On the Universality of Invariant Networks

Haggai Maron[1], **Ethan Fetaya** [2,3], **Nimrod Segol**[1] **& Yaron Lipman**[1]

[1]Weizmann Institute of Science [2]University of Toronto [3]Vector Institute

Introduction

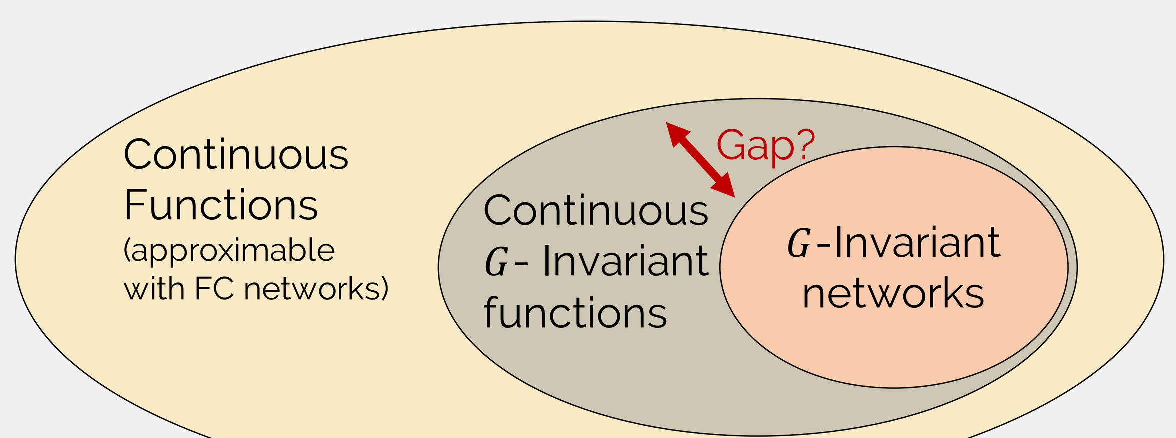
The basic paradigm of deep neural networks is repeatedly composing linear layers interlaced with non-linear, element-wise activation functions to create effective predictive models.

When trying to learn a function (task) f that is known to be invariant to some group of symmetries G (i.e., G-invariant function) it is common to use linear layers that respect this symmetry, namely, invariant and/or equivariant linear layers. Networks with invariant/equivariant linear layers with respect to some group G will be referred here as *G-invariant networks*, e.g., Figure 1.

Goal

When using these networks one important question arises: Can a G-invariant network approximate *an arbitrary* continuous G-invariant function?

The goal of this paper is to address this question for *all* finite permutation groups $G \leq S_n$, where S_n is the symmetric group acting on $[n] = \{1, 2, \dots, n\}$.



Examples

- The archetypal examples of G-invariant networks are Convolutional Neural Networks (CNNs) [3] that are translation equivariant.
- Tasks involving point clouds or sets are in general S_n-invariant [5, 6].
- Learning tasks involving interaction between different sets, where the input data is tabular, requires dealing with different permutations acting independently on each set [1].
- Tasks involving graphs and hyper-graphs lead to symmetries defined by tensor products of permutations [2, 4].

Contributions

1. We prove: any continuous function invariant to an arbitrary permutation subgroup $G \leq S_n$ can be approximated on a compact set to an arbitrary precision using a G-invariant network, using high-order tensors.
2. We prove a lower bound on the order d of tensors used in a G-invariant network so to achieve universality.
3. We prove a necessary condition on groups $G \leq S_n$ so that G-invariant networks, using only first order tensors, are universal. First order tensors are interesting since they lead to computationally tractable algorithms.

