ELLIOTT PAQUETTE, McGill University
High-dimensional Optimization with Applications to Compute-Optimal Neural Scaling Laws

Given the massive scale of modern ML models, we now only get a single shot to train them effectively. This restricts our ability to test multiple architectures and hyper-parameter configurations. Instead, we need to understand how these models scale, allowing us to experiment with smaller problems and then apply those insights to larger-scale models. In this talk, I will present a framework for analyzing scaling laws in stochastic learning algorithms using a power-law random features model, leveraging high-dimensional probability and random matrix theory. I will then use this scaling law to address the compute-optimal question: How should we choose model size and hyper-parameters to achieve the best possible performance in the most compute-efficient manner?