Berfin Simsek: Finite-Width Neural Networks: A Landscape Complexity Analysis

In this talk, I will present an average-case analysis of finite-width neural networks through permutation symmetry. First, I will give a new scaling law for the critical manifolds of finite-width neural networks derived from counting all partitions due to neuron splitting from an initial set of neurons. Zooming into a line of critical points, I will discuss intriguing transitions from saddles to minima through non-strict saddle points. Considering the invariance of zero neuron addition, we will derive the scaling law of the zero-loss manifold that is exact for the population loss. The competition between these two scaling laws gives a notion of landscape complexity of finitely overparameterized neural networks. At the onset of overparameterization, the complexity explodes and then gradually decreases with further overparameterization, dropping to zero for infinitely wide networks. Our results enable a quantitative understanding of (linear) mode connectivity. Finally, based on our theory, we propose an `Expand-Cluster' algorithm for neuron pruning in practice.