The exponential capacity of modern associative memories

Carlo Lucibello
Bocconi University, Milano, Italy

The Hopfield model is a paradigmatic model of associative memory which is able able to retrieve the stored patterns from noisy observations. Thanks to the replica method from spin glass theory, it has been shown that the model is able to store a number of uncorrelated patterns that scales linearly with the size of the system, with the asymptotic threshold that can be computed to high precision. In this talk, we present the statistical physics analysis of a recently proposed generalization of the Hopefield model, named modern Hopfield network (MHN). The MHN has an exponential capacity, i.e. it is able to store $P = exp(\alpha\ N)$ patterns where N is the size of the system, for an exponential rate α low enough. Besides this huge storage capacity, the MHN is linked to the attention mechanism of Transformer architectures in deep learning. In fact, one step in the dynamics of a MHN can be mapped into the forward pass of an attention layer. We provide the phase diagram of the model thanks to a large deviation analysis of a Random Energy Model (REM) related to the problem. Our framework allows the analysis of a large class of pattern ensembles, each one inducing a characteristic distribution in the related REM problem. We derive exact thresholds for single pattern retrieval and lower bounds for all patterns retrieval largely improving the existing ones. Additionally, we are able to compute the exact size of the basins of attraction, and to analyze different scaling regimes of the number of patterns and the coupling strength with N. The generic framework we present is then applied to the cases of gaussian and spherical patterns. For spherical patterns, we find that the lower bound for the full retrieval is sharp, and that one can arbitrarily increase the capacity by increasing the interaction strength. For Gaussian patterns instead we find a richer picture where the the lower bound and the single pattern retrieval threshold do not match and for any interaction strength there is a limit capacity above which the system enters either a "liquid" phase or a "condensed" spin glass phase.