SMALLER GENERALIZATION ERROR DERIVED FOR A DEEP RESIDUAL NEURAL NETWORK COMPARED TO SHALLOW NETWORKS

Aku Kammonen[1], Jonas Kiessling[2, 3], Petr Plecháč[4], Mattias Sandberg[3], Anders Szepessy[3*], Raúl Tempone[1, 2]

[1]KAUST, Saudi Arabia
[2]RWTH, Aachen, Germany
[3]KTH, Stockholm, Sweden
[4]University of Delaware, Newark, DE 19716, USA

* szepessy@kth.se

Residual networks were introduced to improve the training of deep neural networks. Can they also be shown to be more accurate? In this talk I will present a theorem which shows that approximation of a function $f : \mathbb{R}^d \to \mathbb{R}$ by a residual neural network with L random Fourier features layers $\bar{z}_{\ell+1} = \bar{z}_\ell + \mathrm{Re} \sum_{k=1}^{K} \bar{b}_{\ell k} e^{\mathrm{i}\omega_{\ell k}\bar{z}_\ell} + \mathrm{Re} \sum_{k=1}^{K} \bar{c}_{\ell k} e^{\mathrm{i}\omega'_{\ell k}\cdot x}$ has smaller generalization error than the classical estimate $\|\hat{f}\|^2_{L^1(\mathbb{R}^d)}/(KL)$ of the generalization error for random Fourier features with one hidden layer and the same total number of nodes KL, in the case the L^∞-norm of f is much less than the L^1-norm of its Fourier transform $\hat{f}$. I will also present related numerical results.

References

[1] Adaptive random Fourier features with Metropolis sampling, by A. Kammonen, J. Kiessling, P. Plecháč, M. Sandberg, A. Szepessy. In on Foundations of Data Science, 2(3): 309--332, 2020.

[2] Smaller generalization error derived for a deep residual neural network compared to shallow networks, by Aku Kammonen and Jonas Kiessling and Petr Plecháč and Mattias Sandberg and Anders Szepessy and Raúl Tempone, arXiv, 2021, eprint 2010.01887