LA-UR-22-29899

Martin Larocca[1,2]@QTML2023

@martinlaroo

Equivariant Quantum Models

- ❏ ***Theory for Equivariant Quantum Neural Networks***, arXiv:2210.08566 (2022)
  + other GQML:
- ❏ ***Group-Invariant Quantum Machine Learning***, PRX Quantum 3, 030341 (2022)
- ❏ ***Representation Theory for Geometric Quantum Machine Learning***, arXiv:2210.07980 (2022)
- ❏ ***Theoretical Guarantees for Sn-Equivariant Quantum Neural Networks***, arXiv:2210.09974 (2022)
- ❏ ***Building spatial symmetries into quantum circuits for faster training***, arXiv:2207.14413 (2022)

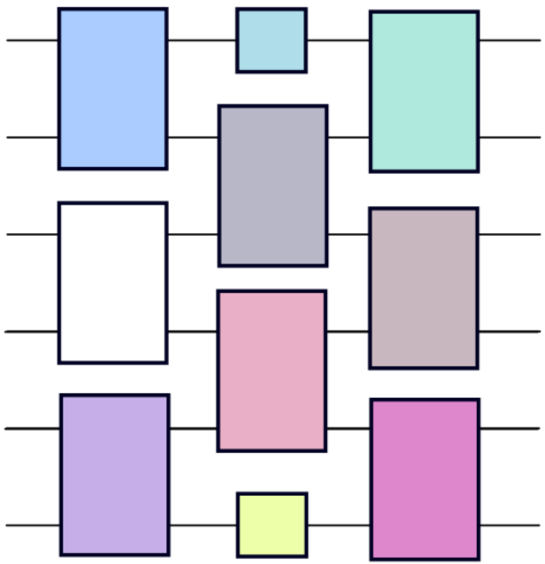
Outline:

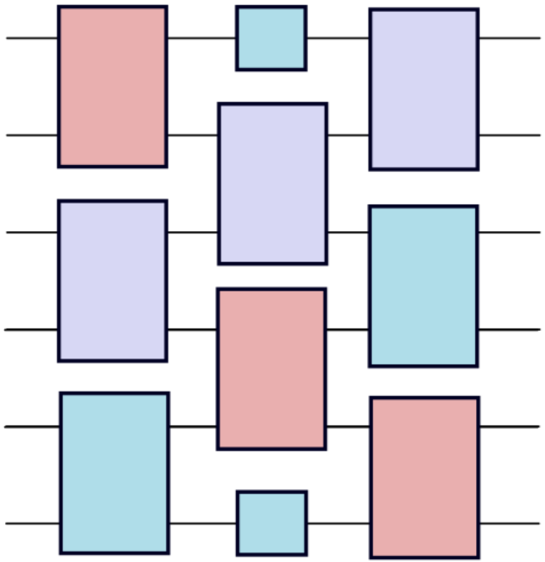
- We will begin by explaining where the **motivation** for this program comes from: classical **geometric deep learning (GDL)**.
- Then we will define what **GQML** intends to be: a set of guidelines / recipes to achieve **invariant/equivariant quantum models**.
- We will argue that symmetry = improved performance in terms of #shots, parameters and data.
- We will end up with discussion and an outlook to what comes next!

1. Theoretical Division, Los Alamos National Laboratory
2. Center for Nonlinear Studies, Los Alamos National Laboratory

Los Alamos NATIONAL LABORATORY

Center for Nonlinear Studies





how do we choose how to correlate gates?

broad motivation: classical

Geometric Deep Learning (GDL)*

Building symmetry into ML has led to major breakthroughs.

- imposing **translational symmetry and parameter sharing** allowed deep CNNs to overthrow fully connected architectures 'essentially solving' computer vision.
- **GDL** conceptualizes CNN success (symmetry exploitation) and generalizes it to tasks with other symmetries.
- As we will see, quantum ml is no different and we can not only verify but also rigorously quantify the **advantages of enforcing symmetry** in our architectures.

* Cohen and Welling, *Group equivariant convolutional networks*, ICML, 2016.
* Bronstein et al, *Geometric deep learning: Grids, groups, graphs, geodesics, and gauges*, *arXiv:2104.13478*, 2021.
* GitHub Chen-Cai-OSU / awesome-equivariant-network for GDL bibliography summary.

example setting: **Supervised Learning**

train a model on labelled data to make predictions on new data.

model

$$h_\theta(x)$$

Prediction

???

