**Machine Learning**

**Syllabus**

Skolkovo Institute of Science and Technology

November-December 2015.

**Instructor:** Victor Kitov

+7-916-343-6903

[v.v.kitov@yandex.ru](mailto:v.v.kitov@yandex.ru)

**Course Description**

The course gives a general introduction to machine learning – a scientific field about algorithms that automatically learn from data, extract information from data and make decisions. All major steps of machine learning application pipeline will be covered, including data preprocessing, features selection and extraction, algorithm selection and tuning and performance assessment. The course will develop both theoretical and practical skills in machine learning. Major algorithms of regression, classification, clustering, feature selection, dimensionality reduction and ensemble learning subfields will be studied, their theoretical properties analyzed and algorithms will be qualitatively compared between each other. At the same time much emphasis will be placed on applying machine learning algorithms to real-world problems using python scientific infrastructure, including numpy, scipy, pandas, matplotlib and scikit-learn libraries.

**Prerequisites**

Linear algebra, mathematical analysis, basic probability and statistics, basic optimization (with and without constraints) and general programming skills.

**Recommended materials**

* Statistical Pattern Recognition, 3rd Edition, Andrew R. Webb, Keith D. Copsey, John Wiley & Sons Ltd., 2011.

<http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470682272.html>

* Машинное обучение. Курс лекций. К.В.Воронцов. (Russian). Available using search on [www.machinelearning.ru](http://www.machinelearning.ru)
* The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman, 2nd Edition, Springer, 2009.

<http://statweb.stanford.edu/~tibs/ElemStatLearn/>

* Pattern Recognition and Machine Learning. Christopher M.Bishop. Springer. 2006.
* Machine Learning: A Probabilistic Perspective. Kevin P.Murphy. Massachusetts Institute of Technology. 2012.

**Grading**

Grading is based on combined student activity throughout the course with the following proportions:

attendance contributes 10%,

home assignments contribute 30%,

exam contributes 30%,

applied project contributes 30%

of the grade.

Additional points for successful participation in machine learning competitions on [www.kaggle.com](http://www.kaggle.com) may be given.

Exam will be written and confirm general knowledge and understanding of all material covered on lectures and seminars.

On applied project students will need to apply theoretical knowledge obtained on lectures and seminars to practical real-world prediction problem. Using a common training dataset, students will need to develop their own prediction algorithm with the help of software tools and libraries of their own choice. The quality of the project will be assessed using a common prediction quality measure and will be graded according to the relative rank of this quality measure in the competition between students. The grade will also depend on the quality of presentation of the developed algorithm and of the breadth of analysis.

Homework will consist of theoretical tasks, confirming understanding of the material, and practical tasks, developing skills of applying python scientific tools for data analysis, exploration and prediction.

**Attendance**

Attendance of lectures and seminars is optional but will contribute to the final mark for the course. It is strongly recommended to attend all classes since specific knowledge and skills apart from general information from textbooks will be given.