Federica Gerace

Title: Gaussian Universality of Perceptrons with Random Labels

Abstract: While classical in many theoretical settings — and in particular in statistical physics-inspired works— the assumption of Gaussian i.i.d. input data is often perceived as a strong limitation in the context of statistics and machine learning. In this study, we redeem this line of work in the case of generalized linear classification, a.k.a. the perceptron model, with random labels. We argue that there is a large universality class of high-dimensional input data for which we obtain the same minimum training loss as for Gaussian data with corresponding data covariance. In the limit of vanishing regularization, we further demonstrate that the training loss is independent of the data covariance. On the theoretical side, we prove this universality for an arbitrary mixture of homogeneous Gaussian clouds. Empirically, we show that the universality holds also for a broad range of real datasets.