

OPTIMAL APPROXIMATION OF PIECEWISE SMOOTH FUNCTIONS USING DEEP RELU NEURAL NETWORKS

TALK BY F. VOIGTLAENDER IN THE AFG SEMINAR ON NOVEMBER 9, 2017

Abstract. Recently, machine learning techniques based on deep neural networks have significantly improved the state of the art in tasks like image classification and speech recognition. Nevertheless, a solid theoretical explanation of this success story is still missing.

In this talk, we will present recent results concerning the approximation-theoretic properties of deep neural networks which help to explain some of the characteristics of such networks; in particular we will see that deeper networks can approximate certain classification functions much more efficiently than shallow networks. We emphasize though that these approximation theoretic properties do not explain why simple algorithms like stochastic gradient descent work so well in practice, or why deep neural networks tend to generalize so well; we purely focus on the expressive power of such networks.

Precisely, as a model class for classifier functions we consider the class of (possibly discontinuous) piecewise smooth functions for which the different "smooth regions" are separated by smooth hypersurfaces. Given such a function, and a desired approximation accuracy, we construct a neural network which achieves the desired approximation accuracy, where the error is measured in L^2 . We give precise bounds on the required size (in terms of the number of weights) and depth of the network, depending on the approximation accuracy, on the smoothness parameters of the given function, and on the dimension of its domain of definition. Finally, we show that this size of the networks is optimal, and that networks of smaller depth would need significantly more weights than the deep networks that we construct, in order to achieve the desired approximation accuracy.