



Contribution ID: 42

Type: not specified

Peter Bartlett - The Dynamics of Sharpness-Aware Minimization.

Thursday, October 6, 2022 3:30 PM (1 hour)

Optimization methodology has been observed to affect statistical performance in high-dimensional prediction problems, and there has been considerable effort devoted to understanding the behavior of optimization methods and the nature of solutions that they find. We consider Sharpness-Aware Minimization (SAM), a gradient-based optimization method that has exhibited performance improvements over gradient descent on image and language prediction problems using deep networks. We show that when SAM is applied with a convex quadratic objective, for most random initializations it converges to oscillating between either side of the minimum in the direction with the largest curvature, and we provide bounds on the rate of convergence. In the non-quadratic case, we show that such oscillations encourage drift toward wider minima by effectively performing gradient descent, on a slower time scale, on the spectral norm of the Hessian. (Based on joint work with Olivier Bousquet and Phil Long)