MATHEMATICS OF NEURAL NETWORKS

OUTLINE

- Groups
- ACTIONS/LINEAR REP.
- EQUIVARIANCE/INVARIANCE
- DATA AUGMENTATION/DAC
- ENSEMBLING/REYNOLD AVERAGING
- IRREDUCIBLE REPRESENTATIONS/ SCHUR'S LEMMA
- EQUIVARIANT MAPPINGS.
- BEYOND IMAGES
  - GRAPHS
  - POINT CLOUDS
  - ETC.

Groups

A GROUP G VERIFIES THE FOLLOWING AXIOMS:

$\exists e \in G : \forall g \in G \quad g \circ e = g$ identity element.

$\forall g, s \in G \quad g \circ s \in G$

$\forall g \in G \ \exists s \in G \quad g \circ s = e$

EXAMPLES:

- $\mathbb{Z}_n = \{e, g, g^2, \ldots, g^{n-1}\}$

  $g^k \circ g^{n-k} = e$

- $(\mathbb{R}_+, \times)$

GROUP ACTION : FOR A VECTOR SPACE V (e.g. $\mathbb{R}^n$)

$T : G \times V \to V$ IS A GROUP ACTION

iif it VERIFIES

$$T_g \circ T_s = T_{g \circ s}$$

$T_e v = v$ identity mapping $\forall v \in V$

LINEAR REPRESENTATIONS = LINEAR GROUP ACTIONS

$$T_g(\alpha v + \beta u) = \alpha T_g v + \beta T_g u$$

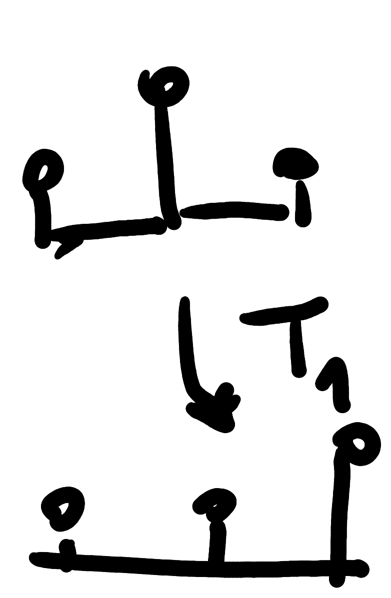
$$\alpha, \beta \in \mathbb{R} \quad u, v \in V$$

EXAMPLE $V = \mathbb{R}^n \Rightarrow T_g \in \mathbb{R}^{n \times n} \quad \forall g \in G$

INVERTIBLE MATRICES

$$T_e = I_n \quad T_g^{-1} = T_{g^{-1}}$$

- SHIFTS OF A DISCRETE TIME SERIES $(n=3)$



$\{T_1, T_2, I\}$ $\qquad T_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ $\qquad \mathbb{Z}_3$

$$T_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

- TRIVIAL REPRESENTATION

$$T_g = I \quad \forall g \in G$$ IS A VALID GROUP ACTION

- REGULAR REPRESENTATION

$$V = \mathbb{R}^{|G|}$$

$$T_g e_s = e_{g \circ s}$$

PERMUTATION MATRICES

$$e_s = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$ → s TH POSITION

<u>EQUIVARIANCE</u>: A FUNCTION $f: \mathbb{R}^m \to \mathbb{R}^p$ IS EQUIVARIANT iif

$$f(T_g x) = \tilde{T}_g f(x) \quad \forall g \in G, \ \forall x \in \mathbb{R}^m$$

if $\underbrace{T \text{ AND } \tilde{T}}_{\text{MIGHT BE DIFFERENT!}}$ ARE TWO VALID GROUP ACTIONS (LINEAR OR NON LINEAR)

<u>EXAMPLE</u>

$f(x) = x$ is G EQUIVARIANT FOR ANY GROUP.

PROPERTIES OF LINEAR REPS.

