COMP760 Week 10 - Equivariant Networks Part II

Joey Bose and Prakash Panangaden

Group Actions

1. We have a set $\mathscr{X}$ and $f : \mathscr{X} \to \mathbb{C}$

2. Group G acts on $\mathscr{X}$

$$T_g : \mathscr{X} \to \mathscr{X} \quad \forall g \in G$$

$$\forall g1, g2 \in G,\ T_{g2g1} : T_{g2} \circ T_{g1}$$

If $\mathscr{X}$ is a (finite) Vector Space then $T_g \in GL(n)$

3. Extending the action to functions

$$\mathbb{T}_g : f \to f' \qquad f'(T_g(x)) = f(x)$$

<u>Groups</u>

1. $e \in G$

   Identity

2. $(a \circ b) \circ c = a \circ (b \circ c)$

   Associativity

3. $\forall a \in G \quad \exists b \in G$

   $a \circ b = e$

   Unique Inverses

Equivariance

$$\begin{tikzcd} L_{(V_1)}(\mathscr{X}_1) \arrow[r, "\mathbb{T}_g"] \arrow[d, "\phi"'] & L_{(V_1)}(\mathscr{X}_1) \arrow[d, "\phi"] \\ L_{(V_2)}(\mathscr{X}_2) \arrow[r, "\mathbb{T}'_g"] & L_{(V_2)}(\mathscr{X}_2) \end{tikzcd}$$

Equivariance Networks Recipe

$$
\begin{tikzcd}
{} & {} \\
\mathcal{X}_2 \arrow[u, "f_2"] \arrow[r, "\mathbb{T}_g^2"] & \mathcal{X}_2 \arrow[u, "f_2"] \\
\mathcal{X}_1 \arrow[u, "f_1"] \arrow[r, "\mathbb{T}_g^1"] & \mathcal{X}_1 \arrow[u, "f_1"] \\
\mathcal{X}_0 \arrow[u, "f_0"] \arrow[r, "\mathbb{T}_g^0"] & \mathcal{X}_0 \arrow[u, "f_0"]
\end{tikzcd}
$$

Equivariance from Invariance:

Lemma: Let $f : \mathbb{R}^d \to \mathbb{R}$ be invariant with respect to G and assume that R_g is orthogonal for $g \in G$. Then $\nabla_u f(u)$ is equivariant with respect to G

$Proof$: Read Equivariance Section in Normalizing Flow Review Paper by Papamakarios et. Al 2019

Representation Theory Perspective

Group Elements

$$R(g_2 g_1) = R(g_2) R(g_1)$$

Matrices! Matrix Multiplication is the Group Operation

Linear Actions

1. G acts on a set S by $x \mapsto R_g(x)$

2. $f : S \to \mathbb{R}$ is a function on S and we want to learn $f \mapsto h(f)$

3. The induced action $\mathbb{T}^{ind}$ on the function space $f \mapsto f'$ where $f'(x) = f(T_g^{-1}(x))$

4. The induced action is $\mathbb{T}^{ind} : L(S) \to L(S)$ is automatically linear

Basic facts on Representations

1. Two representation R and R' are said to be equivalent if

$$R'(g) = UR(g)U^\dagger \text{ for some Unitary Matrix } U$$

2. A representation R is said to be (completely) reducible if

$$R(g) = U \left(\begin{array}{c|c} R_1(g) & \\ \hline & R_2(g) \end{array} \right) U^\dagger$$

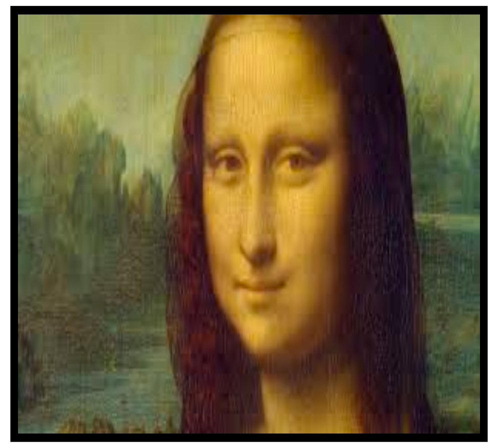
Complete Reducibility

Theorem. Let R be a representation of a compact group G on a vector space V. If R fixes the subspace W, then it also fixes $W^{\perp}$.

$$R(g) = U \left(\begin{array}{c|c} R_1(g) & B(g) \\ \hline & R_2(g) \end{array} \right) U^{\dagger} \qquad \Longrightarrow \qquad B(g) = 0$$

Corollary. Any representation of a compact group is reducible into a direct sum of irreducible representations. This is Maschke's Theorem if the group is finite, and Peter-Weyl (part 2) for continuous.

