Journal of Machine Learning

Journal of Machine Learning (JML) publishes high quality research papers in all areas of machine learning, including innovative algorithms of machine learning, theories of machine learning, important applications of machine learning in AI, natural sciences,social sciences, and engineering etc. The journal emphasizes a balanced coverage of both theory and practice. The journal is published in a timely fashion in electronic form. All articles in JML are open-access and there is no charge for the authors.

On the Existence of Global Minima and Convergence Analyses for Gradient Descent Methods in the Training of Deep Neural Networks

Arnulf Jentzen & Adrian Riekert

DOI: 10.4208/jml.220114a, J. Mach. Learn., 1 (2022), pp. 141-246.

Communicated by: Weinan E

Category: Theory

Summary for general readers:

The training of deep neural networks (DNNs) via gradient-based optimization methods is nowadays a widely used procedure with various practical applications. However, it still remains a widely open problem to mathematically explain why these methods are so successful in practice.

In this article we consider fully connected feedforward DNNs with the rectified linear unit (ReLU) activation. We study solutions of gradient flow (GF) differential equations in the training of such DNNs with an arbitrarily large number of hidden layers and we prove that every non-divergent GF trajectory converges with a polynomial rate of convergence to a critical point. Our assumptions are that the distribution of the input data has a piecewise polynomial density and that the target function (describing the relationship between input data and output data) is piecewise polynomial. One of the main steps in our proof is to verify that the considered risk function satisfies a Kurdyka-Lojasiewicz inequality. A key difficulty in the analysis of the training of ReLU DNNs is the fact that the ReLU function is not differentiable, and thus we need to employ some tools from nonsmooth analysis.

Under the additional assumption that the considered risk function admits at least one regular global minimum, we also establish convergence of the risk of the gradient descent (GD) method with random initializations in the training of ReLU DNNs. Finally, in the special situation of shallow networks with one-dimensional input we verify this assumption by proving that for every Lipschitz continuous target function there exists a sufficiently regular global minimum in the risk landscape.

Sponsored by the Center for Machine Learning Research, Peking University & AI for Science Institute, Beijing.