Dynamics & Phase Transitions in Wide, Deep Neural Networks

Yasaman Bahri

Google Research, Brain Team, USA

The study of deep neural networks whose hidden layer widths are large has been fruitful in building theoretical foundations for deep learning. This has given rise to exact connections between neural networks, Gaussian processes, and kernel methods. Nonetheless, these results hold only up to a critical learning rate used in gradient descent optimization. At larger learning rates with squared loss, a phase transition to a different, nonlinear regime with universal features across architectures and datasets appears to occur, even in the large width limit. I will describe our theoretical understanding of this phase transition through the study of a class of simple dynamical systems distilled from neural network evolution in function space.