
Machine Learning and Applications

Course Outline, Activities and Grading

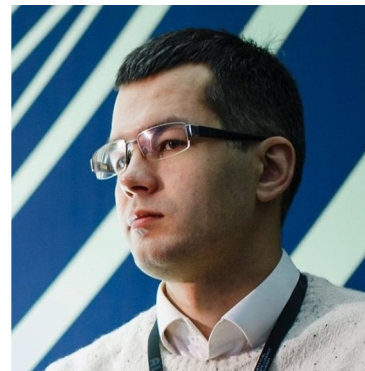
Course Organizers

Course instructor



prof. Evgeny Burnaev

Co-instructor

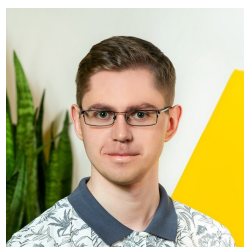


prof. Alexey Zaytsev

Teaching Assistants:



Maria
Ivanova



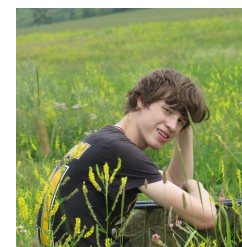
Nikita
Gushchin



Milena
Gazdieva



Daniil
Selikhanovych



Petr
Mokrov

Course Topics

Part 1 (9 lectures & seminars):

- General Introduction (1)
- Regression, Kernel Trick (1)
- Classification (1)
- Support Vector Machines (1)
- Tree-based Methods (1)
- Adaboost (1)
- Gradient boosting (1)
- Advanced Classification (1)
- Model and Feature Selection (1)

Part 2 (7 lectures & seminars):

- Neural Networks (2)
- Sequential data (1)
- Gaussian Processes (1)
- Dimensionality Reduction (1)
- Anomaly Detection (1)
- Clustering (1)

Course Textbooks

Main

1. Hastie, T., and Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 12 print, Springer, 2009
2. Tibshirani, R. and Hastie, T., *An Introduction to Statistical Learning*, Springer 2013
3. Bishop, C.M. *Pattern Recognition and Machine Learning*. Springer, 2007
4. Barber, D. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012

Additional

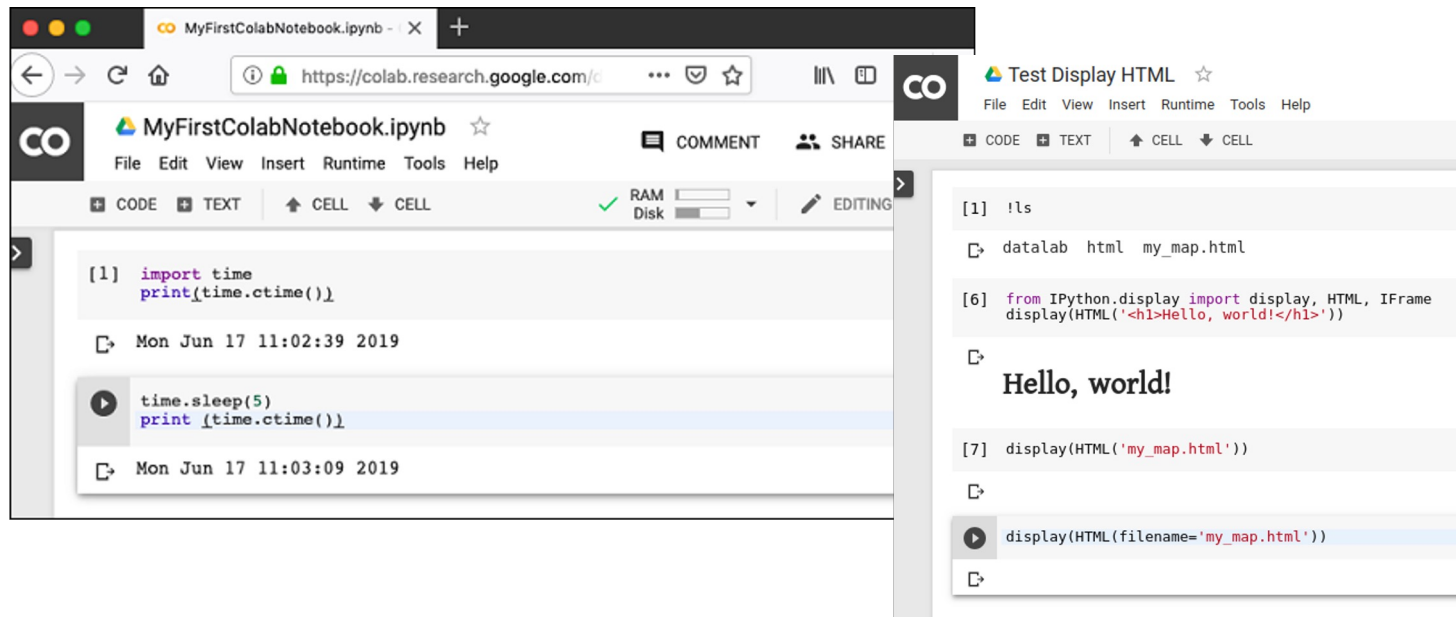
1. Rasmussen, C., and Williams, C. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
2. Mohri, M., and Rostamizadeh, A., and Talwalkar, A. *Foundations of Machine Learning*. MIT, 2012
3. Schapire, R.E., Freund, Y. *Boosting*. MIT, 2012
4. Clarke, B., and Fokoue, E., and Zhang, H.H. *Principles and Theory for Data Mining and Machine Learning*. Springer, 2009

