An interpolation perspective on modern machine learning

Mikhail Belkin Ohio State University

A striking feature of modern supervised machine learning is its consistent use of techniques that fit the training data exactly.
Deep networks often containing several orders of magnitude more parameters than data points, are trained to obtain zero or near zero error on the training set. Yet, at odds with most theory, they show excellent test performance. It has become accepted wisdom that these properties are special to deep networks and require non-convex analysis to understand.

In this talk I will show that classical (convex) kernel machines do, in fact, exhibit these unusual properties. Indeed, kernel machines explicitly constructed to interpolate (achieve zero regression error) the training data, show excellent test performance. I will show empirical and theoretical results indicating that we are unlikely to make progress on understanding deep learning until we develop a fundamental understanding of classical "shallow" kernel classifiers in the "modern" near-interpolated setting. I will proceed to discuss a number of theoretical and practical consequences of interpolation including results on the optimality of certain interpolated classifiers.