Neural Network Architecture Beyond Width and Depth

Shijun Zhang

Duke University

(Joint work with Zuowei Shen and Haizhao Yang)

Motivation

- Deep neural networks have achieved great success in real-world applications.

Motivation

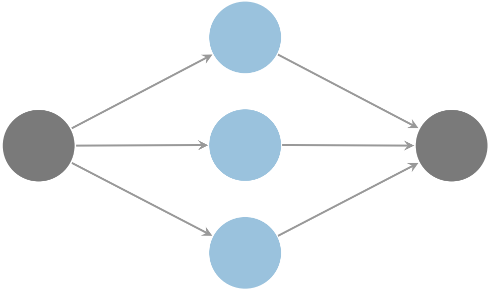
- Deep neural networks have achieved great success in real-world applications.
- Explosive growth of parameters and computation.

Motivation

- Deep neural networks have achieved great success in real-world applications.
- Explosive growth of parameters and computation.
- New network architecture via parameter sharing.

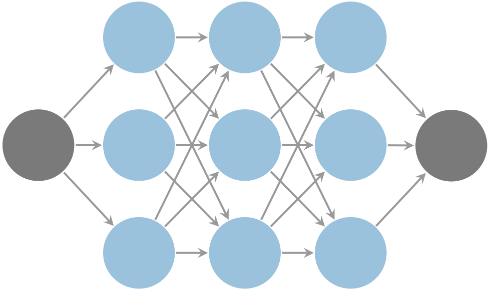
Standard networks

shallow network



width $= 3$, depth $= 1$.

deep network



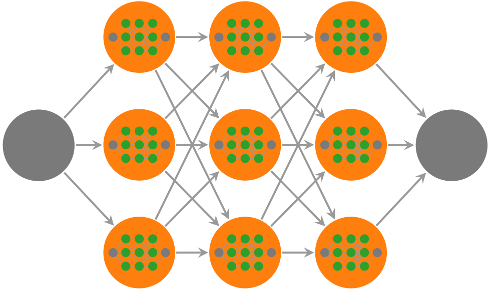
width $= 3$, depth $= 3$.

Concept of height

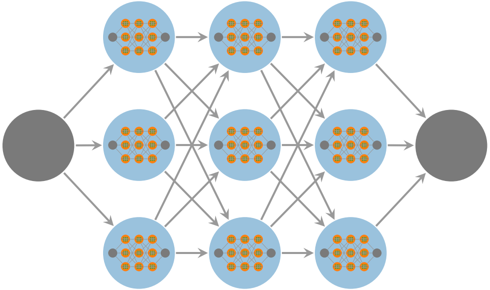
- Introduce **height** beyond width and depth.
- Share parameters via repetitions of activation functions.

Concept of height

- Introduce **height** beyond width and depth.
- Share parameters via repetitions of activation functions.



width $=$ depth $= 3$, height $= 2$.



width $=$ depth $=$ height $= 3$.

Concept of height

- Our new network architecture is recursively constructed with a **nested** structure and hence networks with the new architecture are called nested networks (**NestNets**).

Concept of height

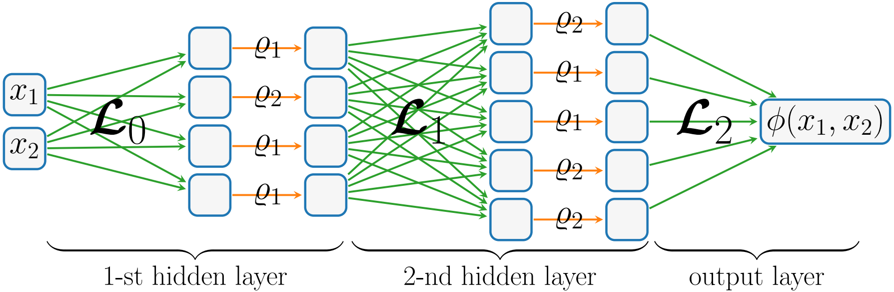
- Our new network architecture is recursively constructed with a **nested** structure and hence networks with the new architecture are called nested networks (**NestNets**).