T_g : ACTION ON $\mathbb{R}^n$

$\tilde{T}_g$ " " $\mathbb{R}^m$

1) $\begin{bmatrix} T_g & \\ & \tilde{T}_g \end{bmatrix}$ LINEAR REP ON $\mathbb{R}^{n+m}$

2) $T_g \otimes \tilde{T}_g$ LINEAR REP ON $\mathbb{R}^{nm}$

KRONECKER PRODUCT

3) FOR INVERTIBLE $A \in \mathbb{R}^{n \times n}$

$A T_g A^{-1}$ LINEAR REP ON $\mathbb{R}^n$

4) T_g LINEAR REP OF G

$\tilde{T}_{g'}$ LINEAR REP OF G'

$\Rightarrow T_g \tilde{T}_{g'}$ LINEAR REP OF $G \times G'$

INVARIANCE: f is invariant iff

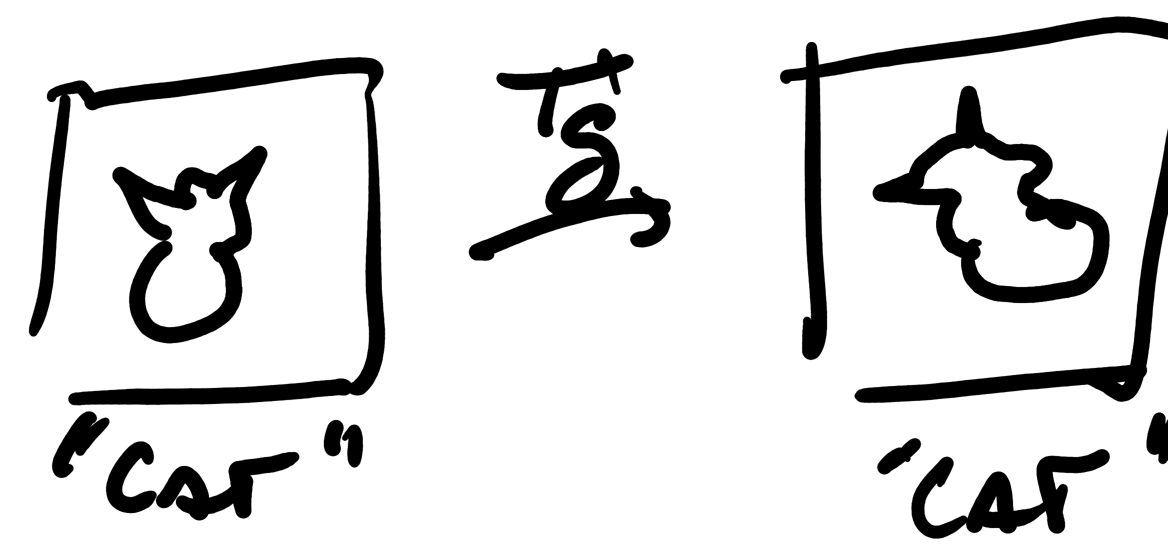
$$f(T_g x) = x \quad \forall g \in G, \ \forall x \in \mathbb{R}^n$$