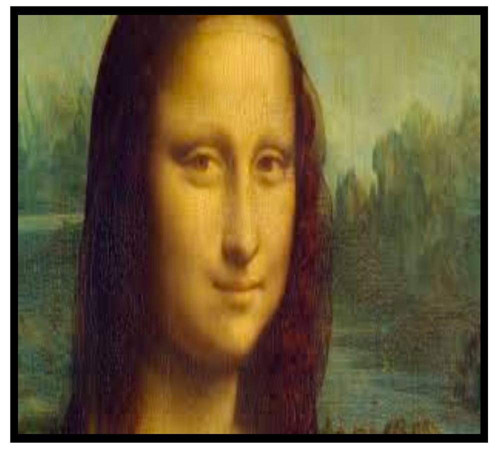
Example: The dihedral group D_4

e 

r 

r^2 

r^3 

s 

rs 

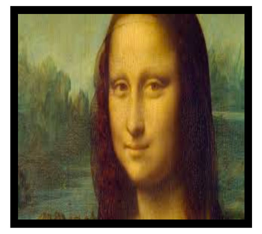
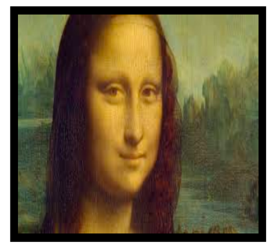
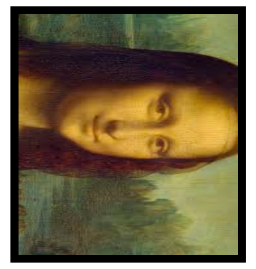
r^2s 

