SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of AI &DS

Course Name – 19AD602 DEEP LEARNING

III Year / VI Semester

Unit 1-INTRODUCTION

Topic: Neural networks as universal function approximates.

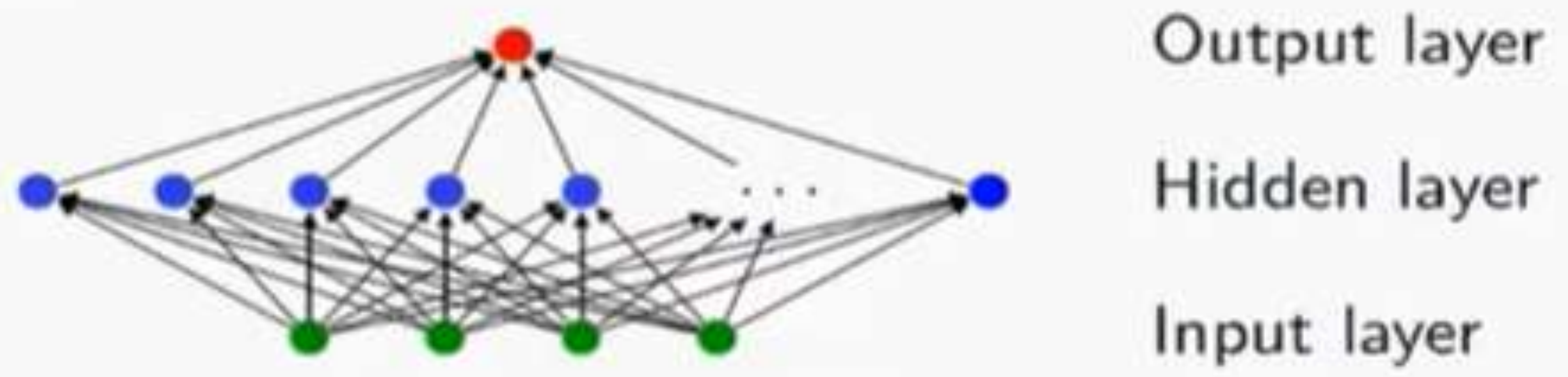
Case Study

A logistics company implemented neural networks to optimize delivery routes. Using a feedforward neural network, the company trained the model to approximate the optimal travel times for various routes, given factors like traffic patterns, weather, and time of day. The neural network accurately approximated these complex relationships, reducing delivery delays by 15% and operational costs by 10%.

Neural networks as universal function approximates.

- Classical universal approximation
  - Informal
  - Formal
- 'Transposed' universal approximation
  - Main result
  - Main trick
  - Sketch proof
- Applications
- Extensions
- Open questions

Neural networks as universal function approximates.



Wide, shallow networks, with nonpolynomial activation function, can approximate any function.

