

Berlin Mathematics Research Center

MATH+

MATH+ Activity Group “Mathematics of Data Science”

Berlin Mathematical School

BMS

Thursday 9 January 2020 at 16:15

Zuse Institute Berlin, Lecture Hall, Takustr. 7, 14195 Berlin

Eric Vanden-Eijnden

(Courant Institute, NYU)

Trainability and accuracy of artificial neural networks

The methods and models of machine learning (ML) are rapidly becoming de facto tools for the analysis and interpretation of large data sets. Complex classification tasks such as speech and image recognition, automatic translation, decision making, etc. that were out of reach a decade ago are now routinely performed by computers with a high degree of reliability using (deep) neural networks. These performances suggest that it may be possible to represent high-dimensional functions with controllably small errors, potentially outperforming standard interpolation methods based e.g. on Galerkin truncation or finite elements. In support of this prospect, in this talk I will present results about the trainability and accuracy of neural networks, obtained by mapping the parameters of the network to a system of interacting particles relaxing on a potential determined by the loss function. Unlike the particles themselves, their empirical distribution evolves on a convex landscape. This observation can be used to prove a dynamical variant of the universal approximation theorem showing that the optimal neural network representation can be attained by (stochastic) gradient descent, with a approximation error scaling as the inverse of the network size. I will also show how these findings can be used to accelerate the training of networks and optimize their architecture, using e.g. nonlocal transport involving birth/death processes in parameter space. Finally I will discuss some implications of these results in the context of scientific computing.



© private