

Consistent Sparse Deep Learning: Theory and Computation

Faming Liang

Purdue University, USA
fmliang@purdue.edu

Abstract

Deep learning has been the engine powering many successes of data science. However, the deep neural network (DNN), as the basic model of deep learning, is often excessively over parameterized, causing many difficulties in training, prediction and interpretation. We propose a frequentist-like method for learning sparse DNNs and justify its consistency under the Bayesian framework: the proposed method can learn a sparse DNN with at most $O(n/\log(n))$ connections and nice theoretical guarantees such as posterior consistency, variable selection consistency and asymptotically optimal generalization bounds. In particular, we establish posterior consistency for the sparse DNN with a mixture Gaussian prior, show that the structure of the sparse DNN can be consistently determined using a Laplace approximation-based marginal posterior inclusion probability approach, and use Bayesian evidence to elicit sparse DNNs learned by an optimization method such as stochastic gradient descent in multiple runs with different initializations. The proposed method is computationally more efficient than standard Bayesian methods for large-scale sparse DNNs. The numerical results indicate that the proposed method can perform very well for large-scale network compression and high-dimensional nonlinear variable selection, both advancing interpretable machine learning. This talk is based on a joint work with Yan Sun and Qifan Song.