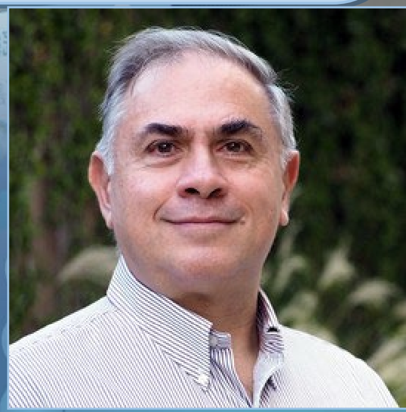
A Short Course on Optimization for Data Science and Machine Learning Problems

Speaker

Prof. George Michailidis

Department of Statistics & Data Science,
University of California, Los Angeles, U.S.A.



Venue

Auditorium, B1F, the Institute of Statistical Science Building

***An electronic certificate is available upon request and can be issued only to those who attend all the lectures in person.**

***Online live streaming through Microsoft Teams will be available.**

***Lunch boxes will be provided on Dec 9 and Dec 10.**

December 9 (Tue.)
11:00 a.m.~12:00 p.m.
1:00 p.m.~ 4:10 p.m.

December 10 (Wed.)
9:00 a.m.~12:10 p.m.

December 12 (Fri.)
1:00 p.m.~ 4:10 p.m.

December 15 (Mon.)
1:00 p.m.~ 4:10 p.m.

Course outline

Part I: Foundations of Gradient-Based Optimization	**Part II: Modern Developments in Optimization for Machine Learning**
(a) Fundamentals of Optimization ● Introduction to unconstrained and constrained optimization problems ● **Gradient Descent:** update rules, step-size (learning rate) selection, convergence analysis, and stopping criteria ● **Stochastic Gradient Descent (SGD):** mini-batching, variance reduction, and convergence properties ● **Accelerated and Adaptive Gradient Methods:** momentum and Nesterov acceleration; adaptive learning-rate algorithms including Adagrad, AMSGrad, Adam, and Adam-W, with discussion of their convergence behavior and empirical performance in large-scale learning **(b) Advanced optimization methods** ● **Proximal** algorithms for nonsmooth and regularized optimization (e.g., Lasso, sparse models) ● **Splitting and decomposition techniques**, including the **Alternating Direction Method of Multipliers (ADMM)** ● **Block Coordinate** and Coordinate Descent methods for large-scale and structured optimization	**(a) Distributed and Federated Learning** ● Introduction to **distributed optimization** and communication-efficient algorithms ● Fundamentals of **Federated Learning**: architecture, privacy, and security considerations ● **Stochastic and asynchronous methods:** distributed SGD, variance reduction, and adaptive updates **(b) Minimax Optimization and Adversarial Learning** ● Optimization algorithms for saddle-point problems: Gradient Descent–Ascent, Extra-Gradient, and Optimistic methods ● Convergence and stability analysis in nonconvex–nonconcave settings ● Applications: Generative Adversarial Networks (GANs), robust training, domain adaptation, and fairness in AI **(c) Bilevel Optimization and Applications in Deep Learning** ● Motivation and formulation of bilevel optimization problems ● Algorithms: approximate gradient methods, implicit differentiation, and hypergradient-based approaches ● Applications: hyperparameter optimization, meta-learning, neural architecture search (NAS), and AutoML ● Computational complexity, scalability, and recent advances in efficient implementations

中央研究院
統計科學研究所