Machine Learning and Applications

Course Outline

Prerequisites

The course requires familiarity with

- **Calculus and Numerical Linear Algebra**
- **Optimization Methods**
- **Probability** and **Statistics**

Recommended software: Anaconda 4.2.0 (Python 3.5 version)

- python>=3.5, matplotlib>=1.5.3, jupyter, ipython;
- scikit-learn>=0.17.1, pandas>=0.18.1, numpy>=1.11, scipy>=0.18.1;

Extra: XGboost (Installation), LibSVM>=3.2 (Installation)

Course Topics

The course will cover:

- Linear classification
- Multi-Class Classification
- On-Line Learning and Ensemble Methods
- Decision Forests for Classification, Regression, and Density Estimation
- Regression
- Neural Networks
- Reinforcement Learning
- Kernel Methods
- Kernel Mean Embedding of Distributions
- Dimensionality Reduction
- Manifold and Semi-Supervised Learning
- Metric Learning
- Anomaly Detection
- Conformal and Probabilistic Prediction

Course Activity and Grading

- Weekly individual home assignments
- A report on a selected original research paper
- Applied course project (individual or in a small group)
  - detailed report with replicable results
  - presentation and defense

Activity	Max. Score
Home assignments	25%
Attendance and participation	5% (+ bonus)
Written Exam	25%
Project	25%
Oral Exam	25%

Final Grade	Total Score
A "Excellent"	above 80%
B "Good"	from 65 to 79%
C "Satisfactory"	from 50% to 64%
D "Poor"	from 30% to 49%
E "Very poor"	from 15% to 29%
F "Unacceptable"	below 14%

Assignment Deadlines

- Weekly assignments shall be distributed via Canvas
- Solutions are to be submitted in PDF, IPYNB, or ZIP
- We adopt soft deadlines

Example: if the deadline is set to **2017-02-19**, then

- Assignments submitted before **2017-02-20 08:59:59 MSK** are not penalized
- Late submissions are penalized **per each day** after the deadline **rounded up**

Projects

Reports are to be 15-20 pages long and include:

1. Problem Statement
2. Overview of the state in this field
3. Solution
4. Comparison to existing methods
5. Conclusion and References

Project presentation with 7-10 slides summarizing the results of the project

Assessment criteria:

- 10% -- general literacy and style of the report
- 20% -- analytical/scientific methods and approaches
- 45% -- depth of the subject understanding
- 25% -- presentation style and Q&A

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 or citing the source**;
- **allowing other students to copy** one's own work, **using another student's solutions or code** without specifying this.

Penalties for misconduct include:

- asking to **redo** the assignment, project, or test **for a reduced grade**;
- **getting a failing grade**.

Please refer to

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

Skoltech
Skolkovo Institute of Science and Technology

Main Course Materials

1. Mohri, M., and Rostamizadeh, A., and Talwalkar, A. *Foundations of Machine Learning*. MIT, 2012
2. Shalev-Shwartz, S., and Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge, 2014
3. Bishop, C.M. *Pattern Recognition and Machine Learning*. Springer, 2007
4. Barber, D. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012
5. Rasmussen, C., and Williams, C. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
6. Hastie, T., and Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, 2009
7. Schapire, R.E., Friend, Y. *Boosting*. MIT, 2012
8. Clarke, B., and Fokoue, E., and Zhang, H.H. *Principles and Theory for Data Mining and Machine Learning*. Springer, 2009