for a valid action T

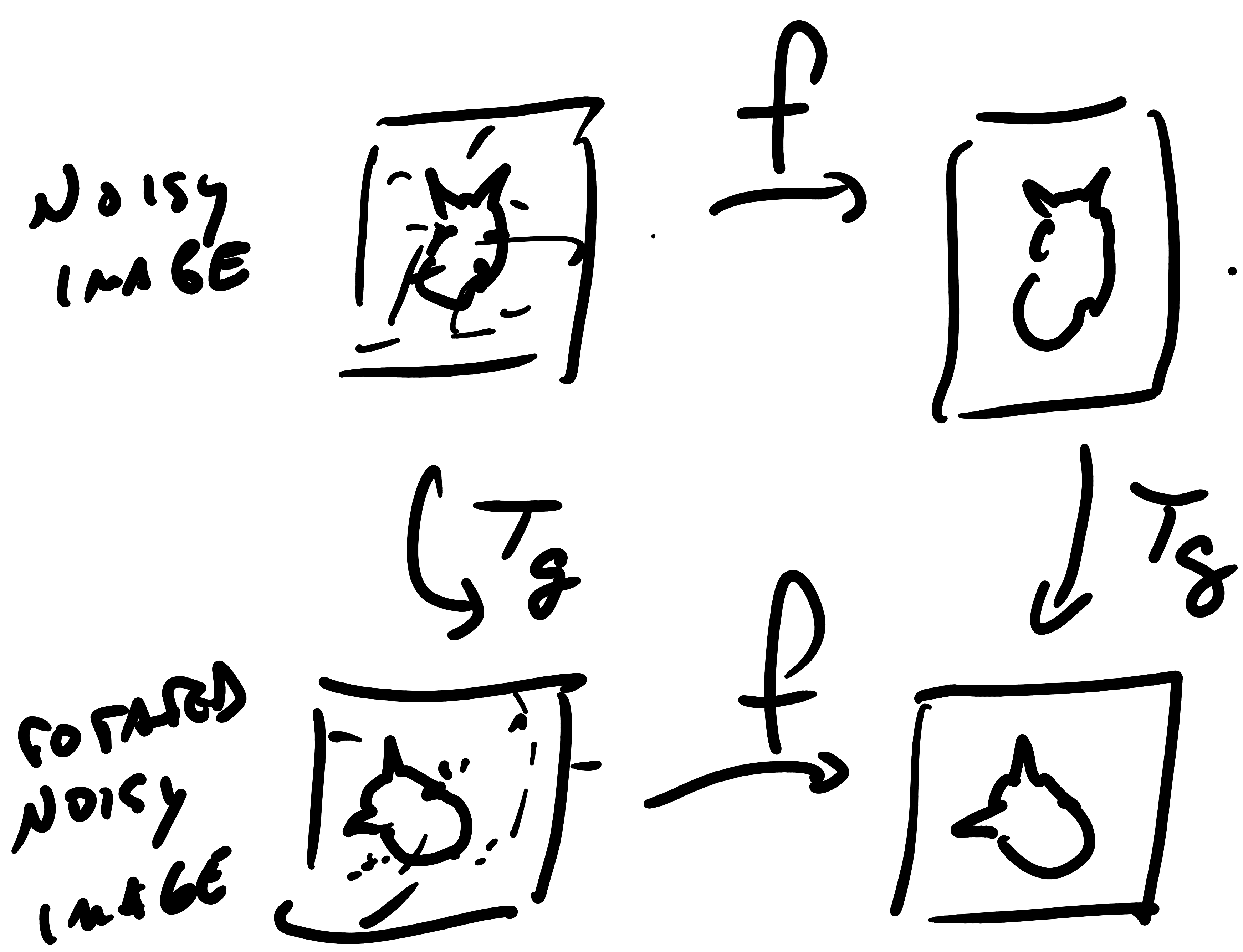
NOTE: this is a special case of equivariance where $\tilde{T}$ is the trivial representation

EXAMPLES IN ML

In classification we look for <u>invariant</u> functions



In regression we look for equivariant functions



DATA AUGMENTATION

$$\arg\min_\theta \sum_{i=1}^{N} \mathcal{L}(f_\theta(x_i), y_i)$$

CLASSICAL EMPIRICAL RISK MINIMIZATION

$$\arg\min_\theta \sum_{i=1}^{N} \underbrace{\sum_{g\in G}}_{\text{GENERALLY HANDLED IN A STOCHASTIC WAY}} \mathcal{L}(f_\theta(T_g x_i), T_g y_i)$$

- DATA AUG. CONSISTENCY

$$\arg\min_\theta \sum_{i=1}^{N} \mathcal{L}(f_\theta(x_i), y_i) + \underbrace{\sum_{g\in G} \mathcal{L}(f_\theta(x_i), f_\theta(T_g x_i))}_{\text{OFTEN USED FOR SELF-SUPERVISED LEARNING?}}$$

Group averaging / ensembling

equivariant
$f: \mathbb{R}^n \to \mathbb{R}^n$

$$f^E(x) = \frac{1}{|G|} \sum_{g \in G} T_g^{-1} f(T_g x)$$

invariant
$f: \mathbb{R}^n \to \mathbb{R}$

$$f^I(x) = \frac{1}{|G|} \sum_{g \in G} f(T_g x)$$

- too expensive for large groups...
- still it is popular in computer vision for 90° rotations + vertical and horiz. flips.

- also called 'Reynolds averaging'.

EQUIVARIANT NETWORKS

→ CONSTRAIN WEIGHTS θ SUCH THAT

f_θ IS G EQUIVARIANT FOR ALL θ.

→ DEEP NETWORKS:

- MAINTAIN EQUIVARIANCE UNTIL THE LAST LAYER + INVARIANCE IN THE FINAL LAYERS.

CNN

CONVOLUTION LAYERS → TRANSLATION EQUIVARIANT
+ POINTWISE NON-LINEARITY

POOLING LAYERS → TRANSLATION INVARIANT

$$x^{\ell+1} = \phi(W x^\ell)$$

EQUIVARIANT LINEAR LAYER.

WHAT IS THE SET
OF EQUIVARIANT MATRICES?

