Machine Learning

András Kornai

Plan of the course

Week 1 What are ML algorithms, how did they come about, where are they today, what to expect in 5-10 years. Computational foundations, societal impact.

Week 2 Data collection, standard data sets, repositories. Survey of major application domains: speech- and character recognition (ASR, OCR), (biometric) identification, patern classification, ranking/recommendation, info extraction, info retrieval, natural language processing (NLP).

Week 3 Basics of descriptive statistics, linear algebra, optimization, information theory, and why we care about these. Student project descriptions due.

Week 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). Midterm exam.

Plan of the course cont'd

Week 5 Data reduction, principal component analysis, linear discriminant analysis, feature engineering.

Week 6 Algorithmic information theory, Kolmogorov complexity, minimum description length.

Week 7 Public presentations of student projects, feedback.

Week 8 Final results of student projects, wrap up.

Week 1: Computational foundations

- deterministic, nondeterministic, probabilistic computing
- The old world: deterministic algorithms
- Early nondeterminism, teleology
- Probabilistic models, random variables, sampling
- Mean, median, quantiles
- Determinism and free will

Readings: Rabin and Scott (1959) Finite Automata and Their Decision Problems; Floyd (1967) Nondeterministic algorithms; Kornai (1996) Comments on Mohri, Pereira and Riley

Week 1: Societal impact

- Relationship of ML to NLP: 'rationalists' vs. 'empiricists'
- Relationship of ML to robotics
- Jobs, what were they?
- Individuals, organizations, institutions, societies

(Note: midterm will only include uncontroversial material, in particular finite automata and transducers, both deterministic and nondeterministic, regular expressions, grep/egrep.)

Week 1: a near-trivial ML algorithm

- Unsupervised and supervised learning
- Fisher Iris data
- Train and test