BIOSTATISTICS SEMINAR

From minimal norm solutions to ensemble learning: A prototypical study

Qiang Sun, PhD
Assistant Professor, Department of Statistics
University of Toronto

Despite of their successes in a number of applications, deep neural networks are not yet fully understood. One of such perhaps surprising statistical phenomena is: Over fitted estimators, especially those achieved by (stochastic) gradient descent, can generalize well. Recent studies have focused studying these phenomena in simpler models such as linear regressions and two layer neural nets and revealed that (stochastic) gradient descent induces implicit regularization and converge to the minimum norm solutions whose out-of-sample risk curves exhibit an interesting double-descent phenomenon. The risk exhibits a U-shape curve in the classical statistical regime, explodes at $d=n$, and starts to drop down again in the interpolation regime. We leverage a dual view to show that random sketched (random sub-sampled) regression not only reduces the computational complexity but also the out-sample performance. Another drawback of the minimum norm estimator is that its risk explodes at $d=n$ due to its unbounded variance, partly explaining the fragility of such estimator to noise. From a different viewpoint, ensemble learning is the industrial standard for prediction tasks involving high dimensional noisy dataset. The intuition is that randomization also induces implicit regularization. Given the same sample budget, we show that ensemble of the one-sample least square estimators exhibit a smoother out-of-sample risk curve when increasing d/n, which is much better that of a minimum-norm solution when the noise is high.

Thursday April 21, 2022, 3:30-4:30 PM Eastern

133 Rosenau Hall

Virtual using link and info below.

Link: https://unc.zoom.us/j/98412143955?pwd=a1p6c3hvZ28wSnk3dVVXQWl0dEpzdz09

Meeting ID: 984 1214 3955 Passcode: 0375501630

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH Biostatistics