Check for
updates

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

Selina Drews[1] · Michael Kohler[1]

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.

✉ Selina Drews
drews@mathematik.tu-darmstadt.de

Michael Kohler
kohler@mathematik.tu-darmstadt.de

[1] Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany