An Introduction to Geometric Deep Learning on Sets, Graphs and Grids

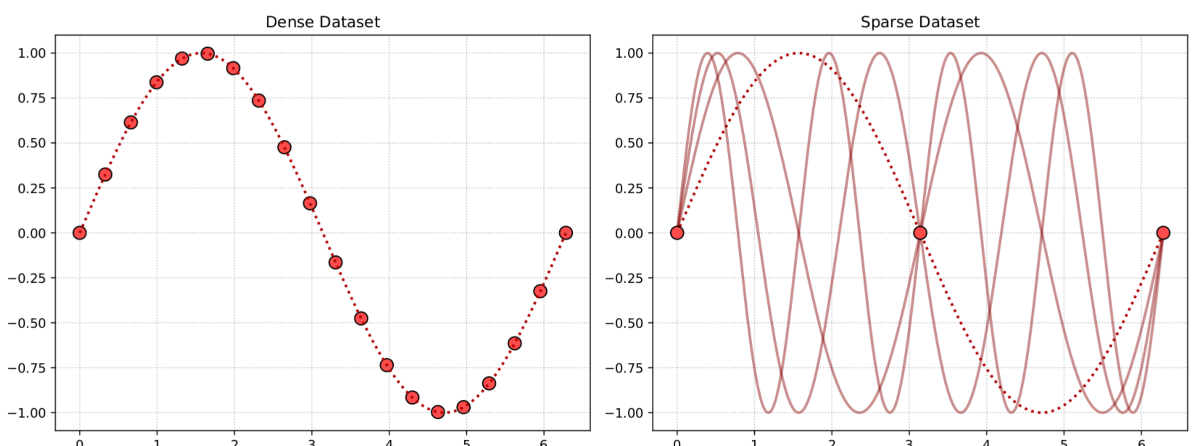
Gabriel Jacob Perin, Advisor: Nina S. T. Hirata
Instituto de Matemática, Estatística e Ciência da Computação - Universidade de São Paulo

IME USP

More Info

Motivation

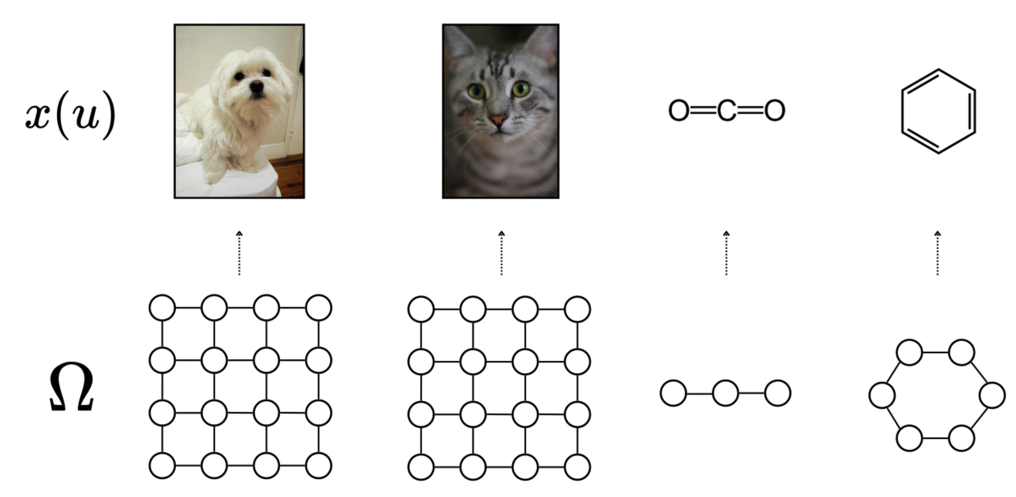
- GDL provides a **unified mathematical framework** for deep learning architectures.
- **The Curse of Dimensionality**: High-dimensional datasets are **sparse**, thus leading to **poor generalization.**