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_c \end{bmatrix}$$

$$S = \{ W \in \mathbb{R}^{m_o \times m_i} : \tilde{T}_g W = W T_g \ \forall g \in G \}$$

EXAMPLE $\tilde{T}_g = T_g$ = Group of SHIFTS.
$m_o = m_i = m$ ~~IN~~ OF m ELEMENTS

$$T_g W = T_g W \iff W = \mathrm{circ}(\omega), \quad \omega \in \mathbb{R}^m$$

$$\dim(S) = m$$

IRREDUCIBLE REPRESENTATIONS:

- A LINEAR REP. IS IRREDUCIBLE iif IT CANNOT BE DECOMPOSED AS

$$T_g = A \begin{bmatrix} \rho_g^1 & \\ & \rho_g^2 \end{bmatrix} A^{-1} \quad \forall g \in G$$

WHERE ρ^1 AND ρ^2 ARE VALID LINEAR REP.

e.g.

ACTION OF $\mathbb{Z}_2$

$$T_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad T_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$T_g = A \begin{bmatrix} 1 & \\ & (-1)^g \end{bmatrix} A^{-1} \quad g = \{0, 1\}$$

CAN BE REDUCED TO

$$T_g = A \begin{bmatrix} \rho^1(g) & \\ & \rho^2(g) \end{bmatrix} A^{-1}$$

WHERE
$$\begin{cases} \rho_1 = 1 \quad \forall g \text{ TRIVIAL REP} \\ \rho^2(g) = (-1)^g \end{cases}$$

THEOREM: A (FINITE) COMPACT GROUP G HAS K **DISTINCT** IRREDUCIBLE REPRESENTATIONS $\{\rho^1, \ldots, \rho^K\}$ SUCH THAT

$$\sum_{i=1}^{K} (\dim \rho^i)^2 = |G|$$

e.g. $\mathbb{Z}_2$ HAS

$$\rho^1(g) = 1$$
$$\rho^2(g) = (-1)^g$$

$$\dim \rho^1 = 1 \qquad |G| = 2 = 1^2 + 1^2$$
$$\dim \rho^2 = 1$$

THEOREM: ANY LINEAR REP. CAN BE DECOMPOSED IN BLOCK DIAGONAL FORM:

$$T_g = A \begin{bmatrix} \rho_1 & & & & & \\ & \ddots & & & & \\ & & \rho_1 & & & \\ & & & \rho_2 & & \\ & & & & \ddots & \\ & & & & & \rho_K \end{bmatrix} A$$

(ρ_1: m_1 TIMES; ρ_2: m_2 TIMES)

WHERE A IS AN ORTHOGONAL BASIS

WHERE ρ_j IS REPEATED m_j TIMES SUCH THAT $\sum m_j \dim \rho_j = n$

EXAMPLE : SHIFT MATRICES (ACTION OF $\mathbb{Z}_m$)

$$T_g = F \begin{bmatrix} \rho_g^1 & & \\ & \ddots & \\ & & \rho_g^m \end{bmatrix} F^*$$

→ F IS THE DISCRETE FOURIER BASIS.

→ $\rho_g^j = e^{-\frac{i 2\pi j}{m} g}$ ARE THE (1-DIM.) IRREPS OF $\mathbb{Z}_m$

THEOREM: EQUIVARIANT MATRICES

$$S = \{ W : \tilde{T}_g W = W T_g \} \text{ WITH } T_g = A^{-1} \begin{bmatrix} \rho_1 & & \\ & \ddots & \\ & & \rho_K \end{bmatrix} A$$

$$\tilde{T}_g = B^{-1} \begin{bmatrix} \rho_1 & & \\ & \ddots & \\ & & \rho_K \end{bmatrix} B$$

if AND ONLY IF WE CAN WRITE W AS

LINEAR SUBSPACE OF $\mathbb{R}^{m_o \times m_i}$

$$W = A^{-1} \begin{bmatrix} \psi_1 & & \\ & \ddots & \\ & & \psi_K \end{bmatrix} B$$

WHERE $\psi_j \in \mathbb{C}^{\tilde{m}_j \dim \rho_j \times m_j \dim \rho_j}$

$$\psi_j = \begin{bmatrix} \tilde{\psi}_j & & \\ & \ddots & \\ & & \tilde{\psi}_j \end{bmatrix} \quad \text{where } \tilde{\psi}_j \in \mathbb{C}^{\tilde{m}_j \times m_j} \text{ are free parameters.}$$

$$\Rightarrow \dim S = \text{DEGREES OF FREEDOM} = \sum_{j=1}^{K} \tilde{m}_j m_j \leq m_i m_o$$

EXAMPLE: T_g AND $\tilde{T}_g$ ARE SHIFTS IN $\mathbb{R}^n$

$$W = F \begin{bmatrix} \omega_1 & & \\ & \ddots & \\ & & \omega_n \end{bmatrix} F^*$$

$$= \operatorname{circ}(\hat{\omega}) \quad \text{CIRCULANT MATRICES!}$$

n dof!