Geometric Deep Learning (GDL): lets exploit symmetry in ML models!

main concept: **Label Symmetry**

certain learning tasks have **labels** that are **invariant** under some **set of transformations** acting on the data

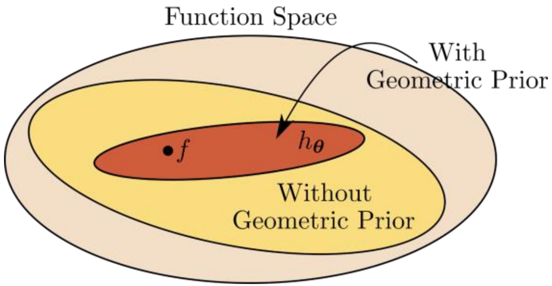
$\mathbb{SO}(2)$

$\xrightarrow{f}$

'cat'

We will say such label-producing f is **group-symmetric** (or **invariant**).

By imposing **invariance** we **constrain the search** to smaller region
→ less **parameters**, better **generalization** with less **data**.

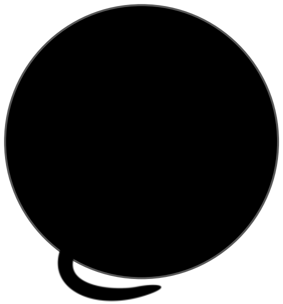


This is the kind of **prior** we are looking for the **parameter correlation** design!

clarification: We are asking *the models* to be invariant, not the data itself!

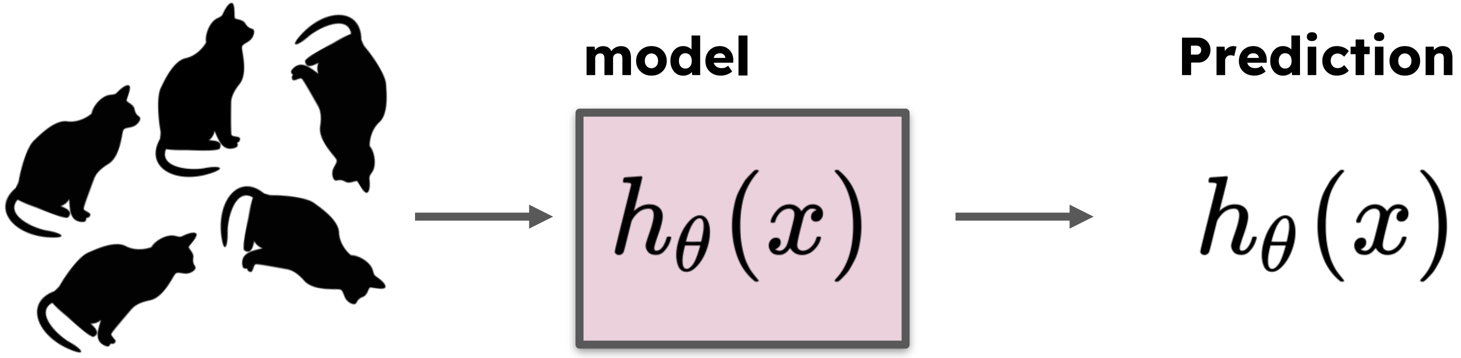
not $\mathbb{SO}(2)$ invariant

$\mathbb{SO}(2)$ invariant



…otherwise we would be looking at **circular** (or even *spherical)* cats like a righteous **physicist** would do. We are **NOT**!

Data Augmentation. *Brute-forcing* group-invariance.



Feed *augmented* data and hope the model **“learns”** the invariance.

issues:

- ❌ data and parameter **inefficient**
- ❌ even if we learn some **theta that makes model invariant**, no guarantee of **invariance over unseen data**.

