

Advanced machine learning

András Kornai

BMETE91MM20

While BMETE95AM36 is formally not a prerequisite, this course builds on material presented there, and students who passed that course will be able to answer the first two exam questions without further preparation.

- 1 Main tasks: classification, regression, generation. Name at least one algorithm for each task in some domain of your choice, explain its mathematical foundations, and demonstrate ability to solve toy problems by using open source software implementing the algorithm.
- 2 Main problem domains: speech- and character recognition (ASR, OCR), (biometric) identification, pattern classification, ranking/recommendation, info extraction, info retrieval, natural language processing (NLP). Name at least one standard data repository for each domain, and demonstrate ability to solve problems by training algorithms on standard data.
- 3 Basics of descriptive statistics, linear algebra, optimization, information theory. Data reduction, principal component analysis, linear discriminant analysis, feature engineering.
- 4 Survey of major machine learners: linear classifiers, maximum entropy, hidden Markov (HMM), nearest neighbor, max margin, genetic/evolutionary, boost, decision tree, Bayesian, neural net (NN).
- 5 Central NLP tasks: sequence tagging (POS, NER), chunking, parsing, anaphora resolution, disambiguation, language identification, role labeling, semantic similarity, paraphrase, dictionary building, machine translation.
- 6 Learning finite automata, transducers, Eilenberg machines
- 7 Algorithmic information theory, Kolmogorov complexity, minimum description length.
- 8 Unsupervised learning, structure learning