r^3s 

$$r^4 = e$$

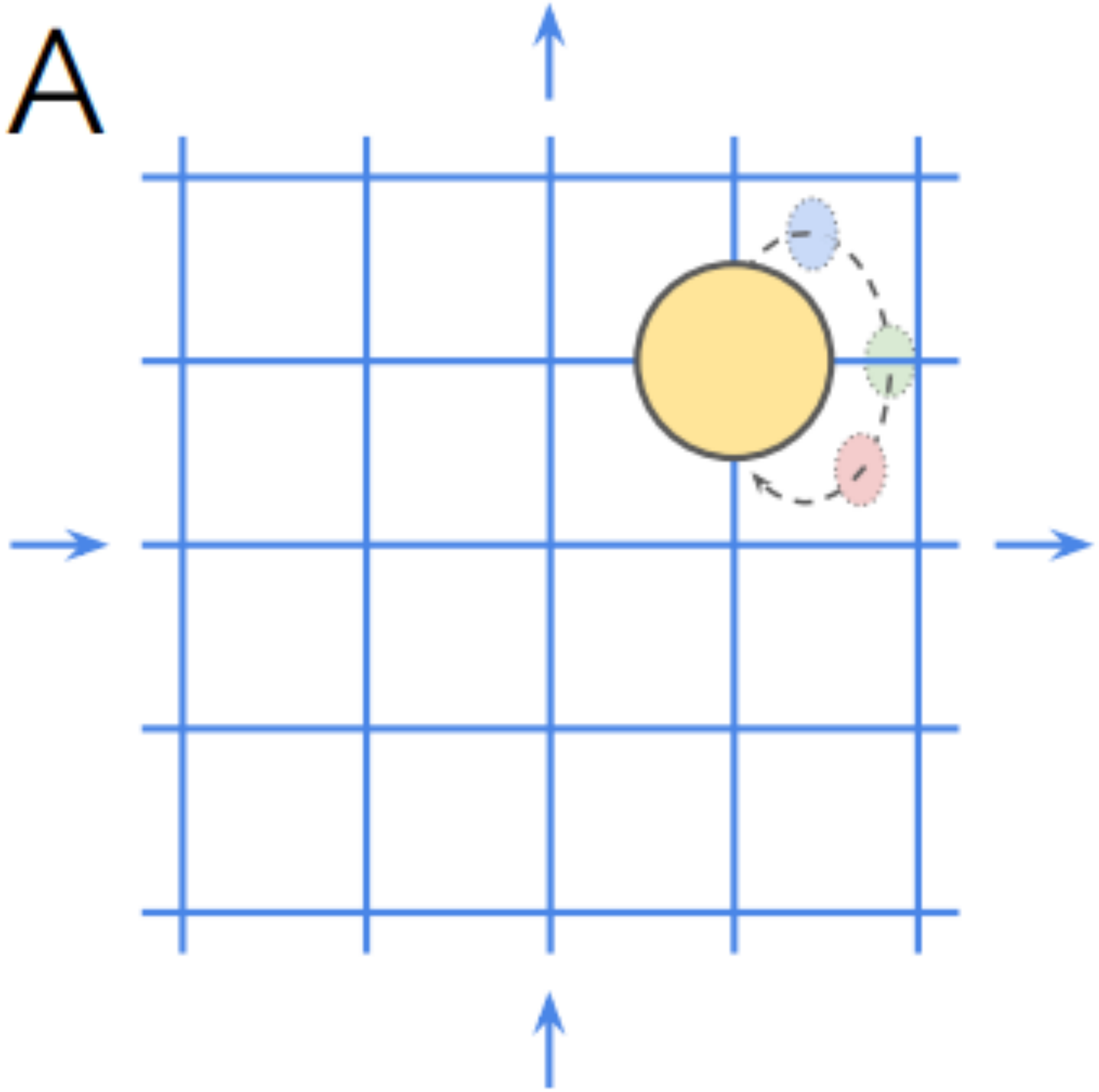
$$s^2 = e$$

$$srs = r^{-1}$$

		R_0	R_1	R_2	R_3	R_4
e		(1)	(1)	(1)	(1)	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
r		(1)	(1)	(-1)	(-1)	$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$
r^2		(1)	(1)	(1)	(1)	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$
r^3		(1)	(1)	(-1)	(-1)	$\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$
s		(1)	(-1)	(1)	(-1)	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
rs		(1)	(-1)	(-1)	(1)	$\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$
r^2s		(1)	(-1)	(1)	(-1)	$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$
r^3s		(1)	(-1)	(-1)	(1)	$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$

Application to ML: Disentanglement

Definition (Informal Higgins et. Al 2018): A vector representation is called a disentangled representation to a particular decomposition of a symmetry group into subgroups, if it decomposes into independent subspaces, where each subspace is affected by the action of a single subgroup and the action of all other subgroups leave the subspace unaffected.