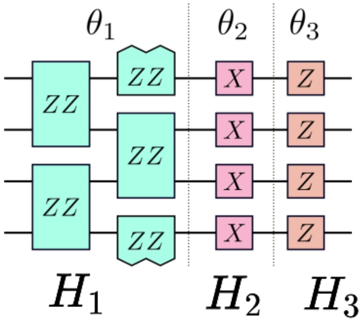
Instead, it is **strongly desirable** to have models that are **invariant *by design***, i.e. for **all theta.**

setting: Quantum Neural Network **(QNN)** -based Supervised Learning

Data

$$\rho$$

QNN



Measurement

$$\begin{aligned} \langle \mathcal{U}_\theta(\rho), O \rangle &= \mathrm{Tr}[\mathcal{U}_\theta(\rho) O] \\ &= h_\theta(\rho) \end{aligned}$$

proposal:

Geometric QML[1] = QML + Geometric DL

We want models that are *by construction* **group-invariant**

$$h_\theta(\sigma \cdot x) = h_\theta(x)$$

How do we achieve this? **GQML's blueprint for invariant models**

$$x \to \rho(x)$$

Equivariant Embedding

$$\rho(\sigma \cdot x) = \sigma \cdot \rho(x)$$

Equivariant circuit

$$\mathcal{U}_\theta(\sigma \cdot \rho) = \sigma \cdot \mathcal{U}_\theta(\rho)$$

$$\mathcal{U}_\theta$$

$$O$$

invariant measurement

$$\langle \sigma \cdot \rho, O \rangle = \langle \rho, O \rangle$$

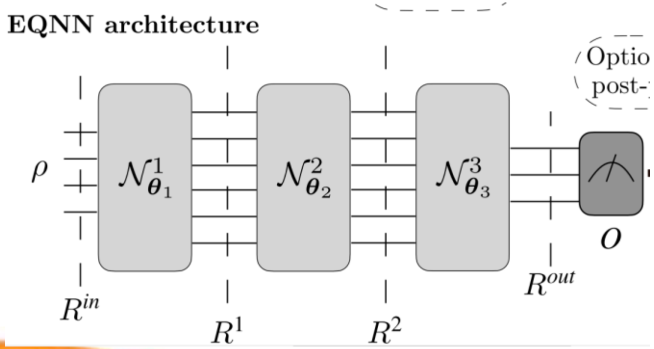
How exactly to **implement** and **parametrize equivariant linear maps**,

and in particular

- The equivariant **encoding**
- The equivariant **unitaries** and **CPTP channels** – aka **EQNN layers**
- The Invariant **measurement**

Each of these can be parametrized!

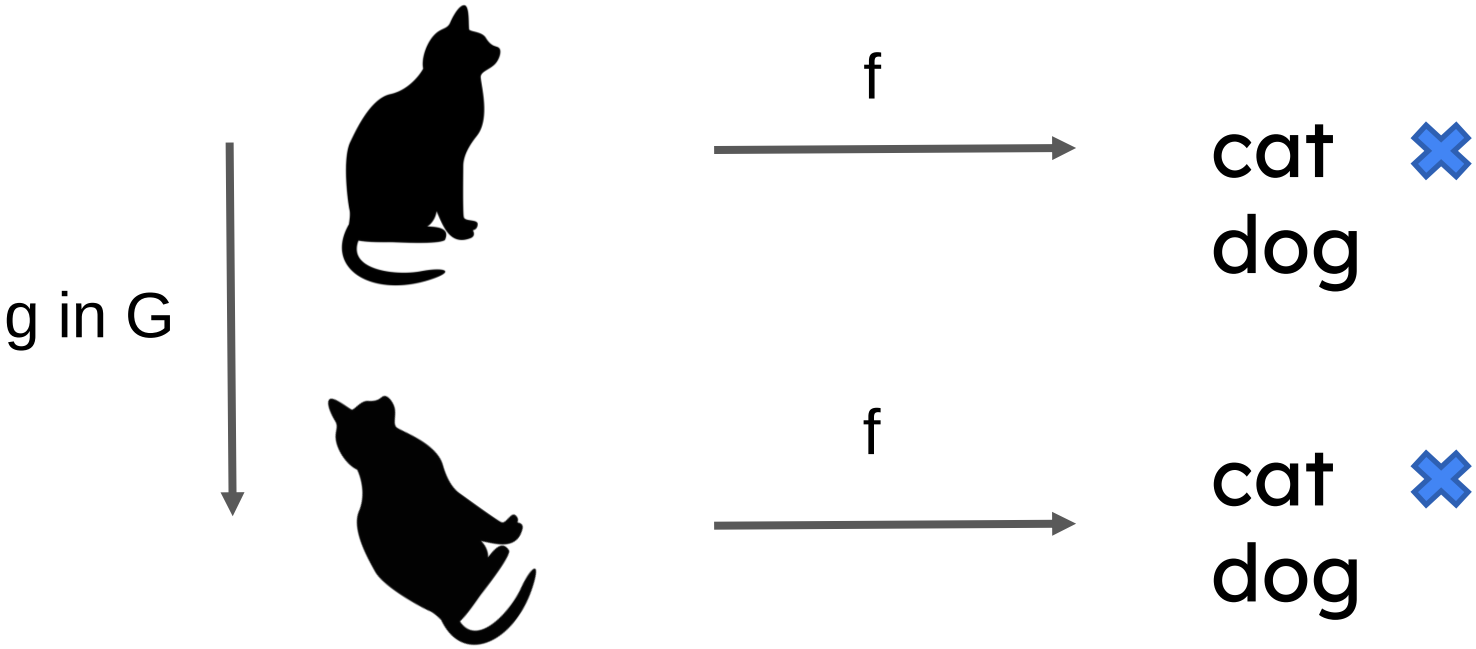
Input and output reps are fixed by problem,
Middle EQNN layers have freedom of rep
Once we fix these, there is a well define space
we can parametrize and explore



