments will generally not be accepted. Exceptions will be made only for exceptional extenuating circumstances. Evaluation Criteria for the Home assignment. Final grade will be calculated as the sum of the points student obtained divided by the maximum sum of the points.
Select Assignment 2 Type	Homework Assignments
Input Example(s) of Assignment 2 (preferable)	Regularization and roughness penalty You need to answer the question in a written form. Exercise 1. Derive the decomposition for the prediction loss and risk in linear regression with roughness penalty. Show that the bias grows with penalty, while the variance decreases. Exercise 2. Compare the RSS (residual sum of squares) with the penalized maximum likelihood
Assessment Criteria for Assignment 2	Each home assignment will have 3-10 problems. Each problem will cost some points according to its difficulty. Problems will be graded in the usual manner: student will receive full or partial credit out of maximum points.
Select Assignment 3 Type	Homework Assignments
Input Example(s) of Assignment 3 (preferable)	Model selection by SURE and AIC, Cross validation. Exercise 1. Consider a problem of linear regression. Show that the statistics used in AIC are unbiased risk estimators.
Assessment Criteria for Assignment 3	Each home assignment will have 3-10 problems. Each problem will cost some points according to its difficulty. Problems will be graded in the usual manner: student will receive full or partial credit out of maximum points.
Select Assignment 4 Type	Homework Assignments
Input Example(s) of Assignment 4 (preferable)	Model section by smallest accepted. Exercise 1. Consider the smallest accepted method in a problem of linear regression. Derive an upper bound on the probability of rejection of the true model m*.
Assessment Criteria for Assignment 4	Each home assignment will have 3-10 problems. Each problem will cost some points according to its difficulty. Problems will be graded in the usual manner: student will receive full or partial credit out of maximum points.

Select Assignment 5 Type	Homework Assignments
Input Example(s) of Assignment 5	Wavelet methods: hard and soft thresholding. Exercise 1. Consider a problem of linear regression with orthonormal design. Show that the l_0-norm penalized maximum likelihood estimate is given by hard thresholding.
Assessment Criteria for Assignment 5	Each home assignment will have 3-10 problems. Each problem will cost some points according to its difficulty. Problems will be graded in the usual manner: student will receive full or partial credit out of maximum points.
Free Style Comments (if any)	Students must upload their solutions of homework exercise in PDF format. The usage of LaTeX is encouraged but not obligatory.