Machine Learning and Applications

Course Outline, Activities and Grading

Skoltech
Skolkovo Institute of Science and Technology

Course Topics

Part 1 (before the midterm):

- Regression, Kernel Trick (1)
- Classification (1)
- Support Vector Machines (1)
- Tree-based Methods (1)
- Advanced Classification (1)
- Model and Feature Selection (1)
- Bagging & Boosting (2)
- Kernel Methods (1)

Part 2 (after the midterm):

- Bayesian ML (1)
- Gaussian Processes (1)
- Neural Networks (2)
- Dimensionality Reduction (1)
- Anomaly Detection (1)
- Clustering (1)
- Active Learning (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., Friend, 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

Skoltech
Skolkovo Institute of Science and Technology

Course Prerequisites

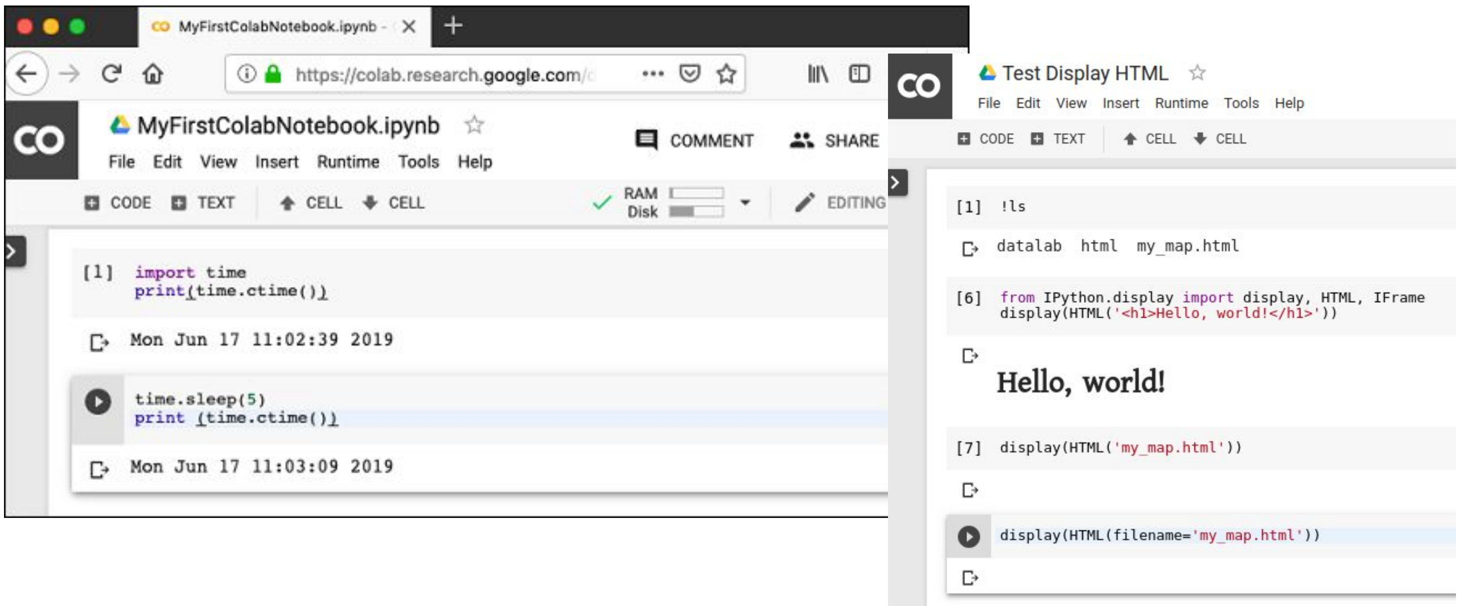
- Adequate understanding of **Calculus** as well as
  - **Probability Theory** and **Statistics**
  - **Linear Algebra (applied and theoretical)**
  - **Optimization Methods**

- Adequate **python programming skills**
  - basic familiarity with **numpy** and **scipy**

- Basic knowledge of **algorithms** and **complexity**

Course Software Requirements

- **<u>Obligatory</u> 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
  - to improve your grade with extra credit assignments

- … **we encourage you to ask** the instructor or the assistants
  - in person after the end of a seminar or lecture
  - through direct messaging or discussion in Canvas
    - communication through other means, e.g. *telegram, whatsapp, vk, fb, or twitter*, is **not welcome** unless agreed upon by both parties

Course Final Score and Grade

- The final score is computed based on activity scores thus

Total = (**S_xtr** x 0.05 + **S_hw** x 0.10)+ **S_mid** x 0.15
+ **S_fin** x 0.20 + **S_prj** x 0.35

  - **S_hw** is the **sum** of %score for each of the 3 core assignments
  - **S_xtr** is the %score of an extra credit assignment (optional)
  - **S_mid, S_fin** -- the %score of the midterm and the final
  - **S_prj** is the %score of the course 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	Weight
Home assignments	30%
Midterm exam	15%
Project	35%
Final exam	20%

Skoltech
Skolkovo Institute of Science and Technology

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**
  - 3 (+1*) assignments, 2 exams and 1 project
- **Zero-tolerance** policy on plagiarism and dishonesty
  - The assignments & exams 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

Three individual assignments (week 1 - week 6):

- each assignment has both theoretical problems and practical tasks
- published in the middle of odd week and stays open for **12 days**
- ipython notebook (colab) + Latex markdowns (within)

Rules

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

Skoltech
Skolkovo Institute of Science and Technology

Course Activity: Exams

- **exams**
  - binary and multiple choice question, theoretical problems
  - **electronic devices and communication are prohibited**
  - one A4 page (two-sided) condensed cheat-sheets are ok

- **the Midterm Exam**
  - examines the topics covered during weeks one, two, and three
  - 75 minutes long on a seminar during week 4

- **the Final Exam**
  - tests the topics covered during the entire course
  - 3 hours long on the 7-th week

Course Activity: the Final Project

- **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 4 -- Submission of project proposals (**Hard Deadline**)
    - Week 5 -- Feedback and approval of the projects
    - Week 7 -- Project Consultations
    - Week 8 -- Project Defence and Final 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

Skoltech
Skolkovo Institute of Science and Technology

The Final Project: Format and Structure

- Project in a **github repo** with the source and the **PDF report**
  - ICML 2020 template has to be used for the report

- Concise report with up to **6 pages** (incl. figures, tables, 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**

Skoltech
Skolkovo Institute of Science and Technology

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 defence
  - science **communication** skills, presentation **quality and clarity**
  - **relevant** content and summary, **knowledge** demonstrated by the team

Skoltech
Skolkovo Institute of Science and Technology

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