NOT TODAY, BUT CHECK OUT

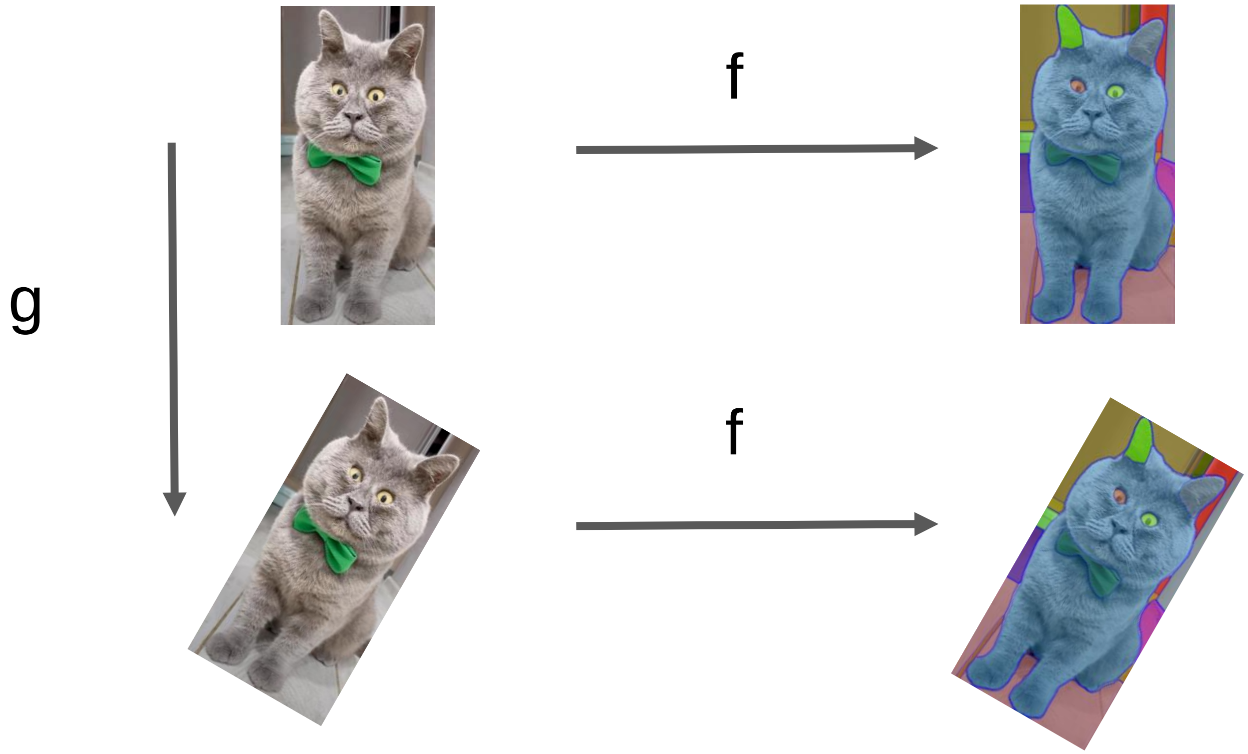
*Nguyen et al., *Theory for equivariant quantum neural networks*, (2022).