Course Prerequisites

- Adequate understanding of **Calculus** as well as
 - **Probability Theory** and **Statistics**
 - **Linear Algebra** (applied and theoretical)
 - **Optimization Methods**
- Adequate **python programming skills**
 - basic familiarity with **numpy** and **scipy**
- Basic knowledge of **algorithms** and **complexity**

Course Software Requirements

- **Obligatory set up** [seminars + homeworks]
 - Google **colab** (colab.research.google.com)



Quick start guide:

https://www.tutorialspoint.com/google_colab/your_first_colab_notebook.htm

Course Assistance and Consultation

- Should you need ...
 - **consultation** on your projects
 - **advice** on solving tough problems in the homework
- ... **we encourage you to ask** the instructor or the assistants
 - online after the end of a seminar or a lecture
 - through course group chat in **Telegram**: t.me/ml2023skoltech
 - through discussions in **Canvas**

communication through private messages (in *telegram, vk, etc.*) is **not welcome** unless agreed upon by both parties

Course Final Score and Grade

- The final score is computed based on activity scores thus

$$\text{Total} = \mathbf{S_hw} \times 0.4 + \mathbf{S_exam} \times 0.25 + \mathbf{S_proj} \times 0.35$$

- **S_hw** is the **sum** of %score for each of the 2 homework assignments
- **S_exam** – the %score of the final exam
- **S_proj** -- the %score of the final project

Final Grade	Total Score
A “Excellent”	86% and above
B “Good”	< 86% to 76%
C “Satisfactory”	< 76% to 66%
D “Poor”	< 66% to 56%
E “Very poor”	< 56% to 46%
F “Unacceptable”	< 46%

Activity	Total weight
Home assignments (2)	40%
Final exam	25%
Final project	35%
Total	100%

Course Activity

- **Out-of-class self study** is very important
 - do homeworks, study the material and reflect on it
 - don't be afraid to ask questions
- **Workload is substantial**
 - 2 assignments, 1 exam, 1 team project
- **Zero-tolerance** policy on plagiarism and dishonesty
 - The assignments and exam are individual. **Any detected plagiarism in will result in an immediate exclusion of the student from the course (with an F grade). There will be no excuses accepted.** Plagiarism includes copying solutions from your peers of this year, participants of previous year classes, explicit copying of etc. Also, references to external sources as solutions (book chapters, papers, websites, etc.) will not be considered and accepted.

