

Stanford University
Department of Statistics

DEPARTMENTAL SEMINAR

4:30pm, Tuesday, November 27, 2018
Sloan Mathematics Center Room 380C

Refreshments served at 4pm in Sequoia Lounge.

Speaker: Rachel Ward, *University of Texas at Austin*

Title: **AdaGrad for strong SGD convergence guarantees without step size tuning**

Abstract:

Stochastic Gradient Descent (SGD) is the basic optimization algorithm behind powerful deep learning architectures which are becoming increasingly omnipresent in society. However, the performance of SGD is quite sensitive to the choice of step size or step size schedule; thus, in practice, the step size is tuned by hand before training, taking days or weeks. In this talk, we discuss a modified SGD algorithm with a simple adaptive step size update rule which can be viewed as a “norm” version of the well-known AdaGrad algorithm. We provide the first theoretical guarantees of convergence for such an adaptive algorithm in the setting of SGD: the adaptive algorithm converges to a near-stationary point of a smooth loss function, at a rate which nearly matches the “oracle” rate as if the Lipschitz constant of the gradient and noise level on the stochastic gradient were known in advance. We provide several numerical experiments on both simple and deep learning problems to complement the theory, and finish by discussing several open questions.