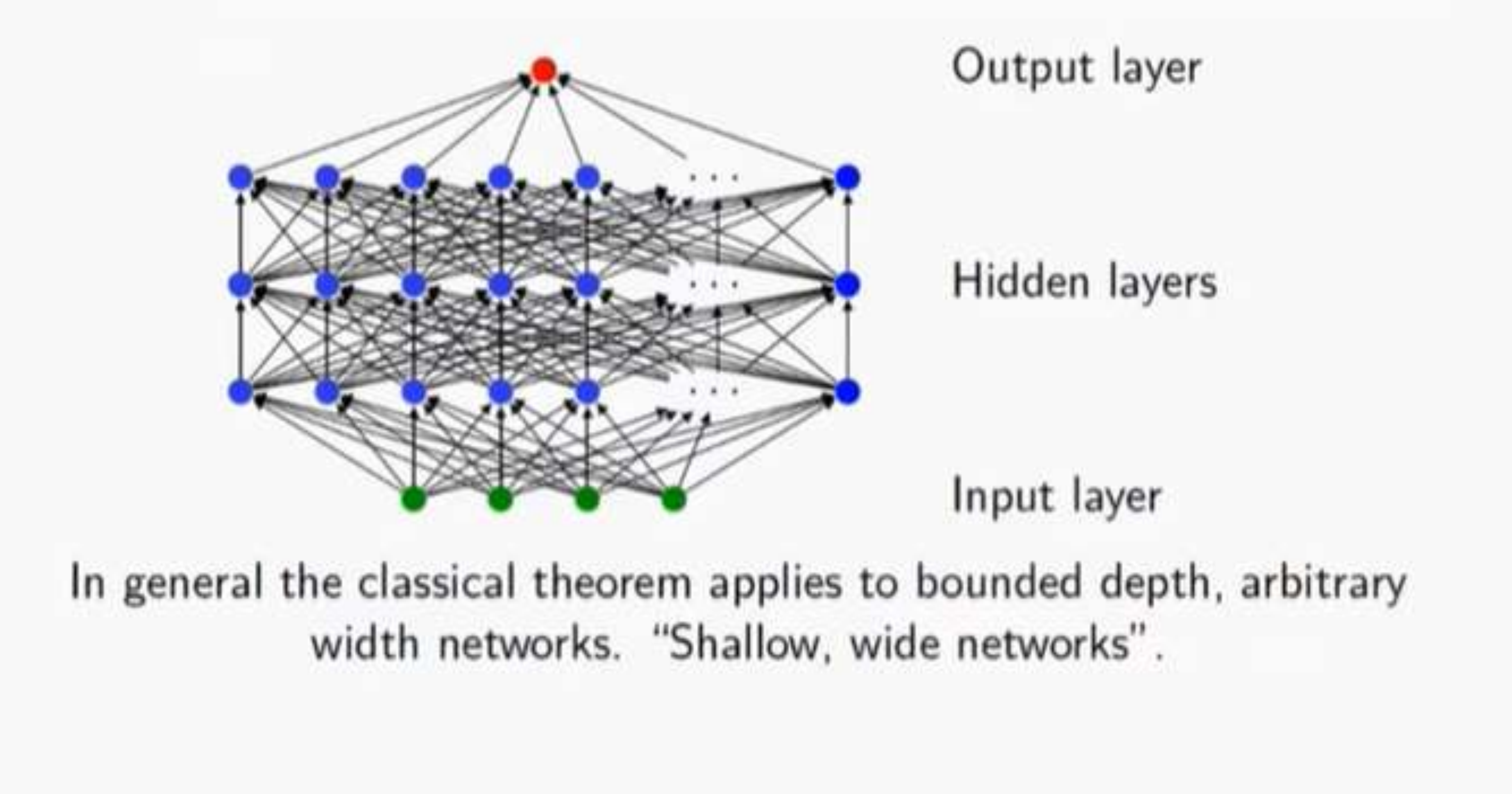
This is a foundational result on "why neural networks work".

Neural networks as universal function approximates.

