FLATIRON INSTITUTE
Center for Computational Mathematics

Invariant and Equivariant Neural Networks

A brief literature review and potential applications to inverse problems

Eftychios Pnevmatikakis
February 28, 2019

CCM - Machine Learning Working Group

Motivation

- Deep neural networks can learn invariances (e.g., translation, scale, rotations) given enough data.
- In unconstrained architectures this can be achieved through data augmentation.
- Parameter sharing can be used to constrain neural architectures to exhibit certain symmetries (e.g., translation invariances with convolutions and pooling).
- This can lead to faster training, better generalization, and requires less training data.

Problem Statement

We focus on symmetries under permutations.

Permutation Invariance/Equivariance

Let $f_{\boldsymbol{\theta}}(\cdot) : \mathbb{X}^N \mapsto \mathbb{Y}^M$ a function parametrized by $\boldsymbol{\theta}$, and $\mathcal{G}$ be the permutation group ($\mathcal{G} = \mathcal{S}_N$).

- If $M = 1$, $f_{\boldsymbol{\theta}}$ is permutation-**invariant** if

$$f_{\boldsymbol{\theta}}(g \cdot \boldsymbol{x}) = f_{\boldsymbol{\theta}}(\boldsymbol{x}), \quad \forall g \in \mathcal{G}, \mathbf{x} \in \mathbb{X}^N$$

- If $M = N$, $f_{\boldsymbol{\theta}}$ is permutation-**equivariant** if

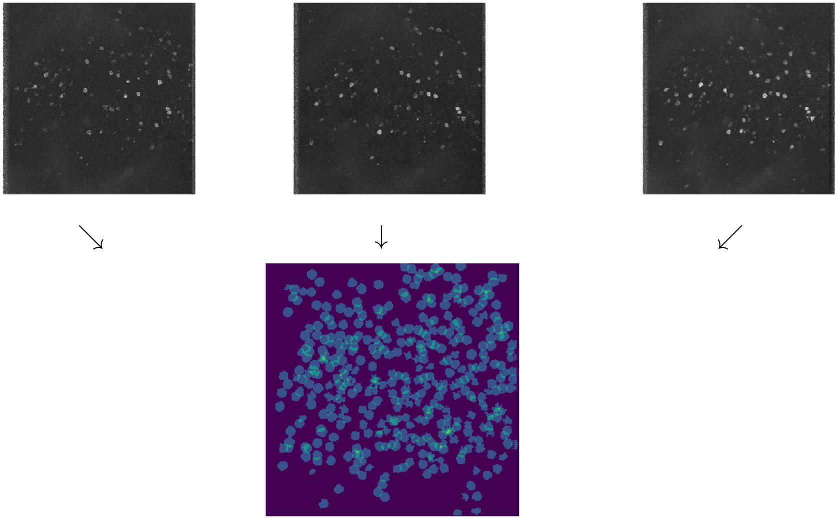
$$f_{\boldsymbol{\theta}}(g \cdot \boldsymbol{x}) = g \cdot f_{\boldsymbol{\theta}}(\boldsymbol{x}), \quad \forall g \in \mathcal{G}, \mathbf{x} \in \mathbb{X}^N$$

More general definitions are available (Ravanbakhsh et al., ICML 2017).

- What is the structure of $f_{\boldsymbol{\theta}}$?

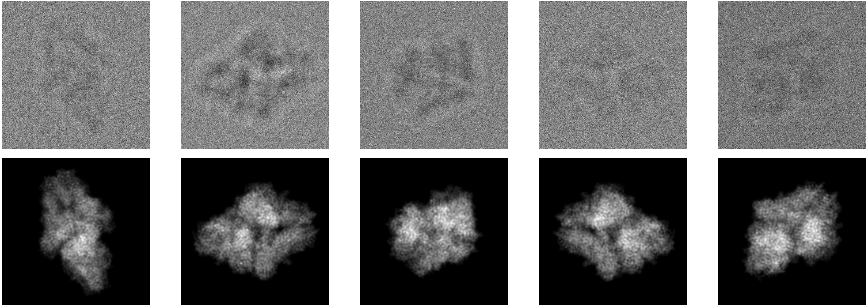
FLATIRON INSTITUTE
Center for Computational Mathematics

Potential applications - Segmentation in calcium imaging



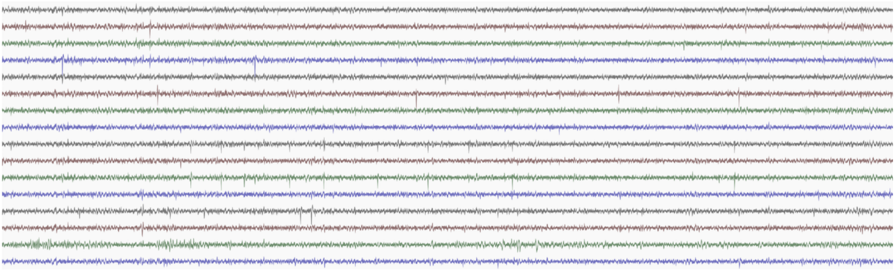
Output is image-permutation invariant.