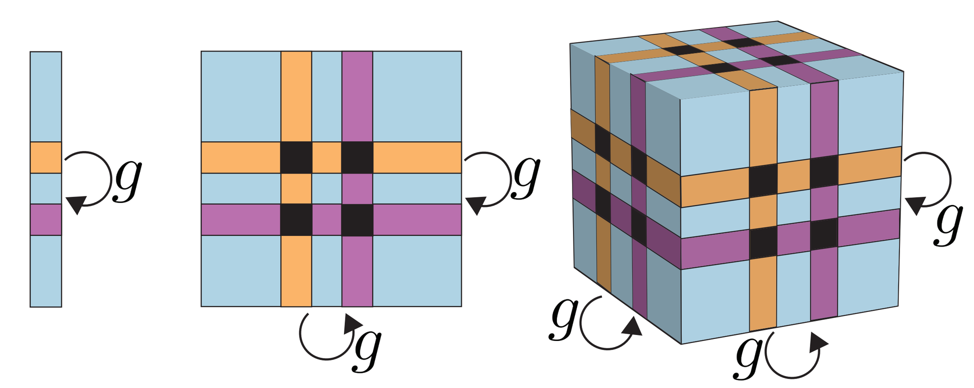
Permutation action

In this paper we consider the case of symmetry defined by any subgroup of the symmetric group, i.e., $G \leq S_n$. The action of G on $x \in \mathbb{R}^n$ used in this paper is defined as

$$g \cdot x = (x_{g^{-1}(1)}, \dots, x_{g^{-1}(n)}),\ g \in G. \tag{1}$$

The action of G on *tensors* $\mathsf{X} \in \mathbb{R}^{n^k \times a}$ (the last index, denoted j represents feature depth) is defined similarly by

$$(g \cdot \mathsf{X})_{i_1 \dots i_k, j} = \mathsf{X}_{g^{-1}(i_1) \dots g^{-1}(i_k), j},\ g \in G. \tag{2}$$



The Figure illustrates the action of g on tensors of order $k = 1, 2, 3$, where the permutation g is a transposition of two numbers and is applied to each dimension of the tensor.

Invariant networks architecture

A **G-invariant function** is a function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies

$$f(g \cdot x) = f(x)$$

for all $x \in \mathbb{R}^n$ and $g \in G$.

A **linear G-equivariant layer** is an affine map $L : \mathbb{R}^{n^k \times a} \to \mathbb{R}^{n^l \times b}$ satisfying

$$L(g \cdot \mathsf{X}) = g \cdot L(\mathsf{X})$$

for all $g \in G$, and $\mathsf{X} \in \mathbb{R}^{n^k \times a}$.

A **linear G-invariant layer** is an affine map $h : \mathbb{R}^{n^k \times a} \to \mathbb{R}^b$ satisfying

$$h(g \cdot \mathsf{X}) = h(\mathsf{X})$$

for all $g \in G$, and $\mathsf{X} \in \mathbb{R}^{n^k \times a}$.

A **G-invariant network** is a function $F : \mathbb{R}^{n \times a} \to \mathbb{R}$ defined by

$$F = m \circ h \circ L_d \circ \sigma \circ \cdots \circ \sigma \circ L_1,$$

where L_i are linear G-equivariant layers, σ is an activation function, h is a G-invariant layer, and m is a Multi-Layer Perceptron (MLP).

