The impact of data structure on learning in two-layer neural networks

Understanding the impact of data structure on learning in neural networks remains a key challenge for the theory of neural networks.
In this talk, we will go beyond the simple i.i.d. modelling paradigm for inputs by studying neural networks trained on data drawn from structured generative models.
We will discuss two results: (1) We give rigorous conditions under which a class of generative models shares key statistical properties with an appropriately chosen Gaussian feature model. (2) We use this Gaussian equivalence to analyse the dynamics of two-layer neural networks trained using one-pass stochastic gradient descent on data drawn from a large class of generators. We complement our theoretical results with experiments demonstrating how our theory applies to deep, pre-trained generative models.
Joint work with Galen Reeves, Marc Mézard, Florent Krzakala and Lenka Zdeborová.