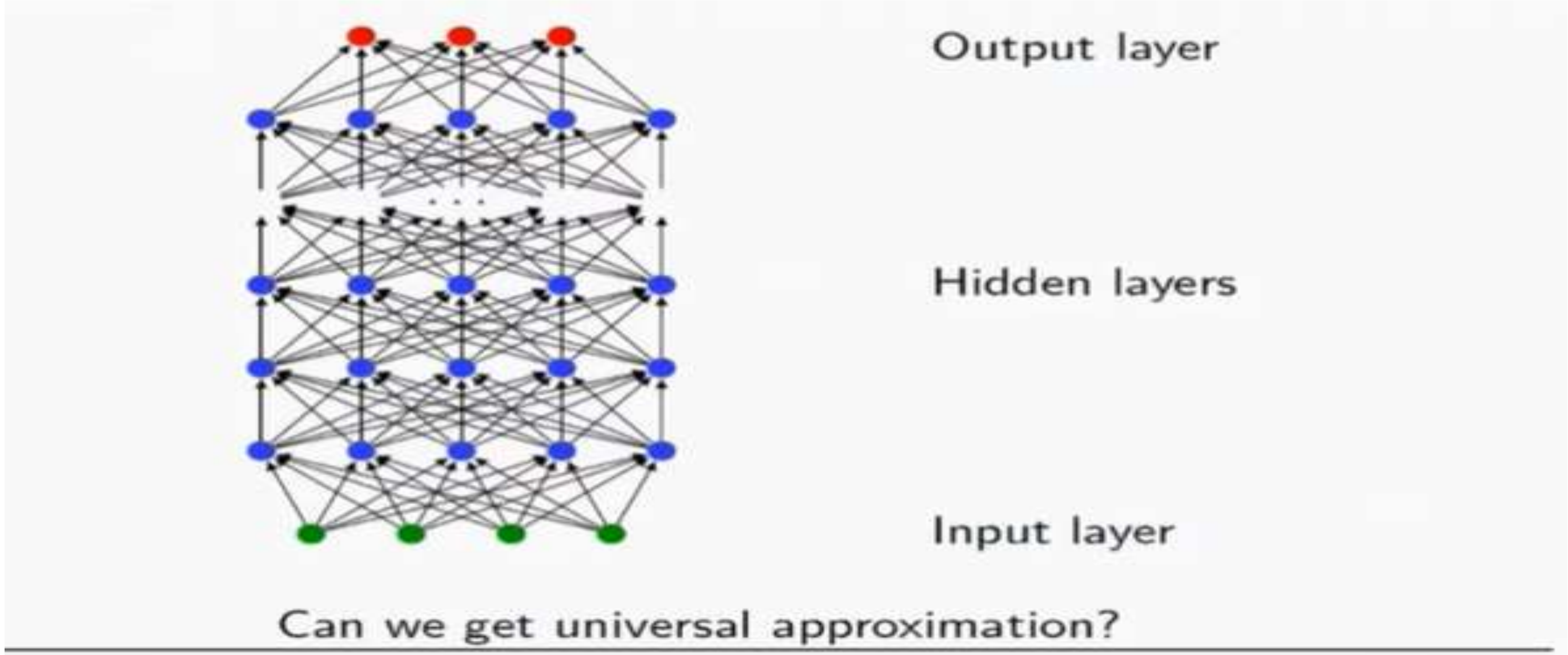
Theorem (Pinkus 1999)

Let $\rho: \mathbb{R} \to \mathbb{R}$ be any continuous function. Let $\mathcal{N}_n^\rho$ represent the class of feedforward neural networks with activation function ρ, with n neurons in the input layer, one neuron in the output layer, and one hidden layer with an arbitrary number of neurons. Let $K \subseteq \mathbb{R}^n$ be compact. Then $\mathcal{N}_n^\rho$ is dense in $C(K)$ if and only if ρ is nonpolynomial.

Neural networks as universal function approximates.



DEEP NARROW NETWORKS:



Neural networks as universal function approximates.

— What activation functions work?

— How narrow can the network be?

— Density w.r.t. which topology?

Neural networks as universal function approximates.

Definition

Let $\mathcal{NN}^{\rho}_{n,m,k}$ represent feedforward neural networks with:

—n input neurons,

—m output neurons,

—k neurons in each hidden layer (and an arbitrary number of hidden layers),

—ρ activation function.

Neural networks as universal function approximates.

Theorem (K., Lyons 2020)

Let $\rho: \mathbb{R} \to \mathbb{R}$ be any nonaffine continuous function which is continuously differentiable at at least one point, with nonzero derivative at that point. Let $K \subseteq \mathbb{R}^n$ be compact. Then $\mathcal{NN}^{\rho}_{n,m,n+m+2}$ is dense in $C(K;\mathbb{R}^m)$ with respect to the uniform norm.

"Deep, narrow networks, with essentially any activation function, can approximate any function."

Neural networks as universal function approximates.

The technical condition on the activation function – "continuously differentiable at at least one point with nonzero derivative at that point" – allows a composition of linear-activation-linear to approximate the identity.

Neural networks as universal function approximates.

Dots represent scalar values.

Arrows represent a linear map composed with an activation function.

Green arrows represent the identity function.

Orange arrows are arbitrary.