Invariance vs Equivariance: (property of a map f)



Output of f is scalar, but more importantly, **the group action is the trivial!**

Invariance vs **Equivariance**



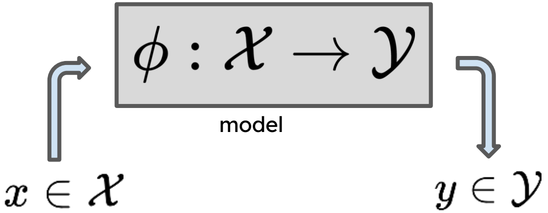
Invariance **AND** Equivariance

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

Invariance is just the case of **\cdot = trivial rep**

From **scalar** to **vector** feature maps

feature maps **need not be scalar-valued.** **Example**: segmentation



Just like before: Equivariant encoding and QNN.
New: replace single **invariant observable** by a tensor of them that holds the same rep as Y.
→ You automatically get ***equivariance*** of the composite **quantum feature map**

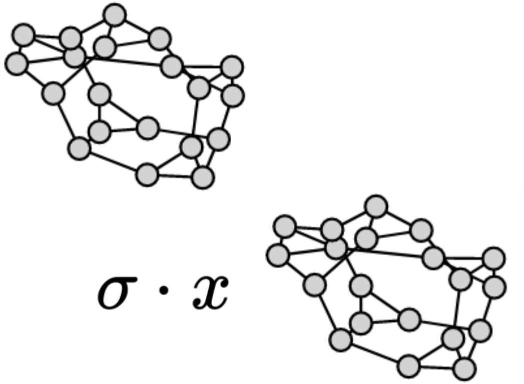
<u>Example: Sn-INvariant problems:</u>

Want to learn a property of n-elements that doesn't depend on ordering (think graphs, n copies of a quantum state, etc., very common!)

Example: learning problems on graphs have **labels** that are **invariant** under Sn action!
e.g. graphs related by node permutations ('relabellings') are isomorphic

x

$\sigma \cdot x$



same number of 4-cliques
#subgraphs = PI property

→ we want a PI model! $h_\theta(\sigma \cdot x) = h_\theta(x)$

Measure PI operator, e.g. sum Xi
→ Quantum graph invariants!

Example: Sn-EQUIvariant problems:

Now, lets say we want **vector** output → measure **set** of **PE** operators, e.g. $\{ X\_i \}$

$$\rightarrow vi(\mathbf{x},th) = < rho\_th(\mathbf{x}), Zi>$$

Now automatically the vector of exp values satisfies

$$v(sg\ \mathbf{x}) = sg\ v(\mathbf{x})$$

This is ideal for integrating in larger scheme, e.g. GNN* → **quantum equivariant features!**

Sanity check: if vector of obs = all trivial irreps, e.g. { sum Xi, sum Yi, sum ZiZj, XXXXXX}

Then

$$v(sg\ \mathbf{x}) = v(\mathbf{x})$$

*Thabet et al, Extending Graph Transformers with Quantum Computed Aggregation." *2022*

Provable Benefits:

- **Gradients:** symmetry = smaller DLA[1,2]

locality

$$\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \frac{1}{\dim(\mathfrak{g})} \mathrm{Tr}\left[O_{\mathfrak{g}}^2\right] \mathrm{Tr}\left[\rho_{\mathfrak{g}}^2\right]$$

expressiveness

entanglement

Resolved Conjecture ML21': Var ~ 1/dimg (more expressive = more concentrating)

it is official: symmetry = smaller DLA
less concentration = less QPU use to resolve loss/gradients.

- **Local minima / overparametrization[3]:** smaller DLA = easier overparametrization
- **Generalization[4]:** less parameters = simpler model = better generalization w less data.

