The out-of-equilibrium landscape of neural networks learning algorithms: from driven stochastic processes to quantum annealing.

R.Zecchina
Bocconi University, Italy

Abstract:

Training deep neural networks is a highly non-convex problem which however is often solved in practice by variants of Stochastic Gradient Descent.

In order to shed light on this phenomenon, we have developed a large deviation analysis based on a local entropy measure which has revealed the existence of subdominant and extremely dense regions of optimal solutions that have a number of highly desirable properties. They are wide minima with good generalization properties. The analysis allowed us to develop a number of novel stochastic algorithms and to find a fundamental connection with quantum annealing. We will show that in spite of the fact that the energy landscape is exponentially dominated by local minima that trap classical thermal annealers, quantum annealing converges efficiently to subdominant dense regions of optimal solutions. This is a concrete example of exponential quantum vs thermal speed up.