Course THE BASICS OF MACHINE LEARNING

Basic Information

This is a course, which contributes to MSc award in Mathematics

Duration	**216 hours (6 ECTS)**
Starting date	September, 1st
Study credits	6 ECTS credits
Language of instruction	English level B1(European Framework of Reference of Communicative Skills)
Academic requirements	– BSc degree in Mathematics, Physics, Computer Science, Engineering or equivalent (a copy of your diplomas from previous university studies and transcripts of completes courses and grades), – Skype interview

Course Description

«The Basics of Machine Learning» is the first course on data analysis, which enables students to master the fundamentals of data visualization, building models and determination of analysis validity.

The most important part of this course is studying the analysis design, which covers all stages from stating problems to the interpretation of results. The scheme of data research will remain almost the same during all the courses of data analysis students do further, therefore this course provides the base for future implementation of advanced instruments in students' projects.

Special Features of the Course

The course focuses on mathematical basis of the studied methods, which enables students to understand relationships between methods, applicability and limitations of different instruments, and develops the depth of knowledge necessary for data scientist.

All programming is assumed to be done in R programming language, which was chosen due to its specialization on data science needs and free availability under the GNU General Public License. Although the course does not provide classes devoted to the syntax of this language, students familiar to any programming language will cope with coding.

Course Aim

- To introduce basic data analysis problems and models.
- To develop skills of setting data analysis tasks, choosing appropriate methods and validation the models built.
- To provide students with knowledge of current technical implementations of some analytical methods.

Course Objectives

- To give students a notion of models and variable types, problems of supervised and unsupervised learning;
- To provide students with detailed knowledge of linear models of predictive analysis;
- To ensure student are familiar to mathematical base for model construction and validating;
- To equip student with abilities of estimation model quality;
- To develop practical skills in programming machine learning algorithms in the R language.

Learning Outcomes of the Course

By the end of the course, students will be able to:

- formulate the problem of analysis for given datasets and goals of research;
- prepare data for analysis;
- choose proper analytical methods and metrics for quality estimation;
- program chosen algorithms using appropriate instruments;
- identify signs and causes of model quality problems;
- interpret the results of model implementation.

Course (module) Structure

Learning Activities	Hours
Lectures	36
Practice sessions / Seminars,	36
Self-study Assignments	108
Final Exam (including preparation)	36
Total study hours	**216**

Course Outline

Week	Lectures	Practice session / Assignments	Hours[1]
1-2	Basic components of machine learning	– Basics of programming using R – Descriptive statistics and visual data analysis – Lab 1 – Lab 2	10
3-5	Linear methods of regression	– Multiple linear regression with Least square method – Quantile regression. – Metrics for regression – Inference for a regression model	24

[1] Hours designed for Classroom sessions, Web-sessions, Home Assignments etc.

		– Lab 3 – Lab 4	
6-10	Linear methods of classification	– Classification task. Classification performance metrics – Linear discriminant analysis – Logistic Regression – Support vector classifier – Support vector machine – Lab 5 – Lab 6	34
11-13	Linear Models and Experimental Design	– Design of experiments – Analysis of variance – Analysis of covariance – Lab 7	12
14-18	Models quality issues	– The problem of multicollinearity – Subset selection – Shrinkage methods: Ridge regression and the Lasso – Principle components regression – Influential points – Lab 8 – Lab 9	24
18	Final exam		4

Assessment

Grade policy for both practical home assignments and the final exam is:

- A (excellent work) 91–100 points
- B (above average work) 81–90 points
- C (average work) 71–80 points
- D (below average work) 50–70 points
- F (failed work) < 50 points

The exam is taken orally. For admission to the final exam, the student must complete and defend at least 8 laboratory works. The exam is an individual problem in the form of a simulated experimental dataset.

The final grade includes grade for homework and exam in a ratio of 50:50.

Attendance Policy

Students are expected to attend classes regularly. However, occasional skipping classes is permissible if a student does all necessary in-class work at home.

Lecturer(s) and Tutors, Contact Information

Tatiana KUSTITSKAYA
Ph.D., Associated Professor at School of Space and Information Technologies, Siberian Federal University
(room 311) 26-1, Kirensky st, Krasnoyarsk, Russia
Tel: +7 391 291 2790,
tkustitskaya@sfu-kras.ru, http://structure.sfu-kras.ru/node/4321