Optimal Bump Functions for Shallow ReLU Networks: Weight Decay, Depth Separation and the Curse of Dimensionality

Friday, September 16, 2022
12:00PM–1:00PM
PGH 232

Abstract: We study how neural networks with a single hidden layer and ReLU activation interpolate data drawn from a radially symmetric distribution with target labels 1 at the origin and 0 outside the unit ball, if no labels are known inside the unit ball. With weight decay regularization and in the infinite neuron, infinite data limit, we prove that a unique radially symmetric minimizer exists, whose weight decay regularizer and Lipschitz constant grow as dimension and \( \sqrt{d} \), respectively. We furthermore show that the weight decay regularizer grows exponentially in \( d \) if the label 1 is imposed on a ball of radius \( \epsilon > 0 \) rather than just at the origin. For comparison, a neural networks with two hidden layers can approximate the target function without encountering the curse of dimensionality. As applications, we discuss approximation rates using mollification and the empirical study of optimization algorithms.