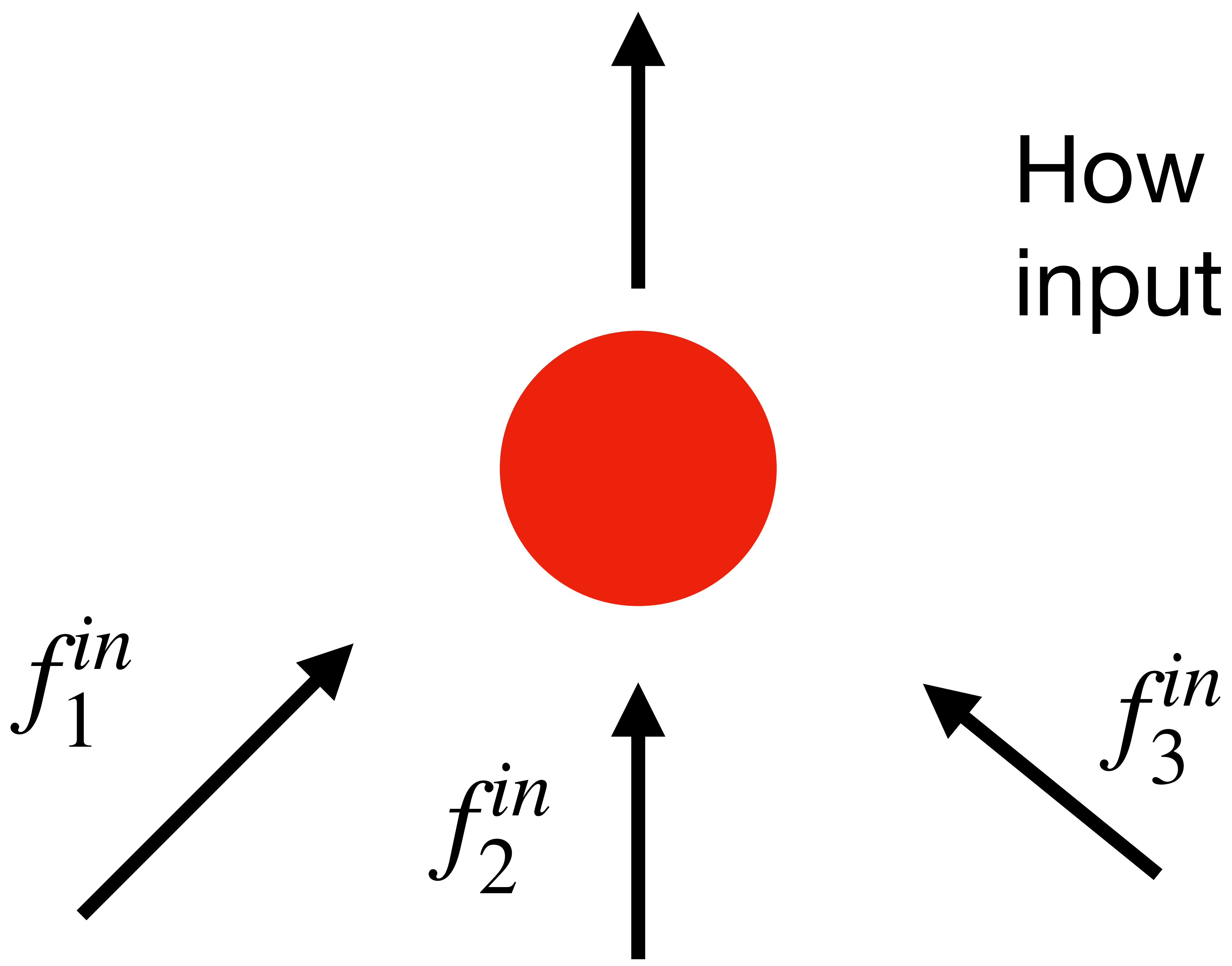


Schur's Lemma

Theorem: Let $\phi : v \mapsto h$ be a linear map and assume that v transforms according to irrep R of a compact group G while w transforms according to irrep R'. Then either $R = R'$ and ϕ is a multiple of the identity map or $\phi = 0$.