- A NestNet of height s: each hidden neuron activated by a NestNet of height $\le s-1$.

Concept of height

- Our new network architecture is recursively constructed with a **nested** structure and hence networks with the new architecture are called nested networks (**NestNets**).
- A NestNet of height s: each hidden neuron activated by a NestNet of height $\leq s-1$.
- When $s=1$, a NestNet degenerates to a standard network.

Height-2 NestNet example



- $\mathcal{L}_0$, $\mathcal{L}_1$, and $\mathcal{L}_2$ are affine linear maps.
- ϱ_1 and ϱ_2 are height-1 networks (i.e., standard ones).
- #parameters $= \sum_{i=1}^{3} \#\mathcal{L}_i \, + \, \sum_{j=1}^{2} \#\varrho_j$.

Approximation of NestNets

Theorem

Given a 1-Lipschitz function f, $\forall\, n, s \in \mathbb{N}^+$, $\exists\, \phi$ realized by a height-s ReLU NestNet with $O(n)$ parameters s.t.

$$\big|\phi(\boldsymbol{x}) - f(\boldsymbol{x})\big| \le n^{-(s+1)/d} \quad \forall\, \boldsymbol{x} \in [0,1]^d.$$

Approximation of NestNets

Theorem

Given a 1-Lipschitz function f, $\forall\ n, s \in \mathbb{N}^+$, $\exists\ \phi$ realized by a height-s ReLU NestNet with $O(n)$ parameters s.t.

$$\left|\phi(\boldsymbol{x}) - f(\boldsymbol{x})\right| \le n^{-(s+1)/d} \quad \forall\ \boldsymbol{x} \in [0,1]^d.$$

- The constant in $O(\cdot)$ is explicitly given in our paper.

Approximation of NestNets

Theorem

Given a 1-Lipschitz function f, $\forall\ n, s \in \mathbb{N}^+$, $\exists\ \phi$ realized by a height-s ReLU NestNet with $O(n)$ parameters s.t.

$$\big|\phi(\boldsymbol{x}) - f(\boldsymbol{x})\big| \le n^{-(s+1)/d} \quad \forall\ \boldsymbol{x} \in [0,1]^d.$$

- The constant in $O(\cdot)$ is explicitly given in our paper.
- Increasing height $\Longrightarrow$ better approximation error.

Approximation of NestNets

Theorem

Given a 1-Lipschitz function f, $\forall\, n, s \in \mathbb{N}^+$, $\exists\, \phi$ realized by a height-s ReLU NestNet with $O(n)$ parameters s.t.

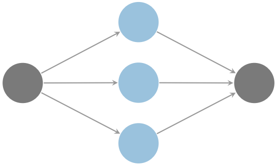
$$\big|\phi(\boldsymbol{x}) - f(\boldsymbol{x})\big| \le n^{-(s+1)/d} \quad \forall\, \boldsymbol{x} \in [0,1]^d.$$

- The constant in $O(\cdot)$ is explicitly given in our paper.
- Increasing height $\Longrightarrow$ better approximation error.
- 1-Lipschitz $\to C([0,1]^d)$, modulus of continuity.

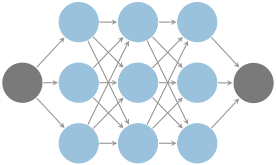
Error comparison

Table: Error comparison: various ReLU networks $\approx$ 1-Lipschitz functions on $[0,1]^d$.

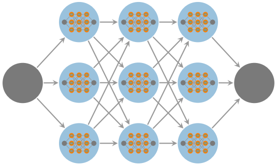
	#parameters	error	remark
shallow network	$O(n)$	$n^{-1/d}$ if $d=1$	linear combination
deep network	$O(n)$	$n^{-2/d}$	composition
height-s NestNet	$O(n)$	$n^{-(s+1)/d}$	nested composition



shallow network



deep network



NestNet

Thank you!

https://shijunzhang.top