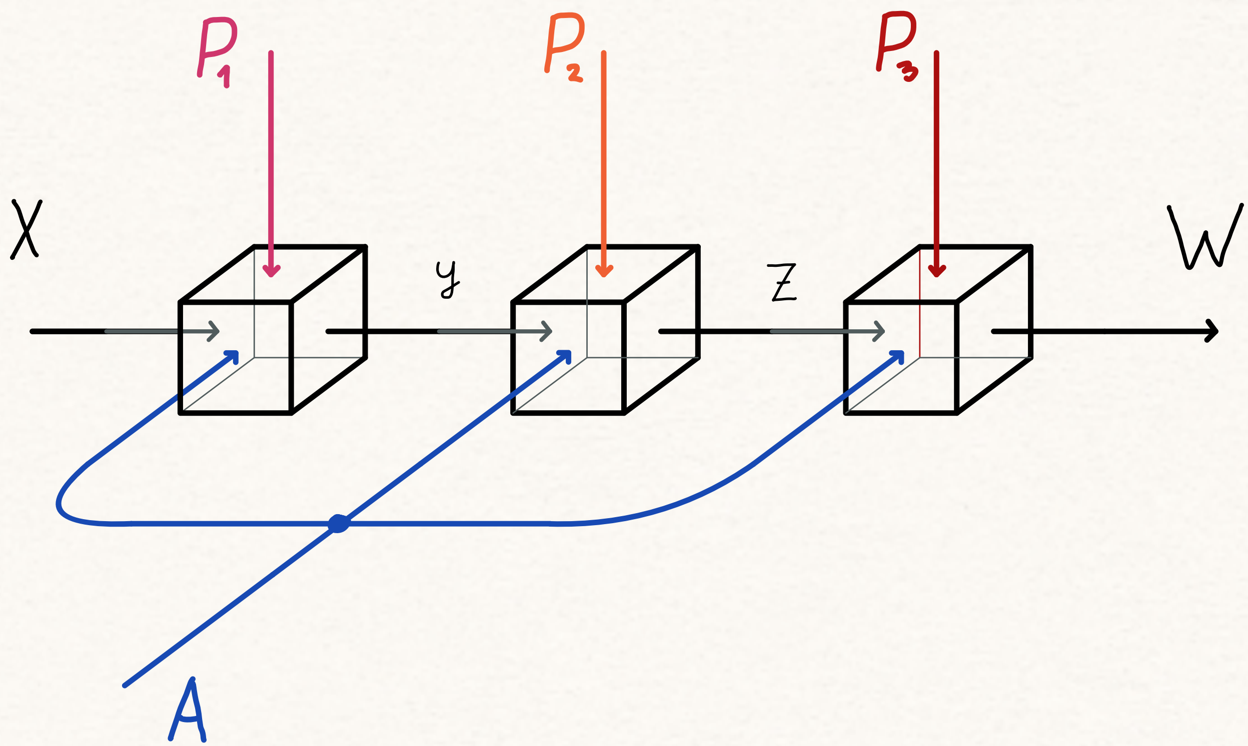
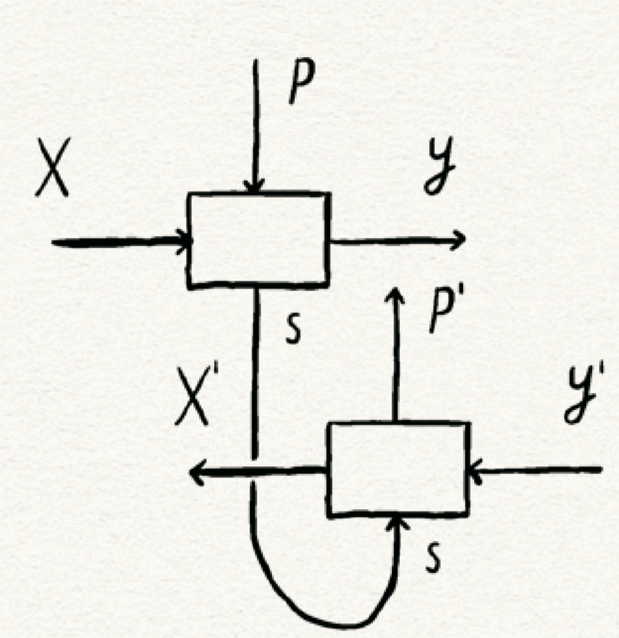
GCNNs AS PARAMETRIC COKLEISLI MORPHISMS

BRUNO GAVRANOVIĆ, MATTIA VILLANI



GOAL:

