武汉大学

随机分析系列报告（十一）

The Heavy-Tail Phenomenon in SGD

报告人： Prof. Lingjiong Zhu (Florida State University, USA)

时间： 2021 年 11 月 8 日，上午 10:00 - 11:30

地点： Zoom meeting ID: 733 955 6904

摘要： In recent years, various notions of capacity and complexity have been proposed for characterizing the generalization properties of stochastic gradient descent (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 to the batch-size, which essentially controls the magnitude of the stochastic gradient noise, and (iii) the tail-index, which measures the heaviness of the tails of the network weights at convergence. In this paper, 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, the distribution of the SGD iterates will converge to a heavy-tailed stationary distribution. We rigorously prove this claim in the setting of quadratic optimization: we show that even in a simple linear regression problem with independent and identically distributed data whose distribution has finite moments of all order, 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 support our theory with experiments conducted on synthetic data, fully connected, and convolutional neural networks. This is based on the joint work with Mert Gurbuzbalaban and Umut Simsekli.

主办单位： 武汉大学数学与统计学院
School of Mathematics and Statistics, Wuhan University