Signal

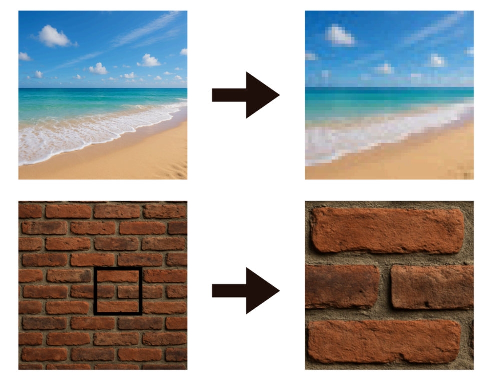
Definition (Signal). *Let Ω be a finite set and C a vector space of dimension d (whose dimensions are called channels). A* ***signal*** *on Ω is a function*

$$x : \Omega \to C.$$



Scale Separation

Key Idea: phenomena operating at different scales can often be modeled independently.



Definition (Coarsening Operator). *Let $\mathcal{S} = \{x : \Omega \to C\}$ and $\mathcal{S}' = \{x' : \Omega' \to C\}$ be sets of signals, with $|\Omega'| \leq |\Omega|$. A Coarsening Operator P is a function $P : \mathcal{S} \to \mathcal{S}'$.*

Symmetry

Key Idea: leverage group theory to formalize symmetries.

- A symmetry is a **transformation** that **preserves** an object's **essential information**.
- for example, a reflected dog image remains recognizable as the same dog.

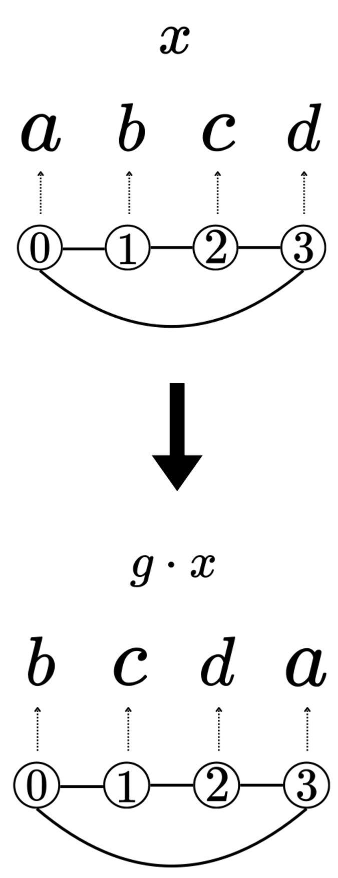
Definition (Invariance). *Let $(\mathcal{G}, \circ)$ be a group that acts on the set X with left group action $\cdot$ and Y a set. A function $f : X \to Y$ is said to be $\mathcal{G}$-invariant if*

$$f(g \cdot x) = f(x) \quad \forall g \in \mathcal{G}, x \in X.$$

Definition (Equivariance). *Let $(\mathcal{G}, \circ)$ be a group acting on the sets X and Y. A function $f : X \to Y$ is said to be $\mathcal{G}$-equivariant if*

$$f(g \cdot x) = g \cdot f(x) \quad \forall g \in \mathcal{G},\ x \in X,$$

where $\cdot$ denotes the corresponding group action on each set.



GDL Blueprint

- Leverage **scale separation** to **reduce** signal **dimension.**
- Leverage **symmetries** to **restrict** the **hypothesis set**.

Definition (Equivariant Block). *Let $\mathcal{S} = \{x : \Omega \to C\}$, $\mathcal{S}' = \{x' : \Omega' \to C'\}$ and $\mathcal{S}'_c = \{x'_c : \Omega'_c \to C'\}$ be sets of signals, with $|\Omega'_c| \leq |\Omega'|$. Let $(\mathcal{G}, \circ)$ be a group acting on $\Omega, \Omega', \Omega'_c$ through $\cdot$. We define the following building blocks:*

