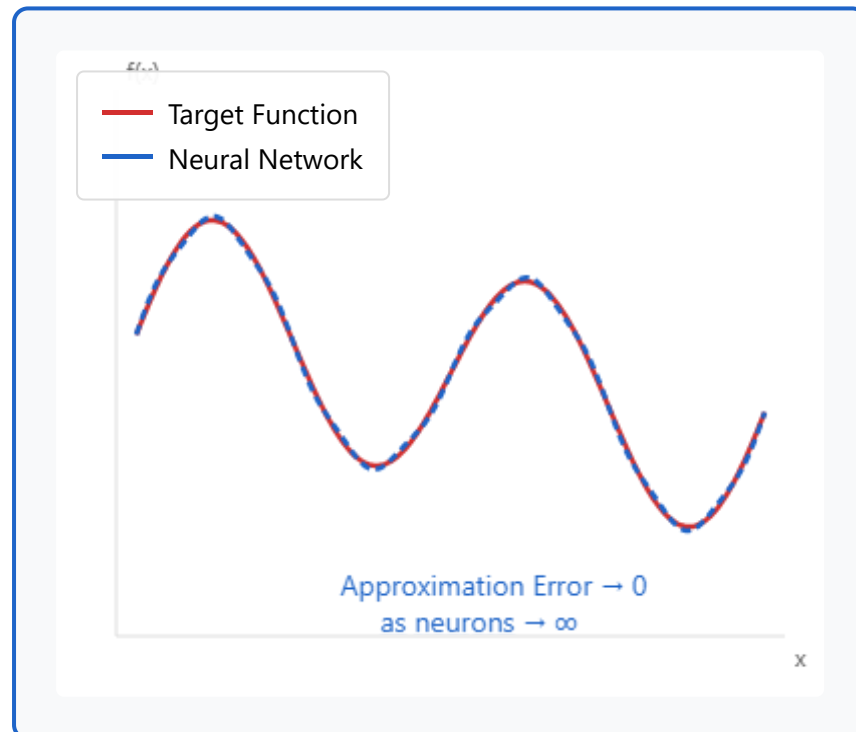


Universal Approximation Theorem

A neural network with a single hidden layer containing a finite number of neurons can approximate any continuous function on a compact subset of \mathbb{R}^n to arbitrary accuracy.

Function Approximation



Key Insights

1 Theoretical Power

Single hidden layer is theoretically sufficient for any continuous function

2 Width vs Depth

May need exponentially many neurons; deeper networks are more efficient

3 Practical Learning

Theorem doesn't guarantee efficient learning or generalization

Practical Implication

While theoretically powerful, we use deep networks in practice because they learn hierarchical features more efficiently and generalize better than wide shallow networks.