Potential applications - Batch denoising in cryo-EM



Output is image-permutation equivariant.

Potential applications - Spike sorting



Output is channel-permutation invariant.

Permutation Invariance I

Permutation Invariance (Zaheer et al., NIPS 2017)

$f_{\boldsymbol{\theta}} : [0,1]^N \mapsto \mathbb{R}$ is a permutation invariant continuous function if and only if it has the representation

$$f(x_1, \ldots, x_N) = \rho\left(\sum_{n=1}^{N} \phi(x_n)\right),$$

for continuous functions $\rho : \mathbb{R}^{K+1} \mapsto \mathbb{R}$ and $\phi : \mathbb{R} \mapsto \mathbb{R}^{K+1}$ with $K \leq N$.

Example

$f(x_1, x_2) = x_1 x_2 (x_1 + x_2 + 3)$ can be represented with $\phi(x) = [x, x^2, x^3]$ and $\rho([u, v, w]) = uv - w + 3(u^2 - v)/2$.

Permutation Invariance II

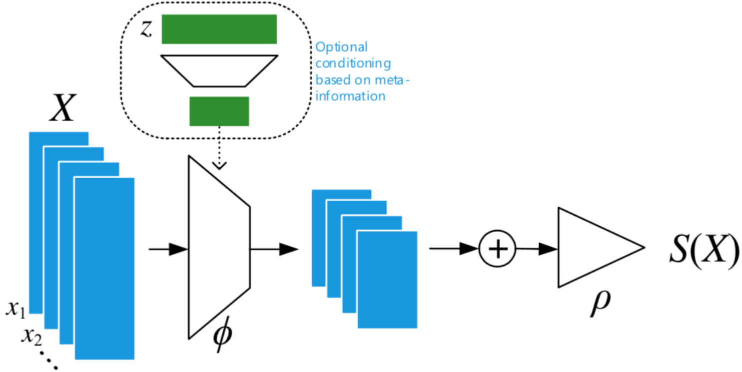


Figure 1: Neural network architecture for permutation invariant functions

Permutation Equivariance

Let $f_{\boldsymbol{\theta}} : [0,1]^N \mapsto \mathbb{R}^N$, with $f_{\boldsymbol{\theta}}(\boldsymbol{x}) = [f_{\boldsymbol{\theta}}^1(\boldsymbol{x}), \ldots, f_{\boldsymbol{\theta}}^N(\boldsymbol{x})]$. $f_{\boldsymbol{\theta}}$ is permutation equivariant if

$$f^i(\boldsymbol{x}) = h(g(x_i), \phi(\boldsymbol{x})),$$

with h, g appropriate functions and ϕ a permutation invariant function.

Equivariant neural network layers (Zaheer et al., NIPS 2017)

- If f_Θ is a standard neural network layer with one input and one output channel
$$f_\Theta(\boldsymbol{x}) = \sigma(\Theta \boldsymbol{x}), \quad \Theta \in \mathbb{R}^{N \times N}.$$
then $\Theta = \lambda I_N + \gamma \mathbf{1_N} \mathbf{1_N}^\top$.
- If $f_{\boldsymbol{\theta}}$ is a standard neural network layer with D input and D' output channels then
$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sigma(\boldsymbol{x}\Lambda + \mathbf{1_N}\mathbf{1_N}^\top \boldsymbol{x}\Gamma),$$
with $\Lambda, \Gamma \in \mathbb{R}^{D \times D'}$.

FLATIRON INSTITUTE Center for Computational Mathematics

Equivariant Architectures

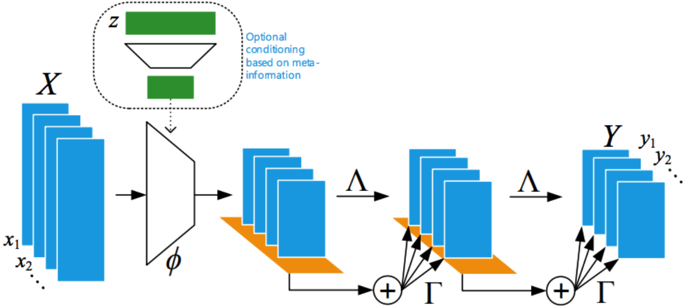
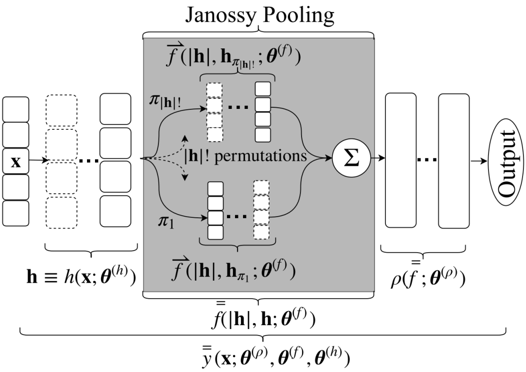


Figure 2: Neural network architecture for permutation equivariant functions

Extensions - modeling interactions



- Instead of pooling over all individual variables, pool over interaction terms over all subsets of fixed cardinality.
- Can be done in a computationally tractable way (Murphy et al., ICLR 2019).

