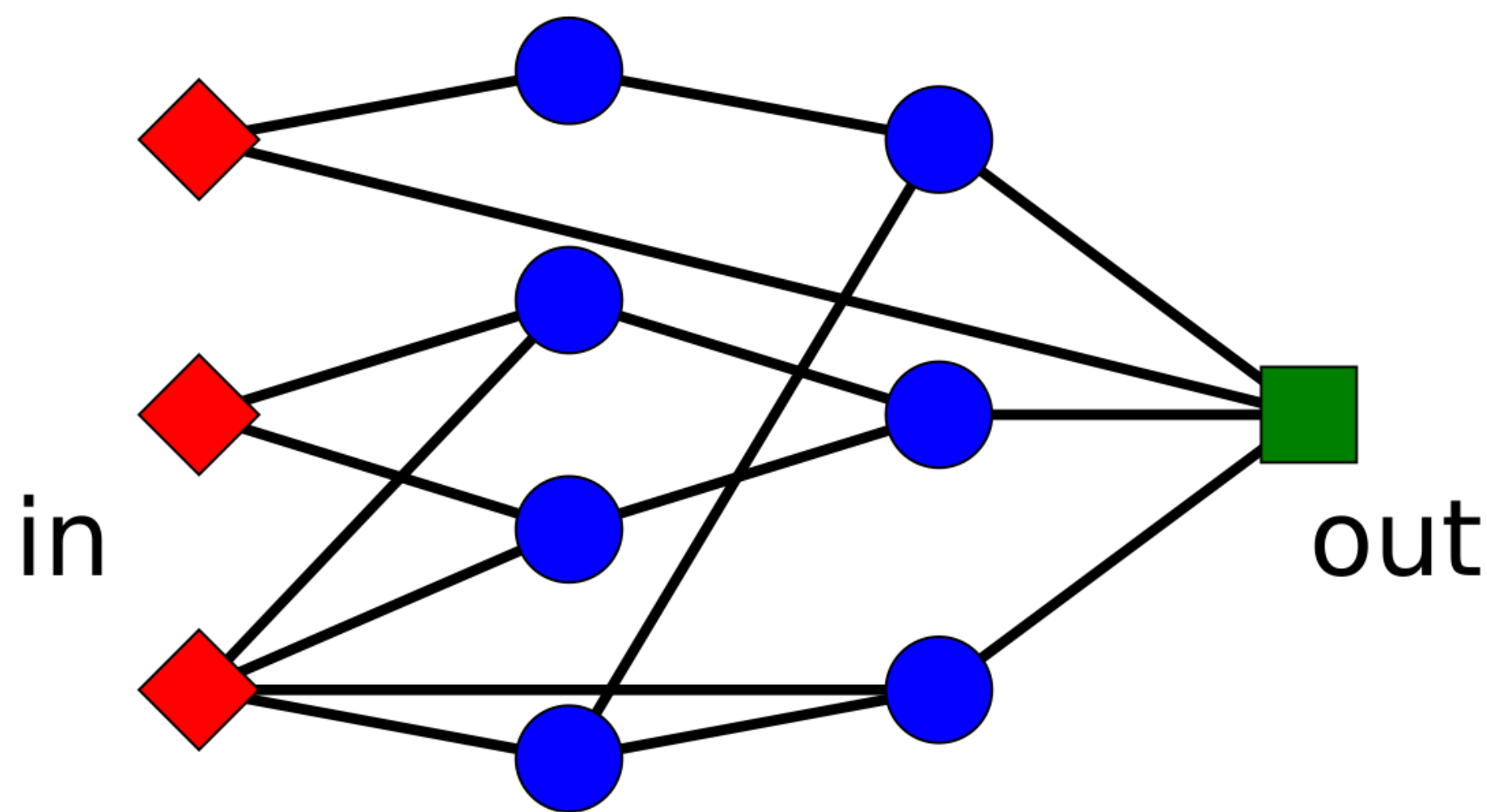


Approximation with neural networks of minimal size: exotic regimes and superexpressive activations

Speaker: Dmitry Yarotsky, Skoltech



Abstract

I will discuss some "exotic" regimes arising in theoretical studies of function approximation by neural networks of minimal size. The classical theory predicts specific power laws relating the model complexity to the approximation accuracy for functions of given smoothness, under the assumption of continuous parameter selection. It turns out that these power laws can break down if we use very deep narrow networks and don't impose the said assumption. This effect is observed for networks with common activation functions, e.g. ReLU. Moreover, there exist some "superexpressive" collections of activation functions that theoretically allow to approximate any continuous function with arbitrary accuracy using a network with a fixed number of neurons, i.e. only by suitably adjusting the weights without increasing the number of neurons. This result is closely connected to the Kolmogorov(-Arnold) Superposition Theorem. An example of superexpressive collection is $\{\sin, \arcsin\}$. At the same time, the commonly used activations are not superexpressive.



[Registration link](#)