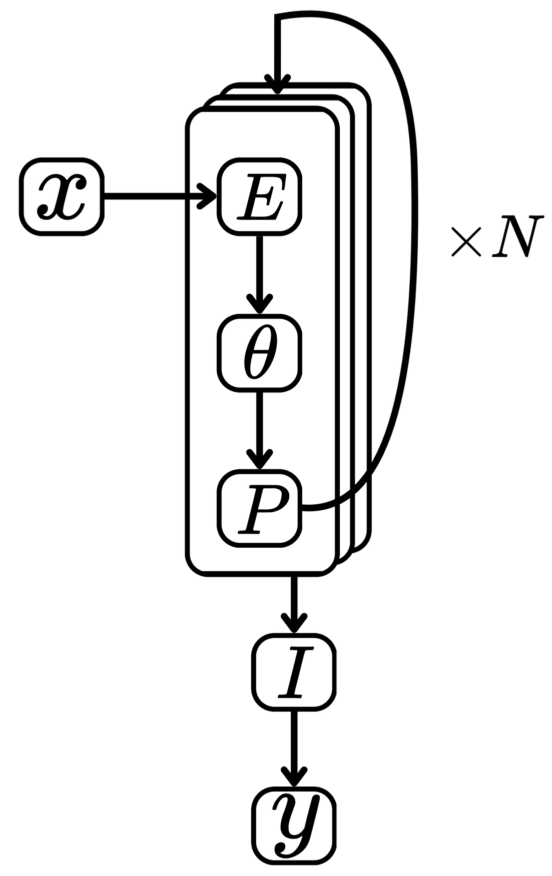
1. *Linear $\mathcal{G}$-equivariant layer $E : \mathcal{S} \to \mathcal{S}'$, satisfying $E(g \cdot x) = g \cdot E(x)$ for all $g \in G$ and $x \in \mathcal{S}$.*
2. *Nonlinearity $\theta : \mathcal{S}' \to \mathcal{S}'$, obtained applying a non-linear function $\vartheta : C' \to C'$ coordinate-wise: $\theta(x')(u') = \vartheta(x'(u'))$, for $x' \in \mathcal{S}'$ and $u' \in \Omega'$.*
3. *Coarsening Operator $P : \mathcal{S}' \to \mathcal{S}'_c$.*

A $\mathcal{G}$-Equivariant Block $B : \mathcal{S} \to \mathcal{S}'_c$ is a composition

$$B(x) = P(\theta(E(x))), \quad x \in \mathcal{S}.$$

Definition (Invariant Layer). *Let $\mathcal{S} = \{x : \Omega \to C\}$ be a set of signals, and $\mathcal{Y}$ the set of labels. Let $(\mathcal{G}, \circ)$ be a group that acts on Ω. A $\mathcal{G}$-Invariant Layer is a function $I : \mathcal{S} \to \mathcal{Y}$ that satisfies*

$$I(g \cdot x) = I(X) \quad \text{for all } x \in S, g \in \mathcal{G}.$$



Sets

Key Idea: When modeling functions on sets, the order of the input should not matter.

- Assume $\Omega = [d]$, $C = \mathbb{R}^k$.
- Representation:

$$X = \begin{bmatrix} \text{---} & x(1) & \text{---} \\ \text{---} & x(2) & \text{---} \\ & \vdots & \\ \text{---} & x(d) & \text{---} \end{bmatrix}$$

Definition (Permutation Invariance). *A function $f : \mathbb{R}^{d\times k} \to \mathcal{Y}$, where $\mathcal{Y}$ is any set, is said to be permutation invariant if, for all permutation matrices $P \in \mathbb{R}^{d\times d}$ and all $X \in \mathbb{R}^{d\times k}$,*

