

Collapse of deep and narrow neural nets

Lu Lu, Brown University & Yeonjong Shin, Brown University

Recent theoretical work has demonstrated that deep neural networks have superior performance over shallow networks, but their training is more difficult, e.g., they suffer from the vanishing gradient problem. This problem can be typically resolved by the rectified linear unit (ReLU) activation. However, here we show that even for such activation, deep and narrow neural networks (NNs) will converge to erroneous mean or median states of the target function depending on the loss with high probability. Deep and narrow NNs are encountered in solving partial differential equations with high-order derivatives. We demonstrate this collapse of such NNs both numerically and theoretically, and provide estimates of the probability of collapse. We also construct a diagram of a safe region for designing NNs that avoid the collapse to erroneous states. Finally, we examine different ways of initialization and normalization that may avoid the collapse problem. Asymmetric initializations may reduce the probability of collapse but do not totally eliminate it.