

Machine Learning and Data Mining

Course syllabus

Maxim Borisyak
National Research University Higher School of Economics (HSE)

Syllabus

Three parts:

- ❖ traditional learning;
- ❖ deep learning;
- ❖ 'big' learning.

Traditional learning

Meta learning:

- ❖ No-Free-Lunch, bias-variance decomposition, regularization;
- ❖ bagging: Random Forest, Extra Trees; stacking: calibration;
- ❖ boosting: AdaBoost, Gradient Boosting;
- ❖ partical session: boosting for computer vision;
- ❖ **homework**: Viola-Jones cascades.

Traditional learning

Algorithm-invariant topics:

- ❖ optimization: gradient and gradient-free methods, global optimization;
- ❖ imbalanced datasets, reweighting, importance sampling;
- ❖ one-class and semi-supervised learning.

Deep Learning

I highly recommend to attend 'Deep Learning' course by Alexander Panin. In order to avoid overlaps, this section is more-or-less complementary to that course.

Deep Learning:

- ❖ Deep Learning and No-Free-Lunch;
- ❖ regularization, pretraining and other tricks;
- ❖ autoencoders;
- ❖ energy-based learning;
- ❖ generative models: Restricted Boltzman Machine;
- ❖ generative models: Generative Adversarial Networks;
- ❖ **homework**: a number of exercises.

Big Data

Big data:

- ❖ introduction to distributed computations with Spark;
- ❖ distributed Machine Learning algorithms;
- ❖ **homework**: distributed logistic regression.

Requirements

For the first part, please, make sure you have installed:

- ❖ python 3;
- ❖ Jupyter Notebook, matplotlib;
- ❖ numpy, scipy, scikit-learn;
- ❖ (optionally) XGBoost;

For the Deep Learning part, please, additionally install:

- ❖ theano;
- ❖ lasagne.

Recommended literature

- ❖ Bishop, C.M., 2006. Pattern recognition and machine learning. springer.
- ❖ Friedman, J., Hastie, T. and Tibshirani, R., 2001. The elements of statistical learning (Vol. 1, pp. 241-249). New York: Springer series in statistics.
- ❖ Bishop, C.M., 1995. Neural networks for pattern recognition. Oxford university press.
- ❖ Jiwei, H. and Kamber, P., 2012. Data Mining concepts and techniques third edition.
- ❖ Wills, J., Owen, S., Laserson, U. and Ryza, S., 2015. Advanced Analytics with Spark: Patterns for Learning from Data at Scale.

Additional materials will be listed on lecture slides.