$\Rightarrow$ RESTRICTING THE NETWORK TO HAVE G-EQUIVARIANT LAYERS REDUCES THE NUMBER OF LEARNABLE PARAMETERS.

$\Rightarrow$ AN EQUIVARIANT NN JUST REQUIRES DEFINING THE IRREPS AND MULTIPLICITIES AT EACH LAYER

$\Rightarrow$ $W^E = \frac{1}{|G|} \sum_g T_g W T_g^{-1}$ is just an orthogonal projection onto S

NON-LINEARITIES:

- IF THE GROUP ACTION AT LAYER L IS REPRESENTED BY PERMUTATION MATRICES: (FOR EXAMPLE, THE REGULAR REPRESENTATION)

$$[T_g x]_s = x_{g \cdot s}$$

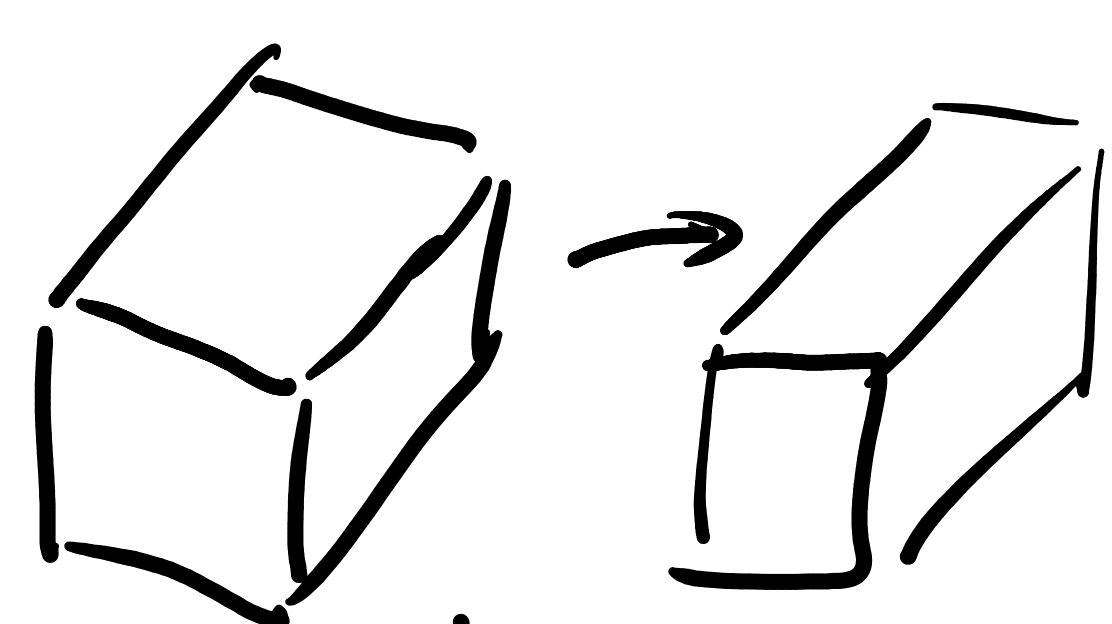
AN ELEMENTWISE NON-LINEARITY IS EQUIVARIANT:
$\phi : \mathbb{R} \to \mathbb{R}$

$$\phi(T_g x) = T_g \phi(x)$$

EXAMPLE CNNs

$$x^{\ell} = \phi(W^{\ell} x^{\ell-1})$$

$$x^{\ell} = \begin{bmatrix} x_1^{\ell} \\ \vdots \\ x_C^{\ell} \end{bmatrix} \begin{array}{l} \to \text{CHANNEL 1} \\ \\ \to \text{CHANNEL C} \end{array}$$



C×H×W   C×H×W

$$T_g^{\ell} = \begin{bmatrix} T_g & & \\ & \ddots & \\ & & T_g \end{bmatrix}$$ → SHIFT MATRICES.