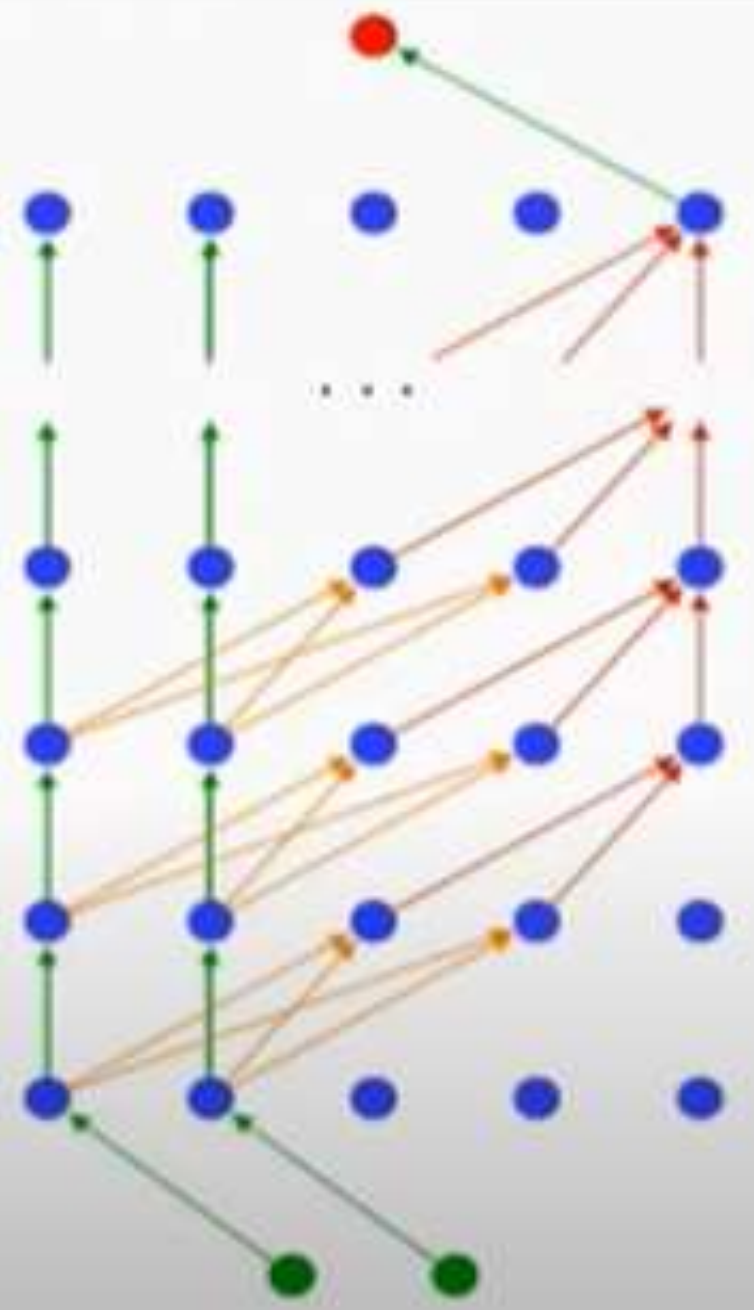
Red arrows are added to each other.

It's like a computer program!

