RUTGERS
School of Arts and Sciences

Department of Statistics

DEPARTMENT OF STATISTICS

Mert Gurbuzbalaban

Department of Management Science and Information Systems
Rutgers University

Heavy-tail Phenomenon in SGD

Wednesday October 21st, 2020
11:30 AM EST

Zoom Meeting: Meeting ID: 940 0298 5296
Password: 369797
Virtual Coffee session before the seminar at 11:30AM EST

Abstract: In the first part of the talk, we focus on the gradient noise (GN) in the stochastic gradient descent (SGD) algorithm for training deep neural networks (DNNs). In this setting, GN is often considered to be Gaussian in the large data regime by assuming that the classical central limit theorem (CLT) kicks in. This assumption is often made for mathematical convenience, since it enables SGD to be analyzed as a stochastic differential equation (SDE) driven by a Brownian motion. We argue that the Gaussianity assumption might fail to hold in deep learning settings and hence render the Brownian motion-based analyses inappropriate. We conduct extensive experiments on common deep learning architectures and show that in all settings, the GN is highly non-Gaussian and admits heavy-tails. We further investigate the tail behavior in varying network architectures and sizes, loss functions, and datasets. By exploiting connections to Levy-driven differential equations and their metastability properties, our results open up a different perspective and shed more light on the belief that SGD prefers wide minima. In the second part of the talk, we question the origins of heavy tails in SGD iterations further and their link to various notions of capacity and complexity that have been proposed for characterizing the generalization properties of SGD in deep learning. Some of the popular notions that correlate well with the performance on unseen data are (i) the 'flatness' of the local minimum found by SGD, which is related to the eigenvalues of the Hessian, (ii) the ratio of the stepsize η to the batch size b, which essentially controls the magnitude of the stochastic gradient noise, and (iii) the 'tail-index', which measures the heaviness of the tails of the eigenspectra of the network weights. We argue that these three seemingly unrelated perspectives for generalization are deeply linked to each other. We claim that depending on the structure of the Hessian of the loss at the minimum, and the choices of the algorithm parameters η and b, the SGD iterates will converge to a heavy-tailed stationary distribution. We rigorously prove this claim in the setting of linear regression: we show that even in a simple quadratic optimization problem with independent and identically distributed Gaussian data, the iterates can be heavy-tailed with infinite variance. We further characterize the behavior of the tails with respect to algorithm parameters, the dimension, and the curvature. We then translate our results into insights about the behavior of SGD in deep learning. We finally support our theory with experiments conducted on both synthetic data and neural networks.

Bio: Mert Gürbüzbalaban is an assistant professor at Rutgers University. Previously, he was a postdoctoral associate at the Laboratory for Information and Decision Systems (LIDS) at MIT. He is broadly interested in optimization and computational science driven by applications in large-scale information and decision systems. He received his B.Sc. degrees in Electrical Engineering and Mathematics as a valedictorian from Boğaziçi University, Istanbul, Turkey, the "Diplôme d'ingénieur" degree from École Polytechnique, France, and the M.S. and Ph.D. degrees in (Applied) Mathematics from the Courant Institute of Mathematical Sciences, New York University. Dr. Gürbüzbalaban received the Kurt Friedrichs Prize (given by the Courant Institute of New York University for an outstanding thesis) in 2013, Bronze Medal in the École Polytechnique Scientific Project Competition in 2006, the Nadir Orhan Bengisu Award (given by the electrical-electronics engineering department of Boğaziçi University to the best graduating undergraduate student) in 2005 and the Bülent Kerim Altay Award from the Electrical-Electronics Engineering Department of Middle East Technical University in 2001. He received funding from a variety of sources including multiple programs at the U.S. National Science Foundation.