Phase diagram of Stochastic Gradient Descent in high-dimensional two-layer neural networks

Bruno Loureiro, École Polytechnique Fédérale de Lausanne (EPFL)

Despite the non-convex optimization landscape, over-parametrized shallow networks are able to achieve global convergence under gradient descent. The picture can be radically different for narrow networks, which tend to get stuck in badly-generalizing local minima. Here we investigate the cross-over between these two regimes in the high-dimensional setting, and in particular investigate the connection between the so-called mean-field/hydrodynamic regime and the seminal approach of Saad & Solla. Focusing on the case of Gaussian data, we study the interplay between the learning rate, the time scale, and the number of hidden units in the high-dimensional dynamics of stochastic gradient descent (SGD). Our work builds on a deterministic description of SGD in high-dimensions from statistical physics, which we extend and for which we provide rigorous convergence rates.