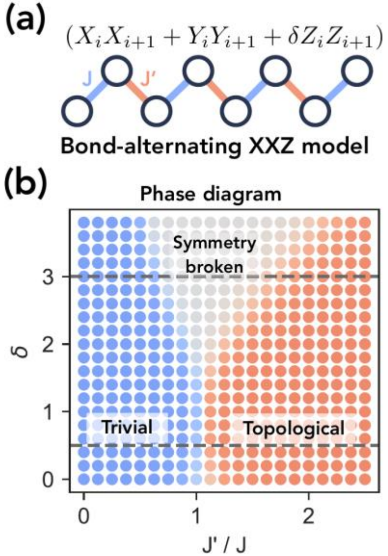
1 Ragone et al, A unified theory of barren plateaus for deep parametrized quantum circuits, *Arxiv,* (2023)
2 Fontana et al, The Adjoint Is All You Need: Characterizing Barren Plateaus in Quantum Ansatze, *Arxiv,* (2023).
3 Larocca et al, Theory of overparametrization in quantum neural networks, *Nature Computational Science* 3.6 (2023): 542-551.
4 Caro et al., Generalization in quantum machine learning from few training data, *Arxiv*, (2021).

Flash Demonstration:

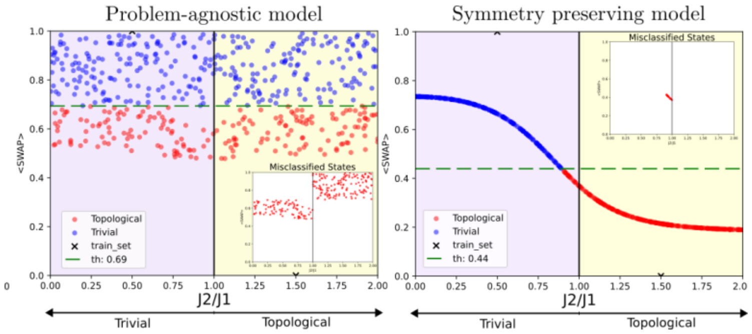
GS classification (qdata) with **SU(2)-equivariant QCNNs**

Problem and model:

- **Classification:** ground state classification problem
- **Symmetry:** SU(2)
- **Model:** SU(2)-equivariant QCNN + SU(2)-equivariant observable



invariant observable
e.g. XXX-interaction



equivariant observables: e.g. <Sx,Sy,Sz>
(or any other subset of operators forming a spin rep)

Outlook / Discussion

We have

- **the why**: motivated the **need** for equivariant models –performance!
- **the how:** layed down a **blueprint** for achieving equivariant models –equivariant encoding, QNN and measurements.
- Established **advantages** of expoiting symmetry:

  Dim (G-equivariant DLA) < dim(DLA) → less **expressiveness** (more problem focused)

    - Larger **gradients**
    - Easier **landscape (local minima/overparametrization)**
    - better **generalization**

Remark: encoding / quantum channel / decoding need not be parametrized!

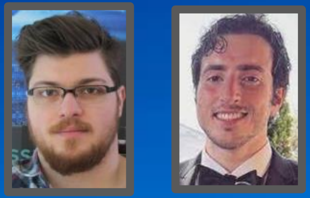
Some open questions / future directions

- **Performance demonstration:** not too many demonstrations of the theoretical outperformance
- **Implementation:– EQUIVARIANTIZE** your favorite model now!
- **Parameter/Data Concentration:** it is crucial that the variance of the model over params and data is not too concentrated, because otherwise (with not too many shots) *"its a machine that produces same feature for different data instances"*, useless.
- **Power and Advantage:** what is the actual power of these. Can we find quantum classical separations?

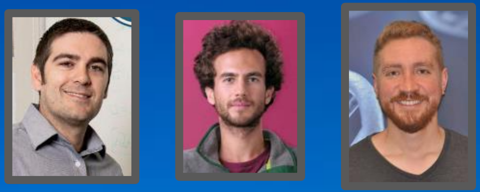
Thank you! Questions?

