The University of Chicago Computer Science Department
PRESENTS:

"Fit without fear: a kernel perspective on modern machine learning"

Mikhail Belkin
The Ohio State University

Abstract:
It has become accepted wisdom that understanding the success of modern machine learning necessitates analysis of inference in non-convex neural network architectures. Indeed, scaling classical convex architectures, such as kernels, to modern big data has proved to be challenging. In this talk I will show that much of the challenge is due not to their intrinsic limitations, but to a mismatch between the algorithms feasible with the current computational technology, such as stochastic gradient descent, and the structure of kernel spaces. We address this mismatch by designing a new spectrally modified kernel. This can be done computationally efficiently with only a small overhead per iteration. The modified (EigenPro) kernel provides 5x-30x acceleration of training over the standard iterative methods. That allows us to exceed or match state-of-the-art results in the kernel literature using a single GPU, a small fraction of their reported computational resources.

Furthermore, this new class of kernel methods gives a different perspective on modern ML. One of the most interesting and innovative aspects of neural networks is the systematic use of over-parametrization in neural networks, where the number of parameters far exceeds the number of data points. This often leads to over-fitting (near interpolation of the training data), yet these over-fitted models still perform well on test data. This phenomenon is not easily explained using classical machine learning analyses. While we still lack a solid understanding of why over-fitted models generalize to unseen data, I will show that kernel methods show very similar behaviour once their computational limitations are addressed. They also reflect other important aspects of neural net inference including effectiveness of stochastic gradient descent and the accelerated (momentum) methods. We argue that many significant aspects of modern Machine Learning are not related to the architectural depth of neural networks, and can now be studied in a much more tractable convex setting.

Bio:
Mikhail Belkin is a Professor in the Department of Computer Science and Engineering and the Department of Statistics at the Ohio State University. He received a PhD from the University of Chicago in Mathematics in 2003. His research focuses on understanding the fundamental structure in data, the principles of recovering these structures and their computational, mathematical and statistical properties. This understanding, in turn, leads to algorithms for dealing with real-world data. A number of his algorithms have been widely used in applications. Prof. Belkin is a recipient of an NSF Career Award and a number of best paper and other awards. He has served on the editorial boards of the Journal of Machine Learning Research and IEEE PAMI.

Monday, November 27, 2017
2:30 pm
Ryerson 251
Host: Risi Kondor