A Characterization Theorem for Equivariant Networks with Point-wise Activations

Marco Pacini, Xiaowen Dong, Bruno Lepri, Gabriele Santin

Fondazione Bruno Kessler & University of Trento

MobS MOBILE AND SOCIAL COMPUTING LAB

Basics of Geometric Deep Learning

How can we build **symmetry sensitive neural networks**? We can build symmetry sensitive linearities and activations. We focus on the latter.

Rotations

Let R_α a rotation of $\mathbb{R}^2$ by an angle $\alpha \in [0, 2\pi)$. It can be represented as the matrix

$$R_\alpha = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

Permutations

Encoding the set of 3 elements $\{1, 2, 3\}$ like

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \in \mathbb{R}^3.$$

Permutations of those elements are represented by matrices like

$$P_{id} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P_{(12)} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P_{(123)} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Equivariance

A feature map $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is **equivariant** with respect to the symmetries P if

$$P \circ \Phi = \Phi \circ P.$$

Equivariant Neural Network

An **equivariant neural network** is a composition

$$\Phi = \phi_m \circ \tilde{f}_{m-1} \circ \phi_{m-1} \circ \cdots \circ \tilde{f}_1 \circ \phi_0,$$

where each *activation* $\tilde{f}_i : \mathbb{R}^n \to \mathbb{R}^n$ is a equivariant function, and

$$\phi_i(x) = Ax + b$$

for $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$ such that ϕ_i is equivariant.

Point-wise Activations

An activation $\tilde{f} : \mathbb{R}^n \to \mathbb{R}^n$ is **point-wise** if there is a real scalar function $f : \mathbb{R} \to \mathbb{R}$ such that

$$\tilde{f}(x_1, \dots, x_n) = (f(x_1), \dots, f(x_n))^t$$

Breaking Symmetry

- ReLU, as many other point-wise activations, is not equivariant with respect to some symmetries such as rotations

$$ReLU \circ R_{\frac{\pi}{2}}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = ReLU\left(\begin{bmatrix} -1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\neq$$

$$R_{\frac{\pi}{2}} \circ ReLU\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = R_{\frac{\pi}{2}}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

- ReLU, as many other point-wise activations, is equivariant with respect to some symmetries such as permutation

A Natural Question Comes to Mind

Which are combinations of **symmetries** and **activation functions** that lead to an **equivariant layer**?

Preliminaries – Activation Functions

- The **b-multiplicative functions:** $f(b^n x) = b^n f(x)$ for each $n \in \mathbb{Z}$ and for each $x \in \mathbb{R}$,
- The **$\pm b$-multiplicative functions:** $f(\pm b^n x) = \pm b^n f(x)$ for each $n \in \mathbb{Z}$ and for each $x \in \mathbb{R}$,
- **Odd functions:** $f(-x) = -f(x)$ for each $x \in \mathbb{R}$,
- **Semilinear functions:** linear on $\mathbb{R}_{>0}$ and on $\mathbb{R}_{<0}$.

Preliminaries – Symmetries

Permutation matrices,**signed permutation matrices**, **b-monomial matrices**, and **$\pm b$-monomial matrices**

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad S = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \quad M_b = \begin{bmatrix} 0 & b & 0 \\ 0 & 0 & b^2 \\ \frac{1}{b^5} & 0 & 0 \end{bmatrix}, \quad M_{\pm b} = \begin{bmatrix} 0 & -\frac{1}{b} & 0 \\ 0 & 0 & b \\ -b & 0 & 0 \end{bmatrix}.$$

The Characterization Theorem

Theorem: Assume activation functions are **not affine** and **continuous**. The following are the only possible combinations of activation functions and symmetries

1. **Continuous functions** and **permutation matrices**,
2. **Odd functions** and **signed permutation matrices**,
3. **Semilinear functions** and **non-negative monomial matrices**,
4. **Continuous b-multiplicative functions** and **b-monomial matrices**,
5. **Continuous $\pm b$-multiplicative functions** and **$\pm b$-monomial matrices**.

Adjacency Matrices and Graph Isomorphism

$$A_{G_1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_{G_2} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

We indicate the space of $n \times n$ matrices as $(\mathbb{R}^n)^{\otimes 2}$ and k-order tensors as $(\mathbb{R}^n)^{\otimes k}$.

The Linear Algebra of Node Permutations

A permutation of nodes induces isomorphic graphs and acts linearly on permutation matrices by conjugation

$$A_{G_2} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = P^t_{(12)} \cdot A_{G_1} \cdot P_{(12)}.$$

Invariant Graph Networks (IGNs)

IGNs are **permutation equivariant neural networks** defined as

$$\mathcal{N} : (\mathbb{R}^n)^{\otimes 2} \otimes \mathbb{R}^f \to \mathcal{Y}$$

where $\mathbb{R}^f$ is a **feature space**. This means

$$\mathcal{N}(P^t_\sigma A P_\sigma, F) = \sigma \mathcal{N}(A, F)$$

where σ is a permutation of the nodes, also acting on $\mathcal{Y}$.

Geometric Relational Structures & IGNs

Geometric graphs or **higher-order structures** are employed in computer graphs, computational biology, and computational chemistry. They can be encoded in a vector divided in a relational part and a geometric part:

$$(\mathbb{R}^n)^{\otimes 2} \otimes \mathbb{R}^3$$

Geometric IGNs are rotation-equivariant IGNS defined as

$$\mathcal{N} : (\mathbb{R}^n)^{\otimes 2} \otimes \mathbb{R}^3 \to \mathcal{Y}$$

A Non-existence Result

Corollary: Every Geometric IGN coupled with non-affine activations is **null**

References

[1] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*.

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

[3] Marco Pacini, Xiaowen Dong, Bruno Lepri, and Gabriele Santin. A characterization theorem for equivariant networks with point-wise activations. *To Appear*.

[4] Jeffrey Wood and John Shawe-Taylor. Representation theory and invariant neural networks. *Discrete applied mathematics*.

mpacini@fbk.eu