$$f(PX) = f(X).$$

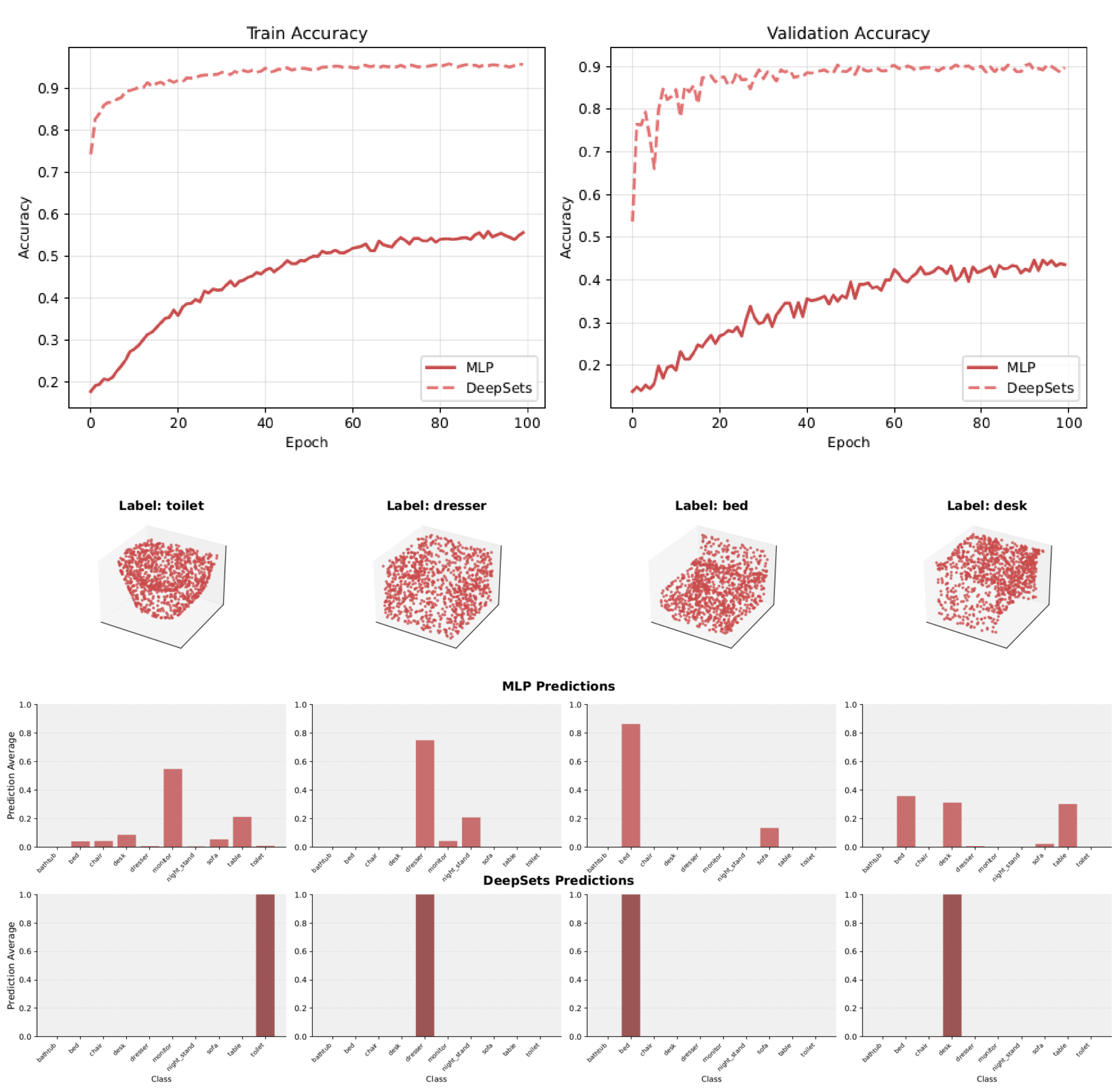
Definition (Permutation Equivariance). *A function $f : \mathbb{R}^{d\times k} \to \mathbb{R}^{d\times k'}$ is said to be permutation equivariant if, for all permutation matrices $P \in \mathbb{R}^{d\times d}$ and all $X \in \mathbb{R}^{d\times k}$,*

$$f(PX) = Pf(X).$$

Definition (DeepSets). *Let $MLP_{\mathcal{W}} : \mathbb{R}^k \to \mathbb{R}^{k'}$ and $MLP_{\mathcal{W}'} : \mathbb{R}^{k'} \to \mathcal{Y}$ be MLPs. Let $\oplus$ denote any permutation-invariant operator. A DeepSet is a function $DS_{\mathcal{W},\mathcal{W}'} : \mathbb{R}^{d\times k} \to \mathcal{Y}$, defined by*

$$DS_{\mathcal{W},\mathcal{W}'}(X) = MLP_{\mathcal{W}'}\left(\bigoplus_{i\in[d]} MLP_{\mathcal{W}}(X_i)\right)$$

Experiment: Point cloud classification on ModelNet10.



