Francesco Camilli

Title: Fundamental limits of shallow neural networks with small training sets

Abstract: In my presentation I shall provide an information-theoretical analysis of a two-layer neural network trained on a relatively small dataset compared to the network size. The dataset is generated by a teacher network with the same architecture. The main finding of this study is the asymptotic equivalence of the two-layer neural network with a Generalized Linear Model, where the Mutual information between the weights and the training set is known. This result in turn yields the Bayes-optimal generalization error, which serves as a lower bound for any neural network with the mentioned architecture. The proof relies on rigorous Mathematical Physics tools used in the study of spin glasses, such as the interpolation scheme, and it is guided by recently conjectured Gaussian Equivalence Principles. With respect to the existing literature, which is either non-rigorous, or restricted to the case of the learning of the readout weights only, the proof addresses the learning of all the network parameters. While our techniques are primarily applicable to the regime of small datasets, it offers the advantage of being self-contained, simple, and it constitutes an independent proof of a Gaussian Equivalence Principle.