Deep Network Approximation in Terms of Intrinsic Parameters

Shijun Zhang

National University of Singapore

(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.

- Network approximation: "important" parameters.

Preliminary

- $\mathscr{F}_{\mathsf{Lip}}$: 1-Lipschitz functions on $[0,1]^d$.

Preliminary

- $\mathscr{F}_{\mathsf{Lip}}$: 1-Lipschitz functions on $[0,1]^d$.

- A ReLU network approximates $f \in \mathscr{F}_{\mathsf{Lip}}$:
  optimal error $\mathcal{O}(n^{-2/d})$ $(n = \#\text{all parameters})$.

6/14

Preliminary

- $\mathscr{F}_{\mathsf{Lip}}$: 1-Lipschitz functions on $[0,1]^d$.

- A ReLU network approximates $f \in \mathscr{F}_{\mathsf{Lip}}$:
  optimal error $\mathcal{O}(n^{-2/d})$ ($n = \#$all parameters).

- Intrinsic parameters:
  those depending on the target function f.

Main results

For $f \in \mathscr{F}_{\mathsf{Lip}}$ and $p \in [1, \infty)$, $\exists\, \phi$ realized by a ReLU network with $n + 2$ intrinsic parameters s.t.

$$\left\| \phi - f \right\|_{L^p([0,1]^d)} \leq 5\sqrt{d}\, 2^{-n}$$

Main results

For $f \in \mathscr{F}_{\mathsf{Lip}}$ and $p \in [1, \infty)$, $\exists \, \phi$ realized by a ReLU network with $n + 2$ intrinsic parameters s.t.

$$\left\| \phi - f \right\|_{L^p([0,1]^d)} \leq 5\sqrt{d}\, 2^{-n}$$

- $\mathscr{F}_{\mathsf{Lip}} \to C([0,1]^d)$, modulus of continuity.

- $L^p \to L^\infty$, additional $\mathcal{O}(n)$ intrinsic parameters.

Main results

$\forall\,\varepsilon > 0$, $\exists\,\phi$ realized by a ReLU network s.t. $\forall\,f \in \mathscr{F}_{\mathsf{Lip}}$, $\exists\,s, v, b \in \mathbb{R}$ satisfying

$$\big|s\phi(\boldsymbol{x}, v) + b - f(\boldsymbol{x})\big| \leq \varepsilon \quad \forall\,\boldsymbol{x} \in [0, 1]^d.$$

Main results

$\forall \, \varepsilon > 0, \, \exists \, \phi$ realized by a ReLU network s.t. $\forall \, f \in \mathscr{F}_{\mathsf{Lip}}$, $\exists \, s, v, b \in \mathbb{R}$ satisfying

$$\big| s\phi(\boldsymbol{x}, v) + b - f(\boldsymbol{x}) \big| \leq \varepsilon \quad \forall \, \boldsymbol{x} \in [0,1]^d.$$

- ϕ is independent of f.

- Three intrinsic parameters are enough.

- High precision is required for v.

Connection to transfer learning

- Transfer learning:

    Apply knowledge gained in a related problem.

Connection to transfer learning

- Transfer learning:

    Apply knowledge gained in a related problem.

- Non-intrinsic parameters:

    Reused and transferred from another target function.

Thank you!

https://shijunzhang.top