Graphs

Key Idea: When modeling functions on graphs, the order (label) of the nodes should not matter.

- Let $G = (V, E)$. Assume $\Omega = V = [d]$.
- Representation:
  - Feature matrix X
  - Adjacency matrix A

Definition (Permutation Invariance). *A function $f : \mathbb{R}^{d\times k} \times \mathbb{R}^{d\times d} \to \mathcal{Y}$, where $\mathcal{Y}$ is any set, is said to be permutation invariant if, for all permutation matrices $P \in \mathbb{R}^{d\times d}$, all $X \in \mathbb{R}^{d\times k}$, and all $A \in \mathbb{R}^{d\times d}$,*

$$f(PX, PAP^\top) = f(X, A).$$

Definition (Permutation Equivariance). *A function $f : \mathbb{R}^{d\times k} \times \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times k'}$ is said to be permutation equivariant if, for all permutation matrices $P \in \mathbb{R}^{d\times d}$, all $X \in \mathbb{R}^{d\times k}$, and all $A \in \mathbb{R}^{d\times d}$,*

$$f(PX, PAP^\top) = Pf(X).$$

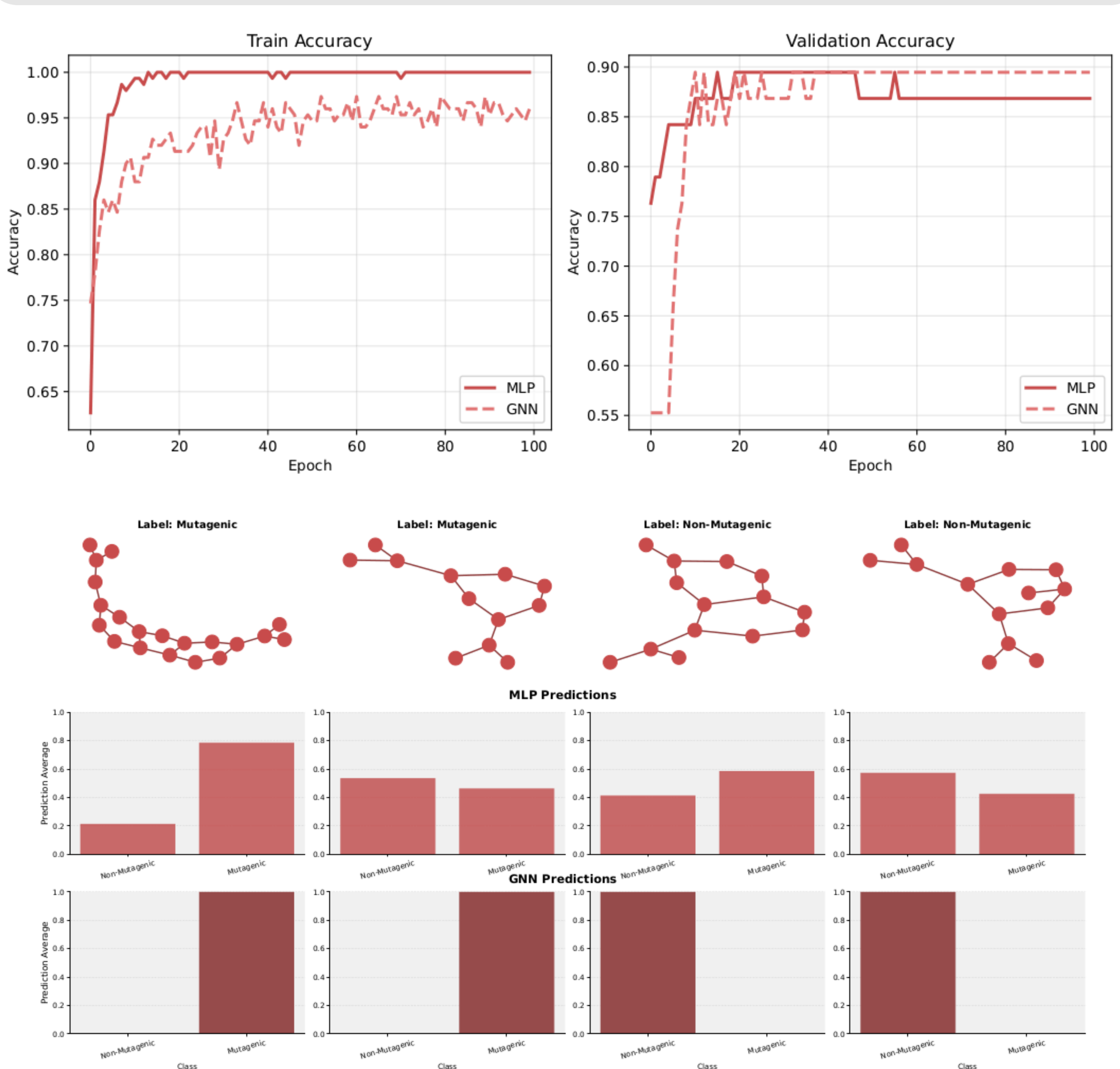
Definition (GNN Layer). *Let $G = (V, E)$ be a graph with corresponding adjacency matrix $A \in \mathbb{R}^{d\times d}$ and feature matrix $X \in \mathbb{R}^{d\times k}$. Define $\mathcal{M}_d = \bigcup_{i\in\{0\}\cup[d]} \mathbb{R}^{i\times d}$ to be the set of all matrices with up to d rows. Let $\sigma : \mathbb{R}^k \times \mathcal{M}_d \to \mathbb{R}^{k'}$ be a permutation invariant function w. r. t. the second argument (i. e. $\sigma(h, PH) = \sigma(h, H)$ for all permutation matrices P). A GNN layer $f : \mathbb{R}^{d\times k} \times \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times k'}$ is a function defined by*

