ADAPTIVE RANDOM FOURIER FEATURES BASED ON METROPOLIS SAMPLING

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

* akujaakko.kammonen@kaust.edu.sa

The supervised learning problem to approximate a function $f : \mathbb{R}^d \to \mathbb{R}$ by a neural network approximation $\mathbb{R}^d \ni x \mapsto \sum_{k=1}^{K} \hat{\beta}_k e^{\mathrm{i}\omega_k \cdot x}$ with one hidden layer is studied as a random Fourier features algorithm. Here the mean square loss problem can be solved easily, since it is convex in the amplitude parameters $\hat{\beta}_k$ given a density $p : \mathbb{R}^d \to [0, \infty)$ for independent frequencies ω_k. It is also well known that the corresponding generalization error is bounded by $K^{-1}\||\hat{f}|^2/p\|_{L^1(\mathbb{R}^d)}$, where $\hat{f}$ is the Fourier transform of f. In this talk I will first show how the constant $\||\hat{f}|^2/p\|_{L^1(\mathbb{R}^d)}$ can be minimized by optimally choosing the density p and then how to approximately sample from this density, only using the data and certain adaptive Metropolis steps. I will also present computational experiments including deep residual networks.

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