EPSTEIN INSTITUTE SEMINAR ▪ ISE 651

Accelerating Stochastic Gradient Descent for Convex and Non-Convex Optimization

ABSTRACT – Stochastic optimization methods, such as the stochastic gradient descent (SGD) method, are the workhorse of modern machine learning for both convex and non-convex optimization. What are the most effective strategies for speeding up, i.e. "accelerating", these methods? This talk will characterize the effectiveness of widely used practical approaches in both convex and non-convex settings, through a mix of theoretical analysis and empirical studies.

In practice, stochastic variants of momentum based gradient methods, including the heavy ball (HB) and Nesterov's accelerated gradient descent (NAG), are widely used for training deep networks and other supervised learning models, due to the observed performance improvements. We provide a counterpoint to the widely held belief that these methods are fundamentally effective for stochastic optimization, by proving that there are simple (non-pathological) problem instances where these methods cannot outperform SGD. Through theory and empirical studies, we argue that practical performance gains of these fast momentum based methods are a by-product of "mini-batching".

As an alternative, we provide a new algorithm for accelerating stochastic optimization; this new algorithm provably improves upon stochastic gradient descent (and provably improves upon the HB and NAG methods in the stochastic regimes); we demonstrate its effectiveness for both convex and non-convex optimization.

Dr. Sham Kakade
Washing Research Foundation Data Science Chair
University of Washington

SPEAKER BIO – Dr. Sham Kakade is a Washington Research Foundation Data Science Chair, with a joint appointment in both the Computer Science & Engineering and Statistics departments at the University of Washington. He completed his Ph.D. at the Gatsby Computational Neuroscience Unit at University College London, advised by Peter Dayan, and earned his B.S. in physics at Caltech. Before joining the University of Washington, Sham was a principal research scientist at Microsoft Research, New England. Prior to this, Sham was an associate professor at the Department of Statistics, Wharton, University of Pennsylvania (2010-2012) and an assistant professor at the Toyota Technological Institute at Chicago (2005-2009). Sham completed a postdoc in the Computer and Information Science Department at the University of Pennsylvania under the supervision of Michael Kearns.

He works in the area broadly construed as data science, focusing on designing (and implementing) both statistically and computationally efficient algorithms for machine learning, statistics, and artificial intelligence. His intent is to see these tools advance the state of the art on core scientific and technological problems.

One line of his work has been in providing computationally efficient algorithms for statistical estimation, which has included the estimation of various statistical models with hidden (or latent) structure (including mixture models, topic models, hidden markov models, and models communities in social networks). More broadly, Sham has made various contributions in various areas including statistics, optimization, probability theory, machine learning, algorithmic game theory and economics, and computational neuroscience. He has had numerous roles in chairing conferences and workshops, has given numerous plenary talks, and has received various awards.

USC Viterbi
School of Engineering
Daniel J. Epstein Department of Industrial and Systems Engineering

TUESDAY, APRIL 3, 2018
11:00AM – 12:00PM
USC ANDRUS GERONTOLOGY CENTER (GER), ROOM 206