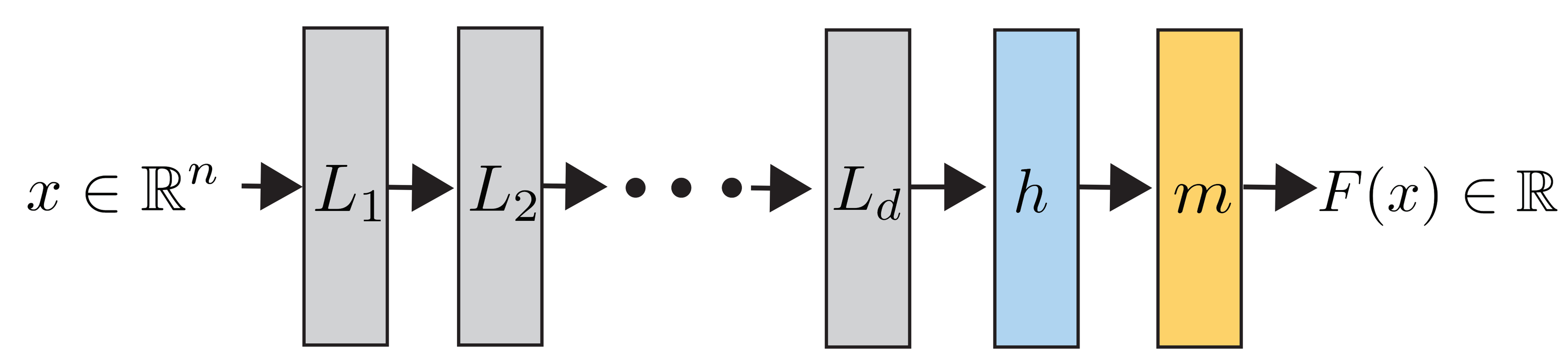


Figure 1. G-invariant network model. By construction, G-invariant networks are G-invariant functions.

Universality of invariant networks

Theorem 1
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous G-invariant function for some $G \leq S_n$, and $K \subset \mathbb{R}^n$ a compact set. There exists a G-invariant network that approximates f to an arbitrary precision on K.

The proof of Theorem 1 is constructive: given a G-invariant function f, it builds an f-approximating G-invariant network with hidden tensors $\mathsf{X} \in \mathbb{R}^{n^d}$ of order d, where $d = d(G)$ is a natural number depending on the group G. It can be shown that $d \leq \frac{n(n-1)}{2}$ for all $G \leq S_n$.

The main idea is to approximate a finite generating set of the ring of G-invariant polynomials, which can be shown to be dense in the space of continuous G-invariant functions on a compact domain.

A lower bound on equivariant layer order

We note that even using tensors $\mathsf{X} \in \mathbb{R}^{n^d}$ with order $d = 2$ could already be computationally challenging. We therefore ask whether we can use G-invariant networks with lower order tensors without sacrificing approximation power.

We show this is generally not the case; for some subgroups a minimal order of $\frac{n-2}{2}$ is needed. Specifically, we prove the following for $G = A_n$, the alternating group:

Theorem 2
If an A_n-invariant network has the universal approximation property then it consists of tensors of order at least $\frac{n-2}{2}$.

The basic idea behind this proof is that both S_n and A_n give rise to the **same** linear equivariant layers for low order tensors. This implies that any A_n-invariant network is also S_n-invariant. We use this fact in order to show that an A_n-invariant network that uses only low order tensors cannot approximate the Vandermonde polynomial $V(x) = \Pi_{1 \leq i < j \leq n}(x_i - x_j)$ (which is A_n-invariant but not S_n-invariant).

Universality of first order networks

Although in general we cannot expect universal approximation of G-invariant networks with inner tensor order smaller than $\frac{n-2}{2}$, it is still possible that for *specific* groups of interest we can prove approximation power with more efficient (i.e., lower order inner tensors) G-invariant networks. Of specific interest are G-invariant networks that use only first order tensors.

Theorem 3
Let $G \leq S_n$. If first order G-invariant networks are universal, then $|[n]^2/H| < |[n]^2/G|$ for any strict super-group $G < H \leq S_n$.

$|[n]^2/G|$ is the number of equivalence classes of $[n]^2$ defined by the relation: $(i_1, i_2) \sim (j_1, j_2)$ if $j_\ell = g(i_\ell)$, $\ell = 1, 2$ for some $g \in G$. Intuitively, this condition asks that super-groups of G have *strictly* better separation of the double index space $[n]^2$.

References

[1] Jason S. Hartford, Devon R. Graham, Kevin Leyton-Brown, and Siamak Ravanbakhsh. Deep models of interactions across sets. In *ICML*, 2018.

[2] Risi Kondor, Hy Truong Son, Horace Pan, Brandon Anderson, and Shubhendu Trivedi. Covariant compositional networks for learning graphs. *arXiv preprint arXiv:1801.02144*, 2018.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[4] Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. In *International Conference on Learning Representations*, 2019.

[5] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 1(2):4, 2017.

[6] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in Neural Information Processing Systems*, pages 3391–3401, 2017.