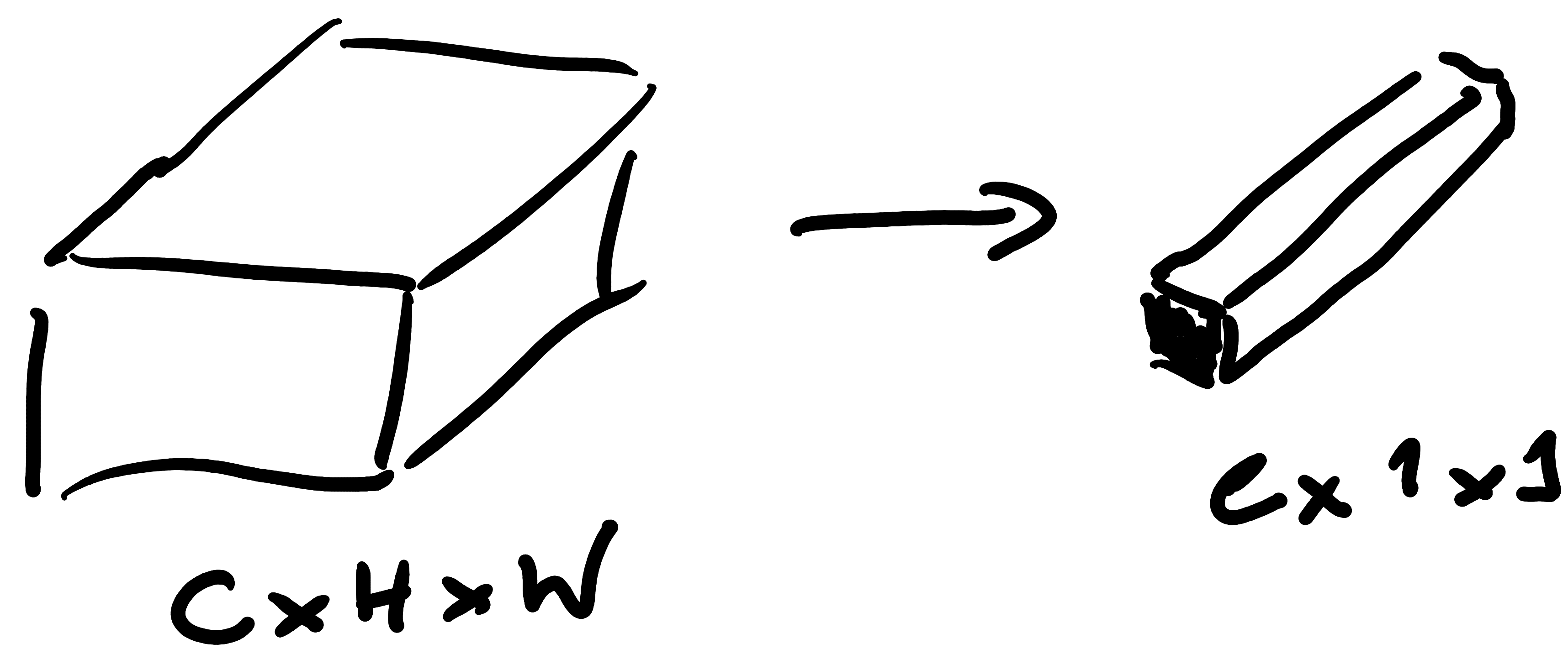
$$W^{\ell} = \begin{bmatrix} \mathrm{circ}\, W_{11} & \cdots & \mathrm{circ}\, W_{1C} \\ \vdots & \ddots & \\ \mathrm{circ}\, W_{C1} & & \mathrm{circ}\, W_{CC} \end{bmatrix}$$

INVARIANT LAYERS

↳ POOLING

OUTPUT ACTION = TRIVIAL ACTION

$$\underbrace{x^{\ell}}_{\in \mathbb{R}^{C}} = [1, \dots, 1] \underbrace{x^{\ell-1}}_{\in \mathbb{R}^{mC}}$$

