A Functional Perspective on Learning Symmetric Functions with Neural Networks

Aaron Zweig[1] Joan Bruna[1,2]

[1]Courant Institute, NYU

[2]Center for Data Science, NYU

Symmetric Neural Networks

- Fix N and let $\boldsymbol{x} = \{x_1, \ldots, x_N\}$
- The DeepSets[1] architecture:

$$f_N(\boldsymbol{x}) = \rho\left(\frac{1}{N}\sum_{n=1}^{N}\Phi(x_n)\right)$$

- Universal approximator for continuous symmetric functions *for fixed N*

Main Question

How to model f_N across varying values of N?

[1] Manzil Zaheer et al. "Deep sets". In: *Advances in neural information processing systems*. 2017, pp. 3391–3401.

DeepSets for probability measures

- Given $\{x_1, \dots, x_N\}$, form the discrete measure $\mu^{(N)} = \frac{1}{N}\sum_{n=1}^{N} \delta_{x_n}$

$$\rho\left(\frac{1}{N}\sum_{n=1}^{N}\Phi(x_n)\right) = \rho\left(\langle \Phi, \mu^{(N)}\rangle\right) = \int_{\mathcal{A}} \sigma(\langle \phi, \mu^{(N)}\rangle)\chi(d\phi)$$

- DeepSets $\iff$ NN that takes $\mu^{(N)}$ as input
- Here χ is a measure over some function space $\mathcal{A}$
- Can we generalize beyond discrete measures $\mu^{(N)}$ to arbitrary μ?

Convex Neural Networks vs. Measure Neural Networks

Convex Neural Network[2]

$$\phi(x) = \int_{\mathbb{S}^d} \sigma(\langle w, \tilde{x} \rangle)\nu(dw)$$

Input: $x \in \mathbb{I} \subset \mathbb{R}^n$
Weights: $w \in \mathbb{R}^{n+1}$
Neurons: $\nu \in \mathcal{M}(\mathbb{R}^{n+1})$

Universal Approximation[3]

Measure Neural Network

$$f(\mu) = \int_{\mathcal{A}} \sigma(\langle \phi, \mu \rangle)\chi(d\phi)$$

Input: $\mu \in \mathcal{P}(\mathbb{I})$
Weights: $\phi \in \mathcal{A}$
Neurons: $\chi \in \mathcal{M}(\mathcal{A})$

Universal Approximation[4] [5]

[2] Francis Bach. "Breaking the curse of dimensionality with convex neural networks". In: *The Journal of Machine Learning Research* 18.1 (2017), pp. 629–681.

[3] George Cybenko. "Approximation by superpositions of a sigmoidal function". In: *Mathematics of control, signals and systems* 2.4 (1989), pp. 303–314.

[4] Gwendoline De Bie, Gabriel Peyré, and Marco Cuturi. "Stochastic deep networks". In: *International Conference on Machine Learning*. 2019, pp. 1556–1565.

[5] Tomas Pevny and Vojtech Kovarik. "Approximation capability of neural networks on spaces of probability measures and tree-structured domains". In: *arXiv preprint arXiv:1906.00764* (2019).

Main Questions

Question

How do we use measure neural networks to learn symmetric families f_N across different N?

Main Questions

Question

How do we use measure neural networks to learn symmetric families f_N across different N?

Proposition (informal)

There exists a continuous extension $\bar{f}$ to probability measures iff the incomplete extension to discrete measures $\hat{f}$ is uniformly continuous with regard to the W_1 metric on its domain.

Main Questions

Question

How do we use measure neural networks to learn symmetric families f_N across different N?

Proposition (informal)

There exists a continuous extension $\bar{f}$ to probability measures iff the incomplete extension to discrete measures $\hat{f}$ is uniformly continuous with regard to the W_1 metric on its domain.

Question

What functional spaces do measure neural networks yield? How does learning work in these spaces?

Functional Spaces of Measure Networks

$$f(\mu) = \int_{\mathcal{A}} \widetilde{\sigma}(\langle \phi, \mu \rangle) \chi(d\phi)$$

Functional spaces determined by restrictions on possible χ:

- Choose if χ is a density w.r.t. some base measure
- Choose the support of χ

Functional Spaces of Measure Networks

$$f(\mu) = \int_{\mathcal{A}} \widetilde{\sigma}(\langle \phi, \mu \rangle) \chi(d\phi)$$

Functional spaces determined by restrictions on possible χ:

- Choose if χ is a density w.r.t. some base measure
- Choose the support of χ

	χ	$Supp(\chi)$	Associated Norm
$\mathcal{S}_1$	Arbitrary	bounded Barron norm	$TV(\chi)$
$\mathcal{S}_2$	Arbitrary	bounded RKHS norm	$TV(\chi)$
$\mathcal{S}_3$	Density	bounded RKHS norm	$\lVert p \rVert_{L_2(\tau)}$ where $p(\phi)\tau(d\phi) = \chi(d\phi)$

Approximation of $\mathcal{S}_i$

Theorem (informal)

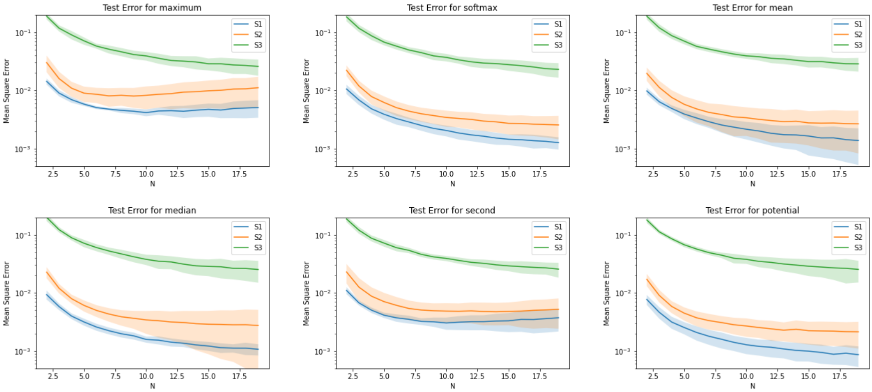
For appropriate choices of the kernel base measures κ and τ, there exist f_1 with $\|f_1\|_{\mathcal{S}_1} \leq 1$ and f_2 with $\|f_2\|_{\mathcal{S}_2} \leq 1$ such that:

$$\inf_{\|f\|_{\mathcal{S}_3} \leq \delta} \|f - f_2\|_\infty \gtrsim d^{-2} \delta^{-5/d} ,$$

$$\inf_{\|f\|_{\mathcal{S}_2} \leq \delta} \|f - f_1\|_\infty \gtrsim |d^{-11} - d^{-d/3}\delta| .$$

- Punchline: $\mathcal{S}_3 \subsetneq \mathcal{S}_2 \subsetneq \mathcal{S}_1$
- Separation is cursed by dimension, but polynomially goes to 0
- Careful analysis of spherical harmonics and parity arguments to map from signed measure to probability measure inputs, requires using the square ReLU as the test function activation

Synthetic Experiments



Synthetic Experiments