



THE UNIVERSITY OF
CHICAGO

Department of Statistics
STATISTICS COLLOQUIUM

SONG MEI

Institute for Computational & Mathematical Engineering and
Department of Statistics, Stanford University

Generalization Error of Linearized Neural Networks:
Staircase and Double-descent

MONDAY, January 13, 2020 at 4:00 PM
Jones 303, 5747 S. Ellis Avenue

ABSTRACT

Deep learning methods operate in regimes that defy the traditional statistical mindset. Despite the non-convexity of empirical risks and the huge complexity of neural network architectures, stochastic gradient algorithms can often find the global minimizer of the training loss and achieve small generalization error on test data. As one possible explanation to the training efficiency of neural networks, tangent kernel theory shows that a multi-layers neural network — in a proper large width limit — can be well approximated by its linearization. As a consequence, the gradient flow of the empirical risk turns into a linear dynamics and converges to a global minimizer. Since last year, linearization has become a popular approach in analyzing training dynamics of neural networks. However, this naturally raises the question of whether the linearization perspective can also explain the observed generalization efficacy. In this talk, I will discuss the generalization error of linearized neural networks, which reveals two interesting phenomena: the staircase decay and the double-descent curve. Through the lens of these phenomena, I will also address the benefits and limitations of the linearization approach for neural networks.

For further information and inquiries about building access for persons with disabilities, please contact Jonathan Rodriguez at 773.702.8333 or send him an email at jgrodriquez@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following website:
<https://lists.uchicago.edu/web/subscribe/statseminars>.