Algorithmic Foundations of Learning

Course level: Part C / MSc course.
Instructor: Patrick Rebeschini.
Term: Michaelmas.

Prerequisites

The course requires a good level of mathematical maturity. Students are expected to be familiar with core concepts in probability (properties of conditional expectations, basic inequalities such as Markov's, Chebyshev's and Cauchy-Schwarz's), statistics (confidence intervals, bias-variance tradeoff), and linear algebra (matrix-vector operations, eigenvalues and eigenvectors). Previous exposure to machine learning (empirical risk minimisation, overfitting, regularisation) is highly recommended.

Students would benefit from being familiar with the material covered in the following courses offered in the Statistics department: SB2a Foundations of Statistical Inference (in particular, Decision Theory) and in SB2b Statistical Machine Learning.

Syllabus

The course is meant to provide a rigorous theoretical account of the main ideas underlying machine learning, and to offer a principled framework to understand the algorithmic paradigms being used, with an emphasis on first-order stochastic optimisation methods.

- Statistical learning framework. Empirical risk minimisation. Error decomposition (generalisation, optimisation, approximation). Bias-complexity tradeoff.
- Generalisation error:
  - Learning via uniform convergence.
  - Concentration inequalities (Azuma-Hoeffding, McDiarmid, Bernstein).
  - Binary classification. Bayes classifier. Slow and fast rate.
  - Elements of VC theory.
  - Rademacher complexity.

  - Covering numbers. Chaining.
  - Convex loss surrogates.
  - Learning via margin bounds.
  - Learning via algorithmic stability.

- Optimisation error:
  - Elements of convex theory (subgradients, projections, Lipschitz, smoothness, strong convexity).
  - Oracle model. Gradient descent. Mirror descent.
  - Stochastic oracle model. Stochastic gradient descent. Stochastic mirror descent. Single pass, multiple pass.

- Examples of error trade-off. Linear predictors, including Boosting. Non-linear predictors, including Support Vector Machines and Neural Networks.

- Structural regularisation (constraints, penalisation). Implicit/algorithmic regularization (early stopping).

- Teacher-student learning framework. Approximate Message Passing.

- Online learning framework. Multi-armed bandits.

Reading Material

- Shai Shalev-Shwartz and Shai Ben-David. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014.

- Ramon van Handel. Probability in High Dimension. Lecture notes available online (http://www.princeton.edu/ rvan/APC550.pdf), 2016.

- Sébastien Bubeck. Convex Optimization: Algorithms and Complexity. Foundations and Trends in Machine Learning, 2015.

- Sébastien Bubeck. Introduction to Online Optimization. Lecture notes available online (http://sbubeck.com/BubeckLectureNotes.pdf), 2011.