Designing an Equivariant Neuron

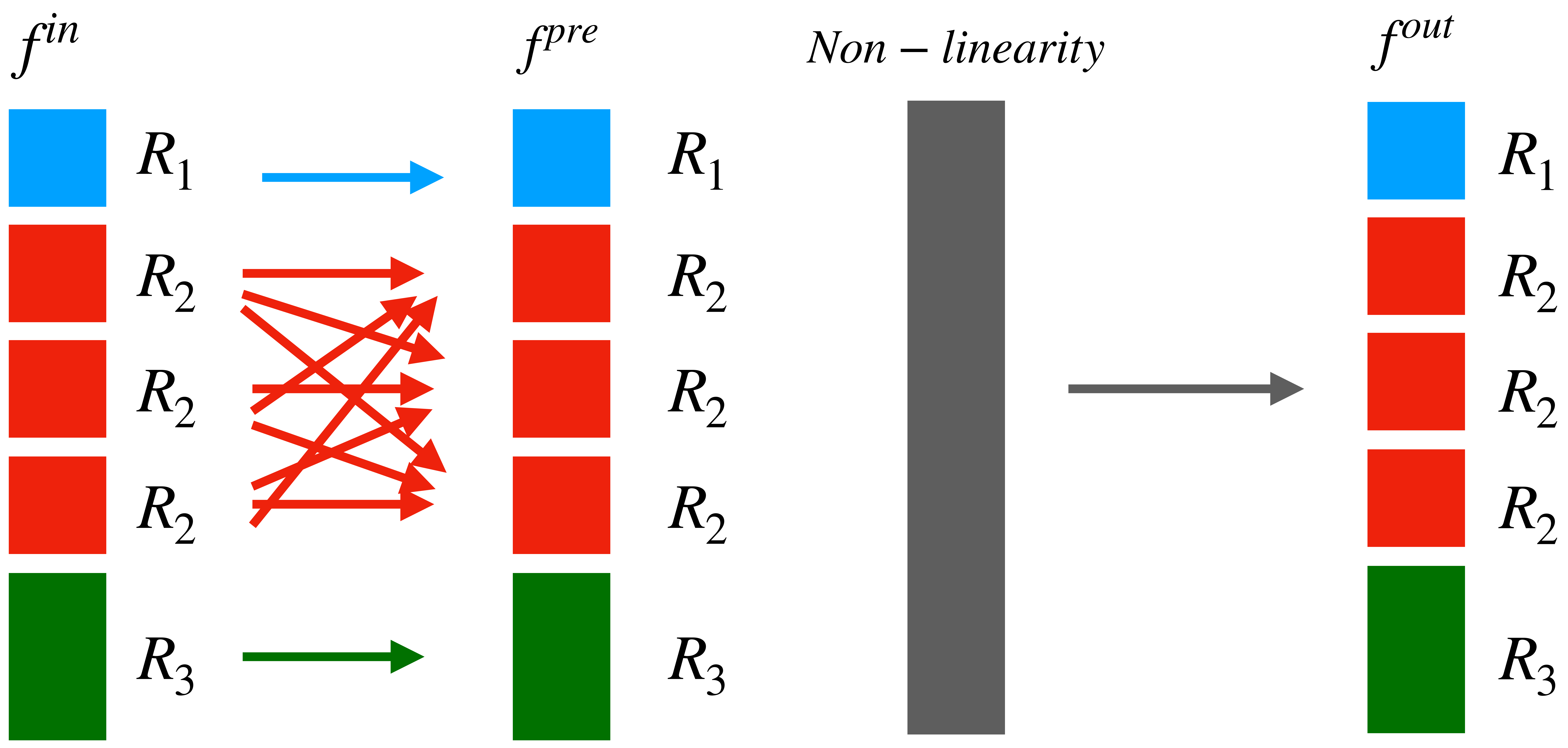
$$f^{out} \mapsto R^{out}(g) f^{out}$$



How does the output transform when the input transforms under irreps?

$$f_1^{in} \mapsto R_1^{in}(g) f_1^{in} \qquad f_2^{in} \mapsto R_2^{in}(g) f_2^{in} \qquad f_3^{in} \mapsto R_3^{in}(g) f_3^{in}$$

Designing an Equivariant Neuron



Equivariant Linear Part:

Theorem. For each irrep R_i we may concatenate the parts of the incoming activations transforming according to R_i into a matrix F_i^{in}. Then the preactivation in a neutron is equivariant iff it is of the form

$$F_i^{pre} = F_i^{in} W_i$$

For learnable weight matrices $W_0, W_1, \cdots$

Equivariant Non-Linearity

1. Express the pre-activations as a function G and apply a point wise non-linearity and transform back.

2. Derive a new non-linear transformation directly expressed using the irreducible parts of the pre-activation.

Tensor product Non-Linearity

$$v_a \mapsto R_a(g)v_a$$

$$v_b \mapsto R_a(g)v_b$$

Must be decomposed into an irrep

$$v_a \otimes v_b \mapsto \boxed{(R_a \otimes R_b)}(g)\boxed{(v_a \otimes v_b)}$$

Representation, but not an irrep