Advanced modelling for Big Data and AI
Instructor: András Lukács

Term: Spring

Weeks: 8-14

Contact hours: 3

Credits: 6

Aim and scope:

Understanding mathematical theory of learning; knowledge of models and algorithmic techniques for analysing data of high volume and large complexity; knowledge of methods for special type of data and domain of problems. Mathematical and statistical foundation of data mining, learning and models. Practices cover the weekly topics in the form of Python notebooks. Trying methods rather than programming.

Syllabus:
Mathematical foundation of learning: learning models, PAC learning No-Free-Lunch theorem, VC-dimension.
High-dimensional data, similarity and distance measures, embeddings. Nearest neighbour problem, minhashing. locality-sensitive hashing, sketches, Johnson-Lindenstrauss theorem. Methods for data streams: sampling, Bloom filter, counting distinct elements, estimating moments.
Large-scale machine learning: map-reduce model, linear algebraic and database tasks with map-reduce. Reinforcement learning, Q-learning.
Models for time-series. Dynamic Time Warping. Symbolic Aggregate Approximation. Graphical models, Markov chains
Link analysis in networks. PageRank, Hyperlink-Induced Topic Search.
Recommendation systems, utility matrix, profiles, collaborative filtering, UV-decomposition.

Grading: exam (incorporating the solution of homeworks)

Literature:

S. Shalev-Shwartz, S. Ben-David: Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014
C. M. Bishop: Pattern Recognition and Machine Learning, Springer, 2006
A. Rajaraman, J. Leskovec, J. D. Ullman: Mining of Massive Datasets, Cambridge Univ. Press, 2014