The separation capacity of random neural networks

Sjoerd Dirksen
Utrecht University, The Netherlands
s.dirksen@uu.nl

Neural networks with random weights appear in a variety of machine learning applications, most prominently as the initialization of many deep learning algorithms and as a computationally cheap alternative to fully learned neural networks. The first goal of this talk is to enhance the theoretical understanding of random neural networks by addressing the following data separation problem: under which conditions can a random neural network make two classes (with positive distance) linearly separable? I will show that a sufficiently large two-layer ReLU-network with Gaussian weights and uniformly distributed biases can solve this problem with high probability. Building on the insights behind this result, I will next present a simple randomized algorithm to produce a small interpolating neural net for a given dataset with two classes. In both results, the size of the network is explicitly linked to geometric properties of the two classes and their mutual arrangement. This instance-specific viewpoint allows to overcome the curse of dimensionality. I will connect the presented results with related work on approximation, memorization, and generalization.

Joint work with Patrick Finke (Utrecht University, The Netherlands), Martin Genzel (Utrecht University, The Netherlands and Merantix Momentum), Laurent Jacques (UC Louvain, Belgium) and Alexander Stollenwerk (UC Louvain, Belgium and KPMG Germany).