On the universal consistency of an over-parametrized deep neural network estimate learned by gradient descent

Selina Drews · Michael Kohler

Received: 13 October 2022 / Revised: 11 September 2023 / Accepted: 28 December 2023 / Published online: 8 April 2024
© The Institute of Statistical Mathematics, Tokyo 2024

Abstract

Estimation of a multivariate regression function from independent and identically distributed data is considered. An estimate is defined which fits a deep neural network consisting of a large number of fully connected neural networks, which are computed in parallel, via gradient descent to the data. The estimate is over-parametrized in the sense that the number of its parameters is much larger than the sample size. It is shown that with a suitable random initialization of the network, a sufficiently small gradient descent step size, and a number of gradient descent steps that slightly exceed the reciprocal of this step size, the estimate is universally consistent. This means that the expected $L_2$ error converges to zero for all distributions of the data where the response variable is square integrable.

Keywords

Neural networks · Nonparametric regression · Over-parametrization · Universal consistency