ECE 592-103
Theoretical Foundations of Large-Scale Machine Learning and Optimization

Instructor(s): Aritra Mitra; email id: amitra2@ncsu.edu

Objective or Description: *Why do optimization algorithms for training large neural nets work? Are there connections between the dynamics of optimization algorithms for supervised learning and common reinforcement learning (RL) algorithms for decision-making?*

This course will explore the above questions (and more) by focusing on the key mathematical techniques at the heart of the rapidly evolving area of large-scale machine learning. We will start with an in-depth complexity analysis of various contemporary first-order stochastic optimization algorithms. To appreciate the applicability of such algorithms in machine learning tasks, we will turn to empirical risk minimization and derive generalization bounds. Building on the stochastic optimization framework, the course will then cover several advanced topics spanning min-max optimization, system-level challenges in distributed and federated learning, and a convergence theory for iterative reinforcement learning algorithms.

The course will focus on highlighting the main theoretical principles underlying the design of modern ML algorithms: sample complexity, statistical accuracy, computational complexity, and the trade-offs therein. The end goal will be to enable students to understand *why* a large class of iterative learning algorithms work by drawing on diverse and often non-traditional viewpoints, including perspectives from the classical Lyapunov stability theory in control.

Prerequisites: Undergraduate probability, calculus, linear algebra, and an introductory course on optimization.

Textbook: Multiple notes and reference papers that are freely available online; these will be made available to the students as and when needed.

Topics:

- **Introduction and Background Material**
    - Motivation: Key optimization challenges in ML.
    - Probability Overview: Concentration inequalities.
    - Convex Analysis Overview: Convex sets and functions.

- **Complexity Theory of Stochastic Optimization Algorithms**
    - Convergence rates for Gradient Descent on various function classes.
    - The Stochastic Gradient Descent (SGD) Method and analysis.
    - Speeding up SGD via Variance Reduction: Incremental algorithms like SAG, SAGA, and SVRG.
    - Non-convexity and the PL land.
    - Understanding optimization algorithms from the lens of Lyapunov Stability Theory.

- **<u>Connections to Learning Theory via Empirical Risk Minimization</u>**
    - Empirical Risk, Rademacher Complexity, and VC Dimension.
    - Algorithmic Stability and Generalization Bounds for SGD.

- **<u>Duality Theory and Min-Max Optimization</u>**
    - Duality, Slater's Condition, and Saddle Points.
    - Min-Max Optimization with applications to Adversarially Robust ML.

- **<u>Challenges in Modern Large-Scale (Federated) Learning</u>**
    - Information constraints: Quantization and Privacy.
    - Security constraints: Robust Aggregation Protocols.
    - Statistical Heterogeneity (Multi-Task Learning and Meta-Learning).

- **<u>Sample Complexity of Reinforcement Learning Algorithms: An Optimization Viewpoint</u>**
    - Introduction to MDPs and basic RL algorithms (e.g., TD- and Q-Learning).
    - Optimization-based analysis of iterative RL algorithms.

<u>Grading:</u> HW (35%); One Mid-Term (30%); Project (35%).