Mathematics of Data: From Theory to Computation (Fall 2024)

Description:	Mathematical Optimization offers a unified framework for obtaining numerical solutions to data analytics problems, oftentimes with provable statistical guarantees of correctness at well-understood computational costs. This course provides an overview of recent advances in mathematical optimization and statistical analysis for machine learning. We review the emerging learning formulations and models as well as their guarantees, describe scalable solution techniques and algorithms, and illustrate the trade-offs involved.
Learning outcomes:	By the end of the course, the students are expected to understand the so-called time-data tradeoffs in data analytics. In particular, the students must be able to: 1. Choose an appropriate convex or non-convex formulation for a data analytics problem at hand; 2. Estimate the underlying data size requirements for the correctness of its solution; 3. Implement an appropriate optimization algorithm based on the available computational platform; 4. Decide on a meaningful level of optimization accuracy for stopping the algorithm; 5. Characterize the time required for their algorithm to obtain a numerical solution with the chosen accuracy.
Prerequisites:	Previous coursework in calculus, linear algebra, and probability is required. Familiarity with optimization is useful. Some familiarity with python, and basic knowledge of one deep learning framework (Pytorch, TensorFlow, JAX) is helpful.
Language:	English
Class Times:	Mondays 9:00-12:00 in CM 1 5, Fridays 16:00-19:00 in CM 1 1 (first three weeks) and remotely on Zoom (`https://go.epfl.ch/mod-zoom`).

Lab & office hours: Fridays (fourth week onward) 16:00-19:00 in CM 1 1 and, when necessary, remotely on Zoom (`https://go.epfl.ch/mod-zoom-lab`).

Instructor: Prof. Volkan Cevher, ELE 233, volkan.cevher@epfl.ch

Head TA: Yongtao Wu (yongtao.wu@epfl.ch)
Pedro Abranches (pedro.abranches@epfl.ch)

Credits: 6

Course Website: `https://go.epfl.ch/mad-moodle`

Resources: Reading resources will be provided during lectures. The recordings of the lectures will be available on Mediaspace: `https://go.epfl.ch/mediaspaceMoD` . It is possible to follow the course remotely.

Honor Code: The EPFL honor code applies to the course: `https://polylex.epfl.ch/wp-content/uploads/2019/01/2.3.1_ch_code_honor_en.pdf`.

Grading: The default grade after registering for the course is 1. The first homework is worth 2 points, second and third homeworks are 1 point each and the written exam is 1 point.

- Three 3-week homework exercises.
- Your answers to the homework questions should be submitted via the Moodle page before the due date. Late submissions are not allowed, and you will get zero point for the corresponding homework.
- Discussing concepts with other students is OK; however, each homework exercise should be attempted and completed individually.
- Copying and cheating on homework will not be tolerated. The first time results in zero point for the corresponding homework, and the second time results in zero point for the whole course.

Course Outline

Lecture 1: Introduction. The role of models and data. Maximum-likelihood formulations. Sample complexity bounds for estimation and prediction.

Lecture 2: Generalized linear models. Logistic regression.

Lecture 3: Linear algebra reminders. Gradients. Reading convergence plots.

Lecture 4: Optimization algorithms. Optimality measures. Structures in optimization. Gradient descent. Gradient descent for smooth functions.

Lecture 5: Optimality of convergence rates. Lower bounds. Accelerated gradient descent. Concept of total complexity. Adaptive methods. Tensor methods.

Lecture 6: Stochastic gradient descent. 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: Variance reduction. Introduction to deep learning. Challenges in deep learning theory and applications.

Lecture 9: Generalization through uniform convergence bounds. Rademacher complexity. Double descent curves and over-parameterization. Implicit regularization. Generalization bounds using stability.

Lecture 10: Escaping saddle points. Adaptive gradient methods.

Lecture 11: Adversarial machine learning and generative adversarial networks (GANs). Wasserstein GAN. Difficulty of minimax optimization.

Lecture 12: Robustness in deep learning. Diffusion models.

Lecture 13: Primal-dual optimization I: Fundamentals of minimax problems. Fenchel conjugates. Duality. Extra gradient method. Chambolle-Pock algorithm. Stochastic primal-dual methods.

Lecture 14: Primal-dual optimization II: Augmented Lagrangian gradient methods. Semi-definite programming. HCGM and CGAL algorithms.

Lecture 15: Language models: Basis of language models. Self-attention and Transformer. GPT family.