How Many Neurons Does it Take to Approximate the Maximum?

Abstract:

Understanding the role of depth is a fundamental endeavor in explaining the practical success of deep learning. In this talk, we will focus on the problem of approximating the maximum function over $d$ inputs using deep ReLU networks and with respect to the uniform distribution over a hypercube. We will show that while approximating the maximum using depth 2 networks to arbitrary accuracy requires arbitrary width, this can be done to arbitrary accuracy using a depth 3 network with a fixed width of $d^2/2$. Additionally, we will also show that this upper bound is tight, namely that width $\Omega(d^2)$ is also necessary when approximating using depth 3. Moreover, using this efficient depth 3 construction, we will show that greater depths result in a lesser width requirement, where width $O(d)$ suffices when we allow depth $O(\log(\log(d)))$. Lastly, we will show a size lower bound of $d$ neurons for approximating the maximum using any depth. These results establish a partial depth hierarchy for approximating a naturally occurring function which helps explain the benefits of depth over width.