

Which ReLU Net Architectures Give Rise to Exploding and Vanishing Gradients?

Boris Hanin, Texas A&M / Facebook

Due to its compositional nature, the function computed by a deep neural net often produces gradients whose magnitude is either very close to 0 or very large. This so-called vanishing and exploding gradient problem is often already present at initialization and is a major impediment to gradient-based optimization techniques. I will give a rigorous answer to the question of which neural architectures have exploding and vanishing gradients for feed-forward neural nets with ReLU activations. The results presented will cover both independent and orthogonal weight initializations. The results are partly joint with Mihai Nica (Toronto and David Rolnick (UPenn)).