**Universal Approximation Properties and Depth**

A feed-forward network with a single hidden layer containing a finite number of neurons can approximate continuous functions on compact subsets of ℝn, under mild assumptions on the activation function

* Simple neural networks can represent a wide variety of interesting functions when given appropriate parameters
* However, it does not touch upon the algorithmic learnability of those parameters

The universal approximation theorem means that regardless of what function we are trying to learn, we know that a large MLP will be able to represent this function. However, we are not guaranteed that the training algorithm will be able to learn that function. Even if the MLP is able to represent the function, learning can fail for two different reasons.

* Optimizing algorithms may not be able to find the value of the parameters that corresponds to the desired function.
* The training algorithm might choose wrong function due to over-fitting

The universal approximation theorem says that there exists a network large enough to achieve any degree of accuracy we desire, but the theorem does not say how large this network will be. provides some bounds on the size of a single-layer network needed to approximate a broad class of functions. Unfortunately, in the worst case, an exponential number of hidden units may be required. This is easiest to see in the binary case: the number of possible binary functions on vectors v∈{0,1}nv∈{0,1}n is and selecting one such function requires 2nbits, which will in general require O(2n) degrees of freedom.

A feedforward network with a single layer is sufficient to represent any function, But the layer may be infeasibly large and may fail to generalize correctly. Using deeper models can reduce no.of units required and reduce generalization error