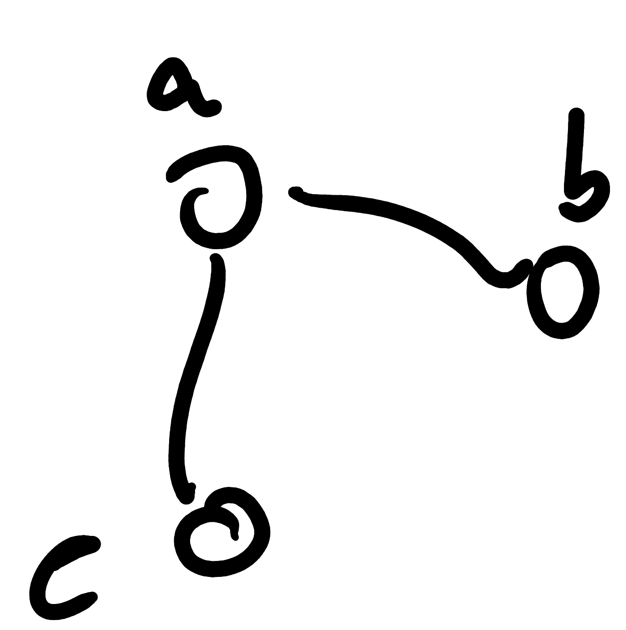


EXAMPLES:

- STEERABLE CNNs : EQUIVARIANCE TO 90° ROTATIONS AND HORIZONTAL AND VERTICAL FLIPS. (HOMEWORK).

- GRAPH NEURAL NETWORKS

GRAPHS ARE REPRESENTED BY ADJACENCY MATRICES



$$A = \begin{array}{c|ccc} & a & b & c \\ \hline a & 0 & 1 & 1 \\ b & 1 & 0 & 0 \\ c & 1 & 0 & 0 \end{array}$$

$$\tilde{A} = \begin{array}{c|ccc} & b & c & a \\ \hline b & 0 & 0 & 1 \\ c & 0 & 0 & 1 \\ a & 1 & 1 & 0 \end{array}$$

THEY REPRESENT THE SAME GRAPH

ANY PERMUTATION MATRIX T_g $g \in S_n$ DEFINES THE SAME GRAPH.

$$A' = T_g A T_g^T$$

GRAPH CLASSIFICATION PROBLEM:

LEARN $f(A) =$ GRAPH LABEL.

f CAN BE DEFINED VIA PERMUTATION INVARIANT LAYERS.

WE CAN WRITE THE PROBLEM IN VECTOR FORM

$$\underbrace{\mathrm{vec}(A')}_{A \in \mathbb{R}^{n^2}} = \underbrace{T_g \otimes T_g}_{\tilde{T}_g \text{ group action on } \mathbb{R}^{n^2}} \underbrace{\mathrm{vec}(A)}_{a \in \mathbb{R}^{n^2}}$$

$$\Longleftrightarrow \quad a' = \tilde{T}_g a$$

SET OF PERMUTATION INVARIANT FUNCTIONS

POINT CLOUDS

$$x = \begin{bmatrix} p_1^T \\ \vdots \\ p_m^T \end{bmatrix} \in \mathbb{R}^{m \times 3}$$ SET OF POINTS $p_i \in \mathbb{R}^3$

$$x \in \mathbb{R}^{m \times 3}$$

GROUP ACTIONS:

- PERMUTATIONS OF POINTS S_m

$P_g W(x) R_g = W(P_g x R_g)$

- GLOBAL ROTATION + TRANSLATION OF POINTS.

$SO(3)$ $(\mathbb{R}^3, +)$

THIS IDEAS CAN BE EXTENDED TO ANY TYPE OF DATA AS LONG AS WE KNOW WHICH

EXAMPLES

- DATA ON A SPHERE (e.g. GLOBAL CLIMATE DATA)
- SETS
- TIME SERIES
- DATA ON GENERAL MANIFOLDS.

APPENDIX

2D TRANSLATIONS:

m × m PIXEL IMAGE

$$\mathbb{Z}_m \times \mathbb{Z}_m$$

$$(g,s) \longrightarrow \overbrace{\tilde{T}_{g,s}}^{m^2\times m^2} = \underbrace{\overbrace{T_g}^{g \text{ vertical shifts}}}_{m\times m} \otimes \underbrace{\overbrace{T_s}^{s \text{ horizontal shifts.}}}_{m\times m}$$

where

$$T_g = F \begin{bmatrix} e^{-\frac{i2\pi g}{m}} & & \\ & \ddots & \\ & & e^{-\frac{i2\pi g m}{m}} \end{bmatrix} F^*$$

$$\Rightarrow \tilde{T}_{g,s} = \underbrace{(F \otimes F)}_{\tilde{F}} \begin{bmatrix} e^{-\frac{i2\pi g s}{m^2}} & & \\ & \ddots & \\ & & e^{-\frac{i2\pi m^2}{m^2}} \end{bmatrix} \underbrace{(F^* \otimes F^*)}_{\tilde{F}^*}$$