$$f(X, A) = \begin{bmatrix} \text{---} & \sigma(X_1, X_{N_A(1)}) & \text{---} \\ \text{---} & \sigma(X_2, X_{N_A(2)}) & \text{---} \\ & \vdots & \\ \text{---} & \sigma(X_d, X_{N_A(d)}) & \text{---} \end{bmatrix}.$$

One example is the attentional GNN, used by the Transformer:

$$\sigma(X_i, X_{N_A(i)}) = \vartheta\left(X_i, \bigoplus_{j\in N_A(i)} a(X_i, X_j)\varrho(X_j)\right),$$

Experiment: Molecule classification on MUTAG.



Grids

Key Idea: When modeling functions on images, where the object appears in the image should not matter.

- Assume $\Omega = \{0, ..., d-1\}\times\{0, ..., d-1\}, C = \mathbb{R}$
- Representation: $X \in \mathbb{R}^{\Omega}$ s. t. $X_{ij} = x(i, j)$

Definition (Translation Invariance). *A function $f : \mathbb{R}^{\Omega} \to \mathcal{Y}$, where $\mathcal{Y}$ is any set, is said to be translation invariant if, for all $m, n \in \{0, ..., d-1\}$, all $X \in \mathbb{R}^{\Omega}$,*

$$f(S^m X (S^n)^\top) = f(X).$$

Definition (Translation Equivariance). *A function $f : \mathbb{R}^{\Omega} \to \mathbb{R}^{\Omega}$ is said to be translation equivariant if, for all $m, n \in \{0, ..., d-1\}$, all $X \in \mathbb{R}^{\Omega}$,*

$$f(S^m X (S^n)^\top) = S^m f(X)(S^n)^\top.$$

Where S^m, S^n denote (cyclic) shift matrices.

Definition (Convolution). *Let $x : \Omega \to \mathbb{R}$ be the input signal and $w : \Omega \to \mathbb{R}$ be the convolution kernel. The convolution operator $*$ is defined as:*

$$(x * w)(i, j) = \sum_{m=0}^{k-1}\sum_{n=0}^{k-1} w(m, n)\, x(i-m, j-n),$$

where $x * w : \Omega \to \mathbb{R}$.

Experiment: Digit classification on MNIST.