Extensions to matrices/tensors (Hartford et al., ICML 2018)

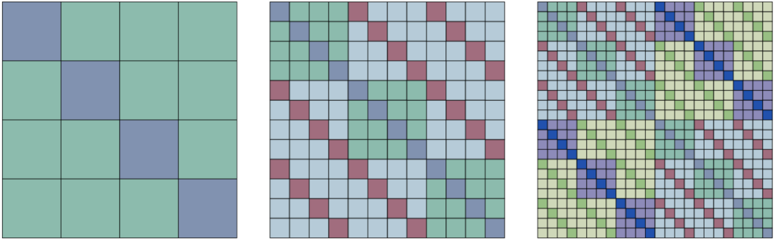


Figure 3: Parameter sharing for equivariant NN layers between tensors

Similar architectures can be defined for NN models over matrices (or higher order tensors) with invariance/equivariance properties:

$$f_{\boldsymbol{\theta}}(P_1 X P_2) = f_{\boldsymbol{\theta}}(X) \qquad \text{(invariance)}$$

$$f_{\boldsymbol{\theta}}(P_1 X P_2) = P_1 f_{\boldsymbol{\theta}}(X) P_2 \qquad \text{(equivariance)}$$

Applications to matrix completion, factorization etc.

Exchangeability of random variables

de Finetti theorem

If $X = (X_1, X_2, \ldots, X_N)$ is an exchangeable sequence of random variables then it can be represented as

$$p(X|\alpha) = \int_{\boldsymbol{\theta}} \left[\prod_{n=1}^{N} p(x_i|\boldsymbol{\theta}) \right] p(\boldsymbol{\theta}|\alpha)\, \mathrm{d}\boldsymbol{\theta}.$$

Example: exponential families

$$p(x|\boldsymbol{\theta}) \propto \exp\left(\langle \phi(x), \boldsymbol{\theta} \rangle - g(\boldsymbol{\theta})\right)$$

$$p(\boldsymbol{\theta}|\alpha, m_0) \propto \exp\left(\langle \boldsymbol{\theta}, \alpha \rangle - m_0 g(\boldsymbol{\theta}) - h(\alpha, m_0)\right)$$

$$p(X|\alpha) \propto \exp\left(h\left(\alpha + \sum_{x \in X} \phi(x), m_0 + m \right) - h(\alpha, m_0) \right)$$

Probabilistic symmetry (Bloem-Reddy and Teh, arXiv 2019)

Let X, Y random variables and $\mathcal{G}$ an appropriate group.

- $P_{Y|X}$ is conditionally $\mathcal{G}$-invariant if and only if $(X, Y) \stackrel{d}{=} (g \cdot X, Y), \forall g \in \mathcal{G}$.
- $P_{Y|X}$ is conditionally $\mathcal{G}$-equivariant if and only if $(X, Y) \stackrel{d}{=} (g \cdot X, g \cdot Y), \forall g \in \mathcal{G}$.

Maximal Invariance

A statistic $\mathcal{M}$ is maximally invariant if

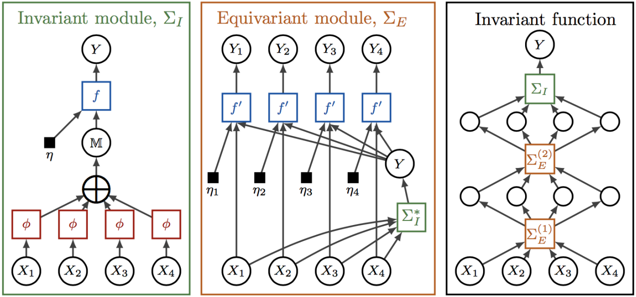
$$M(x_1) = M(x_2) \Rightarrow g \cdot x_1 = x_2 \text{ for some } g \in \mathcal{G}.$$

Modeling invariant distributions

$\mathcal{G}$-Invariance (Bloem-Reddy and Teh, arXiv 2019)

$P_{Y|X}$ is $\mathcal{G}$-invariant if and only if there exists $f : [0,1] \times \mathbb{X} \mapsto \mathbb{Y}$ such that

$$(X, Y) \stackrel{d}{=} (X, f(\eta, M(X))), \text{ with } \eta \sim U[0,1]$$



Predict output with variable cardinality

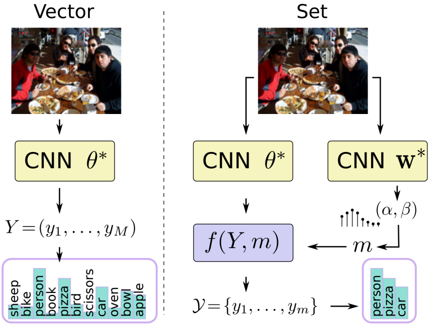


Figure 4: Rezatofighi et al., ICCV 2017

References

- Ravanbakhsh et al., ICML 2017
- Zaheer et al., NIPS 2017
- Murphy et al., ICLR 2019
- Hartford et al., ICML 2018
- Bloem-Reddy and Teh, arXiv 2019
- Rezatofighi et al., ICCV 2017