2D Fourier transform!

$$= \tilde{F} \begin{bmatrix} e^{-\frac{i2\pi (g\cdot 1)(s\cdot 1)}{m^2}} & & \\ & \ddots & \\ & & e^{-\frac{i2\pi (gm)(sm)}{m^2}} \end{bmatrix} \tilde{F}^*$$

Kron prop.

$$(AC)\otimes(BD) = (A\otimes B)(C\otimes D)$$

$\Rightarrow$ irreps: $\rho^{j,k}_{g,s} = e^{-\frac{i2\pi(jg+ks)}{m^2}}$

$j = 0, \dots, m-1$

$k = 0, \dots, m-1$

THEY ALL HAVE MULTIPLICITY 1!

2D TRANSLATIONS

$[0, m-1) \times [0, m-1)$ $\quad g \in [0, m-1)$ VERTICAL TRANSLATION

$s \in [0, m-1)$ HORIZONTAL TRANSLATIONS.

$$g_1 \odot g_2 = \mathrm{mod}(g_1 + g_2, m-1)$$

$$s_1 \odot s_2 = \mathrm{mod}(s_1 + s_2, m-1)$$

$$\Rightarrow \quad \tilde{T}_{g,s} = \tilde{F} \begin{bmatrix} e^{-\frac{i2\pi(g+s)}{m^2}} & & \\ & \ddots & \\ & & e^{-\frac{i2\pi(g+s)k}{m^2}} \end{bmatrix} \tilde{F}^*$$

irreps $\qquad$ THEY ALL HAVE MULTIPLICITY ONE?

$$\rho^{j,k}_{s,g} = e^{-\frac{i2\pi(g+s)(j+k)}{m^2}}$$

$j = 0, \dots, m-1$ $\qquad \Rightarrow$ SAME AS WITH SHIFTS.

$k = 0, \dots, m-1$

2D SHIFTS + 90° ROTATIONS

GROUP $\mathbb{Z}_m \times \mathbb{Z}_m \times \mathbb{Z}_4$

WHY 90° ROTATIONS? ROTATIONS BY $\alpha \in (0, 90°)$ REQUIRE INTERPOLATION

$\Downarrow$

T_g IS NOT INVERTIBLE....
AND THUS NOT
THE ACTION OF A GROUP?

$$\tilde{T}_{g,s,p} = R_p \tilde{T}_{g,s}$$

NON COMMUTATIVE ACTION $\Rightarrow$ THERE ARE SOME IRREPS WITH $\dim > 1$

WHERE R_p IS THE ROTATION

FOR $m=2$

$$R_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} = B \begin{bmatrix} e^{-i\pi} & & \\ & & \\ & & \end{bmatrix} B^{-1}$$

$$\begin{array}{|c|c|} a & b \\ \hline c & d \end{array} \xrightarrow[R_1]{} \begin{array}{|c|c|} c & a \\ \hline d & b \end{array}$$

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \xrightarrow[R_1]{} \begin{bmatrix} c \\ a \\ d \\ b \end{bmatrix}$$

$$R_2 = R_1^2$$

$$R_3 = R_1^3$$

$$R_4 = R_1^4 = I$$

A ROTATION EQUIVARIANT OPERATOR

IDEAS FOR DM

- COMPUTE THE IRREDUCIBLES AND MULTIPLICITIES.
- $S_1 = \{ W \in \mathbb{R}^{m \times m} : \tilde{T}_g W = W \tilde{T}_g \}$

  $S_2 = \{ W \in \mathbb{R}^{m \times m} : R_g W = W R_g \}$

  $\Rightarrow S = S_1 \cap S_2$ EQUIVARIANT TO BOTH ACTIONS.

- APPLY REYNOLDS AVERAGING TO PROJECT ONTO S_2 (or S_1)

$$= \frac{1}{|G|} \sum_{g \in G} R_g^T W R_g$$

↳ YOU CAN TRY THIS OUT COMPUTATIONALLY

- THINK INTUITIVELY ABOUT THE FAMILY OF OPERATORS THAT YOU GET.