Your 598 ID: 5107

Title of paper: Learning Functions: When is Deep Better than Shallow?

What is their primary result? The paper compares shallow (one hidden layer) networks with deep networks (in particular, the idealized model of a deep network as a binary tree) and shows that although both the shallow and deep neural networks are capable of achieving the same degree of accuracy the number of parameters, VC-dimension, and fat-shattering dimension are much smaller for the deep neural network.

Why is this important? The paper attempts to explain why deep neural networks, in general, perform better than shallow networks.

What are their key ideas? Their first result is the following: If the activation function is smooth and not a polynomial anywhere, then any function f in the Sobolev space $W^{r,p}(\mathbb{R}^d)$ (the author uses unorthodox notation for this) can be approximated with error $\mathrm{O}(n^{-r/d})$ by a deep neural network.

Their second collection of results involve Gaussian neural networks and is a bit more mathematically involved. In a nutshell, the authors show that a sufficiently deep Gaussian neural network achieves some stated degree of accuracy (see the paper to determine how the accuracy is being tested) if and only if the function being approximated is in some smoothness space (the smoothness space in question is not traditionally a Sobolev space; see the paper for the details).

Their last result is shows that the VC-dimension can be bounded, by their previous results. They also reference the work of Anthony—Bartlett, who show that the fat-shattering dimension is bounded above by the VC-dimension.

What are the limitations, either in performance or applicability? Most of the results in this paper require that the function the neural network is approximating be in some smoothness space.

Grader's 598 ID: