CAMERON JAKUB, University of Guelph

Depth Degeneracy in Neural Networks: Vanishing Angles in Fully Connected ReLU Networks on Initialization

Despite remarkable performance on a variety of tasks, many properties of deep neural networks are not yet theoretically understood. One such mystery is the depth degeneracy phenomenon: the deeper you make your network, the closer your network is to a constant function on initialization. In this paper, we examine the evolution of the angle between two inputs to a ReLU neural network as a function of the number of layers. By using combinatorial expansions, we find precise formulas for how fast this angle goes to zero as depth increases. These formulas capture microscopic fluctuations that are not visible in the popular framework of infinite width limits, and leads to qualitatively different predictions. We validate our theoretical results with Monte Carlo experiments and show that our results accurately approximate finite network behaviour. We also empirically investigate how the depth degeneracy phenomenon can negatively impact training of real networks. The formulas are given in terms of the mixed moments of correlated Gaussians passed through the ReLU function. We also find a surprising combinatorial connection between these mixed moments and the Bessel numbers that allows us to explicitly evaluate these moments.