CHRISTOPHER MUSCO, New York University

Robust Active Learning via Leverage Score Sampling

Active learning is a promising approach to fitting machine learning models in "data starved" applications, where the cost of collecting data labels is the primary cost of model training. In many of these applications, including in computational science and ML guided engineering, we need active learning methods that work in the challenging agnostic or "adversarial noise" setting. In this setting, collected labels might not match the model being trained, even in expectation. Nevertheless, we seek methods that are robust enough to find the best possible fit with as little data as possible. In this talk, I will discuss recent developments on a flexible class of active learning algorithms based on so-called "leverage score sampling". I will show how leverage score based methods can provably address the challenging agnostic learning problem in a variety of settings, including for linear models, kernel regression models, and also simple neural networks with non-linearities. I will highlight future directions for research and challenging open directions. Based on joint work with Aarshvi Gajjar, Tamás Erdélyi, Chinmay Hegde, Raphael Meyer, Cameron Musco David Woodruff, Taisuke Yasuda, and Samson Zhou.