

DIMENSION-FREE LIMITS OF STOCHASTIC GRADIENT DESCENT FOR TWO-LAYERS NEURAL NETWORKS

Bruno Loureiro

École Normale Supérieure, France
bruno.loureiro@di.ens.fr

Stochastic gradient descent and its variants are the workhorse of modern machine learning models. Despite its current (and successful) use for optimising non-convex problems associated with the training of heavily overparametrised models, most of our theoretical understanding is bound to the context of convex problems. In this talk, I will discuss some recent progress in understanding the SGD dynamics of perhaps one of the simplest non-convex problems: two-layers neural networks. In particular, I will discuss different regimes (classical, high-dimensional and overparametrised) where one can derive a set of low-dimensional “state evolution” equations describing the evolution of the sufficient statistics for the weights. Finally, I discuss some interesting behaviour associated to each regime, and the connections to other descriptions such as the mean-field limit.