Course Activity: Home Assignments

Two individual assignments (week 1 - week 6):

- each assignment has practical (coding) tasks
- published in the middle of week 1 and week 3 and stays open for **~3 weeks**
- ipython notebook (colab) + Latex markdowns (within)
- Assignments costs in sum **40% of the final grade**

Rules

- **hard deadline** assignments (**Plan your work ahead!!!**)
- only **the most recent submission** is graded
 - **no submission** means **zero grade** (**no excuses!**)

Course Activity: Final Exam

- **Exam**

- 1 hour written exam will be held on week 7
- Several theoretical questions, mostly multiple choice, based on the lecture materials
- Exam costs **25% of the final grade**

- **Rules**

- Exam is an **individual** assignment
- There will be **no way** to rewrite a missed exam

Course Activity: the Final Project

- **35% of the final grade**
- **teams** of 3 -- 5 students
- may be combined with **currently running parallel or already taken** courses
 - ***must be explicitly disclosed, failing to do so is plagiarism***
- A comprehensive test of
 - teamwork organization and research engineering
 - knowledge of ML, insight, validation and evaluation
 - research presentation and communication skills
- The project timeline:
 - Week 3 -- we release suggested project topics
 - Week 5 -- Submission of project proposals (**Hard Deadline**)
 - Week 5 -- Feedback and approval of the projects
 - Weeks 7-8 -- Project Consultations
 - Week 8 -- Reviews of Projects
 - Week 8 – Presentation, Repo and Report Submission

The Final Project: Topics

- Final Project types
 - **Applied:** pick an interesting application and figure out how to apply machine learning algorithms to solve it
 - **Algorithmic:** propose a new learning algorithm, or a variant of some existing one to solve a general problem or group thereof
 - **Replication study:** pick a fresh preprint or an accepted conference paper, replicate its results and discuss the outcomes

The Final Project: Format, Structure and Grading

- Project in a **github repo + PDF report**
 - ICML 2020 template has to be used for the report
 - Report costs 25% out of 35% of the final grade
- Project **in-class presentation + PDF presentation**
 - 10-15 mins per group in-class presentation on week 8
 - Presentation costs 10% out of 35% of the final grade
- Concise report with **4-6 pages** (excl. appendices)
 - **Introduction**, motivation and problem statement
 - **Related work** and brief literature overview
 - **Dataset Description**
 - **ML Methods** and algorithms, proposed algorithm modifications, etc.
 - **Experiments / Discussion**: details about (hyper) parameters and how you picked them, cross-validation metrics and details, discussion of failures and successes, equations, results, visualizations, tables, etc.
 - **Conclusion** and directions for further research
 - **References**, acknowledgements and **contributions of each team member**

The Final Project: Evaluation

- structure and clarity of the project repository
 - **reproducible** and well defined **ML pipeline**: *data acquisition, processing, modelling, validation, and report generation*
- the quality and relevance of the PDF report
 - **relevance and novelty**: *toy/real problem or common/unexplored method*
 - **technical quality**: *insightful choice of clever reasonable methods, cross-validation and general assessment of the tools/methods used*
 - literacy, quality of figures/tables and general narrative **structure**
- the project presentation (pdf + in-class presentation)
 - science **communication** skills, presentation **quality and clarity**
 - **relevant** content and summary, **knowledge** demonstrated by the team

Reminder: Student Academic Integrity

Disciplinary penalties are imposed for

- **cheating, plagiarism**, fabrication or falsification of data or results
- **copying**, rewriting, paraphrasing, or summarizing of text, discoveries, or insights without **acknowledging and / or citing the source**;
- **allowing other students to copy** one's own work, **using another student's** solutions or code

Penalties include, but are not limited to

- **getting no grade** for the project, assignment, or exam
- **redoing** an assignment or test **for a significantly reduced grade**

If you have any question, please, refer to

"Student Academic Integrity Regulations". Department of Education, Skoltech. Moscow, 2014

end of this presentation