GQML collabs

SUMMER SCHOOL



GOOGLE



LANL

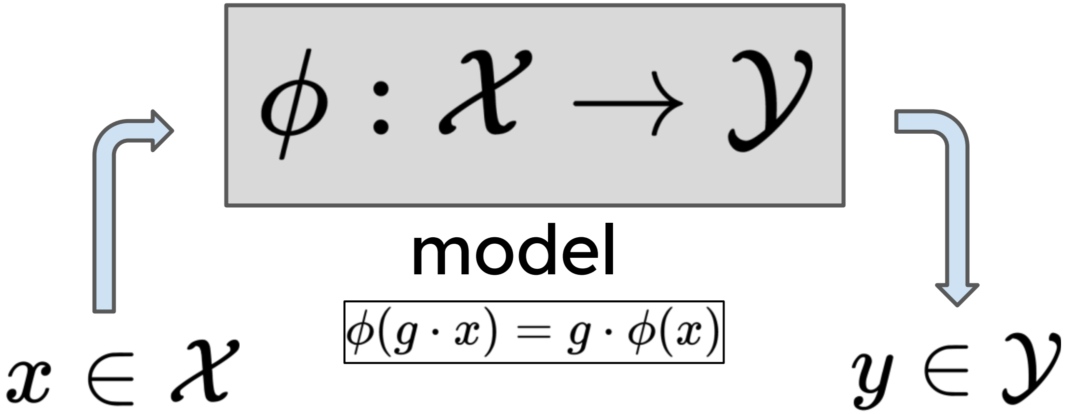
Los Alamos NATIONAL LABORATORY

Center for Nonlinear Studies


IS THIS
EQUIVARIANCE
HERE IN THE ROOM
WITH US?

EQUIVARIANTIZE
YOUR MODEL!

Blueprint for *Equivariant* Models / Features



Just like before: Equivariant encoding and QNN.
New: measure not one but a set of equivariant operators {Oj} with the property that the vector space span{Oj} is a G-module.

Sn-EQUIvariant problems:

instead of predicting some scalar function for the whole graph, we may want some ‘vector’ function, e.g. a scalar for each node, for each edge, for each k-clique.

Instead of measuring single observable measure a vector of them

because the set of vectors already holds a representation of Sn, in this case C{Z_i} = standard sn irrep

we automatically get that v(sg x) = sg v(x)

sanity check: if we choose set of observables forming trivial rep, e.g. a single observable that commutes with G, we get back invariance!

vi(x) = < rho_th(x), Zi> →
if rho_th(x) is equivariant and i consider sg on x
sg * v(x) transforms as the std rep of Sn

→ the vector of expectation values will have a Sn action and we can ask the global model to commute with the symmetry action.

The reason this are called Fourier basis is direct generalization of actual Fourier transform.
Given a f(t) in H holding an action of u1 group via g*f = f(g-1)
We.can find a new basis for H called.fourier such that
The action of the group is now block diagonal. Here diag.
Now, if we want an invariante action we make sure we act on the trivial subps.

At this point, we need to:

- explain how exactly to achieve equivariance of a linear map, and in particular
  - equivariant encoding
  - equivariant unitaries and CPTP channels
  - invariant measurement

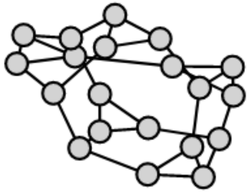
  This produces invariant model h

$$\boxed{f(g \cdot x) = g \cdot f(x)}$$

- More generally, we can in aim at equivariant models satisfying.
- Two reasons:
  - we actually care about tensor outputs, e.g. predict vector field.
  - we care about scalar output but want to integrate this as a component in a larger model – e.g. in a classical ENN , input is vector data not scalars!
- Benefits of equivariance

Example : graph classification (cdata) with **Sn-equivariant QNNs [1]**.

Permutation **invariant** problem (graph learning)



$A \in \mathbb{R}^{n\times n}$

$f(\sigma \cdot A) = f(A)$

Embedding: $H(A) = \sum_{ij} A_{ij} Z_i Z_j$

$$H(\sigma \cdot A) = \sum_{i,j} A_{\sigma(i)\sigma(j)} Z_i Z_j = \sum_{kl} A_{kl} Z_{\sigma(k)} Z_{\sigma(l)} = \sigma \cdot H(A)$$

QNN: generators = { sum Xi , sum Yi, sum ZiZj }

Measurement: sum XiXj

[5] L. Schatzki, **ML**, F. Sauvage, and M. Cerezo, 'Theoretical guarantees for permutation-equivariant quantum neural networks', (2022).