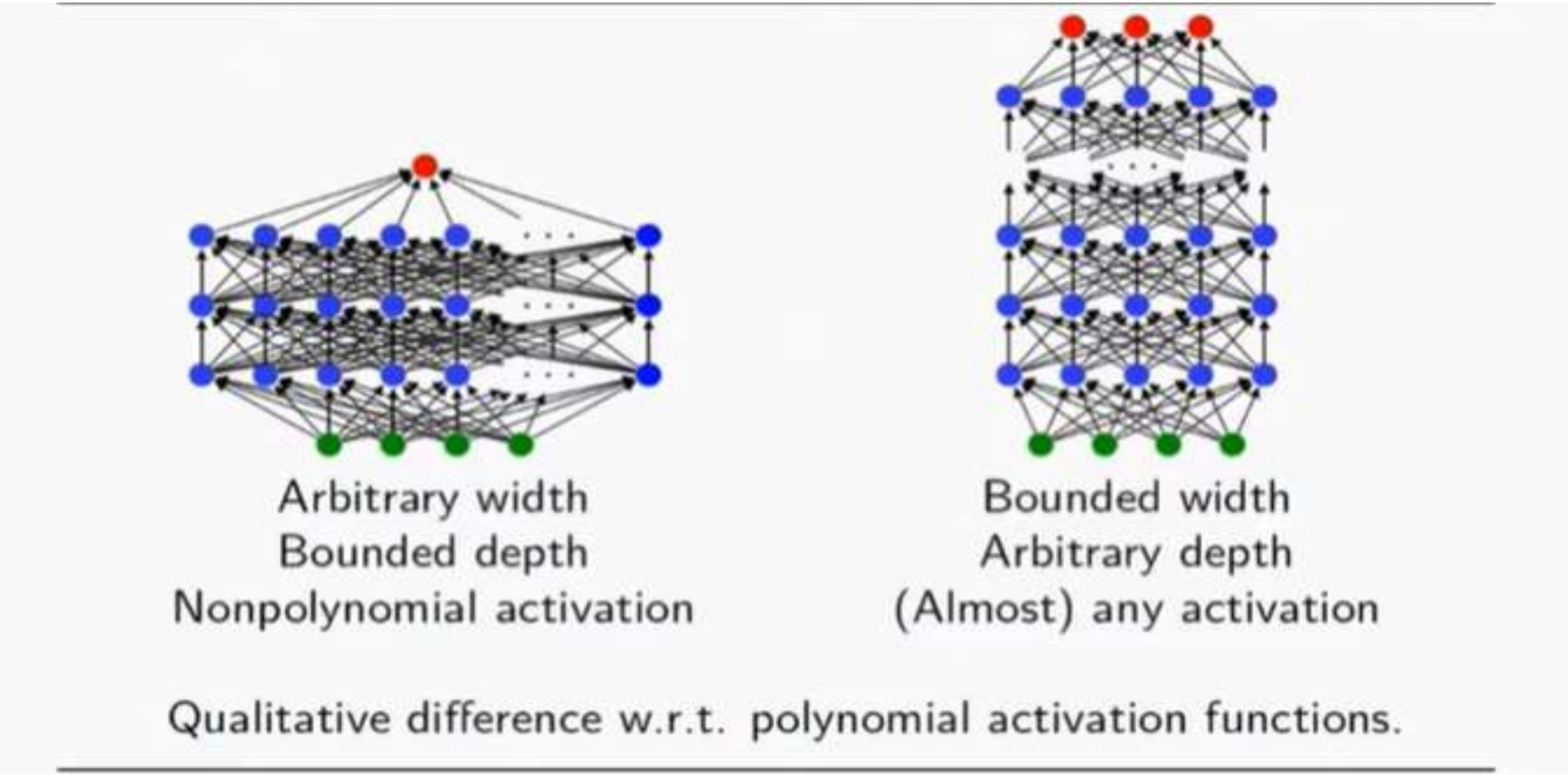
Sparsity

Constructive

Polynomial vs non-polynomial activation functions are handled differently.



Neural networks as universal function approximates.



Neural networks as universal function approximates.

— If proposing a new model, it's desirable to demonstrate universal approximation.

— This gives new ways to do that.

— Gives an understanding of bottlenecks.

— Handles classes of networks which increase in width and depth simultaneously. (The classical theorem doesn't.)

Neural networks as universal function approximates.

Proposition: Can show for "most" activation functions that $\mathcal{NN}^{\rho}_{n,m,n+m+1}$ exhibits universal approximation. (Main result: width $n + m + 2$.)

Proposition: Certain ρ non-differentiable $\implies$ can show $\mathcal{NN}^{\rho}_{n,m,n+m+1}$ exhibits universal approximation. (Main result: assumes a technical condition on the derivative.)

Theorem: Can show universal approximation in $L^p(\mathbb{R}^n; \mathbb{R}^m)$, for the ReLU. (Main result: in $C(K; \mathbb{R}^m)$.)

Neural networks as universal function approximates.

— Can the width be reduced further? (Yes for the ReLU, to $n + m$. (Hanin and Sellke 2017))

— What is the precise minimum width? (It's known that n is too small. (Johnson 2019))

— What about general non-differentiable activation functions?

— Can density in $L^p(\mathbb{R}^n; \mathbb{R}^m)$ be established for non-ReLU activations?

— Sobolev topologies?

Title: Exploring Neural Networks as Function Approximators

Objective: Train a neural network to approximate an unknown function based on data points.

Steps:

1. Generate or provide a dataset with inputs and corresponding outputs (e.g., sales data versus advertising spend).
2. Split the dataset into training and testing subsets.
3. Build and train a simple neural network to learn the pattern in the data.
4. Evaluate the network's ability to approximate the outputs by visualizing predictions against actual values.
5. Discuss how neural networks can generalize from the patterns in the training data to unseen data.

THANK YOU