Baldassi, Carlo; Pittorino, Fabrizio; Zecchina, Riccardo

Shaping the learning landscape in neural networks around wide flat minima. (English)


Summary: Learning in deep neural networks takes place by minimizing a nonconvex high-dimensional loss function, typically by a stochastic gradient descent (SGD) strategy. The learning process is observed to be able to find good minimizers without getting stuck in local critical points and such minimizers are often satisfactory at avoiding overfitting. How these 2 features can be kept under control in nonlinear devices composed of millions of tunable connections is a profound and far-reaching open question. In this paper we study basic nonconvex 1- and 2-layer neural network models that learn random patterns and derive a number of basic geometrical and algorithmic features which suggest some answers. We first show that the error loss function presents few extremely wide flat minima (WFM) which coexist with narrower minima and critical points. We then show that the minimizers of the cross-entropy loss function overlap with the WFM of the error loss. We also show examples of learning devices for which WFM do not exist. From the algorithmic perspective we derive entropy-driven greedy and message-passing algorithms that focus their search on wide flat regions of minimizers. In the case of SGD and cross-entropy loss, we show that a slow reduction of the norm of the weights along the learning process also leads to WFM. We corroborate the results by a numerical study of the correlations between the volumes of the minimizers, their Hessian, and their generalization performance on real data.

MSC:

92B20 Neural networks for/in biological studies, artificial life and related topics
68T05 Learning and adaptive systems in artificial intelligence
82D30 Statistical mechanics of random media, disordered materials (including liquid crystals and spin glasses)
90C26 Nonconvex programming, global optimization
90C90 Applications of mathematical programming

Software:

Fashion-MNIST

Full Text: DOI

References:


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.