EE-556

Mathematics of data: from theory to computation

Cevher Volkan

Cursus	Sem.	Type
Computational science and Engineering	MA1, MA3	Opt.
Computational science and engineering minor	H	Opt.
Data Science	MA1, MA3	Opt.
Data science minor	H	Opt.
Electrical Engineering		Opt.
Electrical and Electronical Engineering	MA1, MA3	Opt.
Electrical and electronic engineering minor	H	Opt.
MNIS	MA3	Opt.
Managmt, tech et entr.	MA1, MA3	Opt.
Minor in statistics	H	Opt.
Neuro-X minor	H	Opt.
Neuro-X	MA1, MA3	Opt.
Quantum Science and Engineering	MA1, MA3	Opt.
Robotics	MA1, MA3	Opt.
Statistics	MA1, MA3	Opt.

Language of teaching	English
Credits	6
Session	Winter
Semester	Fall
Exam	Written
Workload	180h
Weeks	14
Hours	**6 weekly**
Courses	3 weekly
Project	3 weekly
Number of positions	

Summary

This course provides an overview of key advances in continuous optimization and statistical analysis for machine learning. We review recent learning formulations and models as well as their guarantees, describe scalable solution techniques and algorithms, and illustrate the trade-offs involved.

Content

The course consists of the following lectures (2h each)
Lecture 1: Introduction. The role of models and data. Maximum-likelihood formulations. Error decomposition for estimation and prediction.
Lecture 2: Generalized linear models. Logistic regression.
Lecture 3: Linear algebra reminders. Computation of Gradients. Reading convergence plots.
Lecture 4: The role of computation. Challenges to optimization algorithms. Optimality measures. Structures in optimization. Gradient descent. Convergence rate of gradient descent for smooth functions.
Lecture 5: Optimality of convergence rates. Lower bounds. Accelerated gradient descent. Concept of total complexity. Stochastic gradient descent.
Lecture 6: Concise signal models. Compressive sensing. Sample complexity bounds for estimation and prediction. Challenges to optimization algorithms for non-smooth optimization. Subgradient method.
Lecture 7: Introduction to proximal-operators. Proximal gradient methods. Linear minimization oracles. Conditional gradient method for constrained optimization.
Lecture 8: Time-data trade-offs. Variance reduction for improving trade-offs.
Lecture 9: Introduction to deep learning. Generalization through uniform convergence bounds. Rademacher complexity.
Lecture 10: Double descent curves and over-parameterization. Implicit regularization. Generalization bounds using stability.
Lecture 11: Escaping saddle points. Adaptive gradient methods.
Lecture 12: Adversarial machine learning and generative adversarial networks (GANs). Wasserstein GAN. Difficulty of minimax optimization. Pitfalls of gradient descent-ascent approach.
Lecture 13: Primal-dual optimization-I: Fundamentals of minimax problems. Fenchel conjugates. Duality.
Lecture 14: Primal-dual optimization-II: Extra gradient method. Chambolle-Pock algorithm. Stochastic primal-dual methods.
Lecture 15: Primal-dual III: Lagrangian gradient methods. Lagrangian conditional gradient methods.

Keywords

Machine Learning. Signal Processing. Optimization. Statististical Analysis. Linear and non-linear models. Algorithms. Data and computational trade-offs.

Learning Prerequisites

Required courses

Previous coursework in calculus, linear algebra, and probability is required. Familiarity with optimization is useful.
Familiarity with python, and basic knowledge of one deep learning framework (Pytorch, TensorFlow, JAX) is needed.

Learning Outcomes

By the end of the course, the student must be able to:

- Choose an appropriate convex formulation for a data analytics problem at hand
- Estimate the underlying data size requirements for the correctness of its solution
- Implement an appropriate convex optimization algorithm based on the available computational platform
- Decide on a meaningful level of optimization accuracy for stopping the algorithm
- Characterize the time required for their algorithm to obtain a numerical solution with the chosen accuracy

Resources

Moodle Link

- https://go.epfl.ch/EE-556