PROVIDE A CATEGORICAL FRAMEWORK



FOR DEEP LEARNING

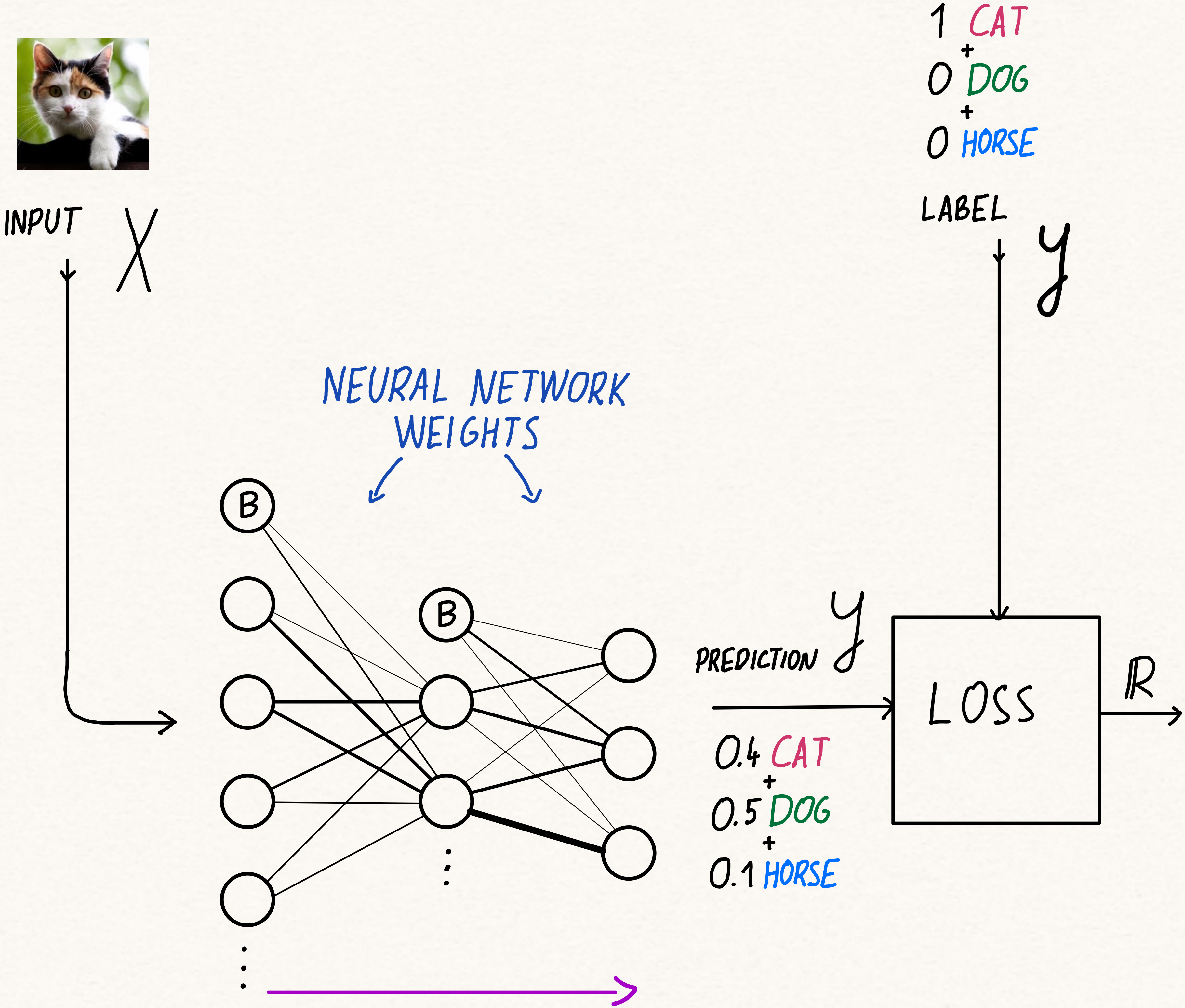
SUPERVISED LEARNING WITH NEURAL NETWORKS

IN ONE SLIDE:

$$x \longrightarrow \boxed{?} \longrightarrow y$$

TASK: FIND A FUNCTION $X \to Y$ THAT BEST FITS

A DATASET: List $X \times Y$


1 CAT
+
0 DOG
+
0 HORSE
LABEL
y
INPUT
X
NEURAL NETWORK
WEIGHTS
B
B
PREDICTION
y
LOSS
R
0.4 CAT
+
0.5 DOG
+
0.1 HORSE

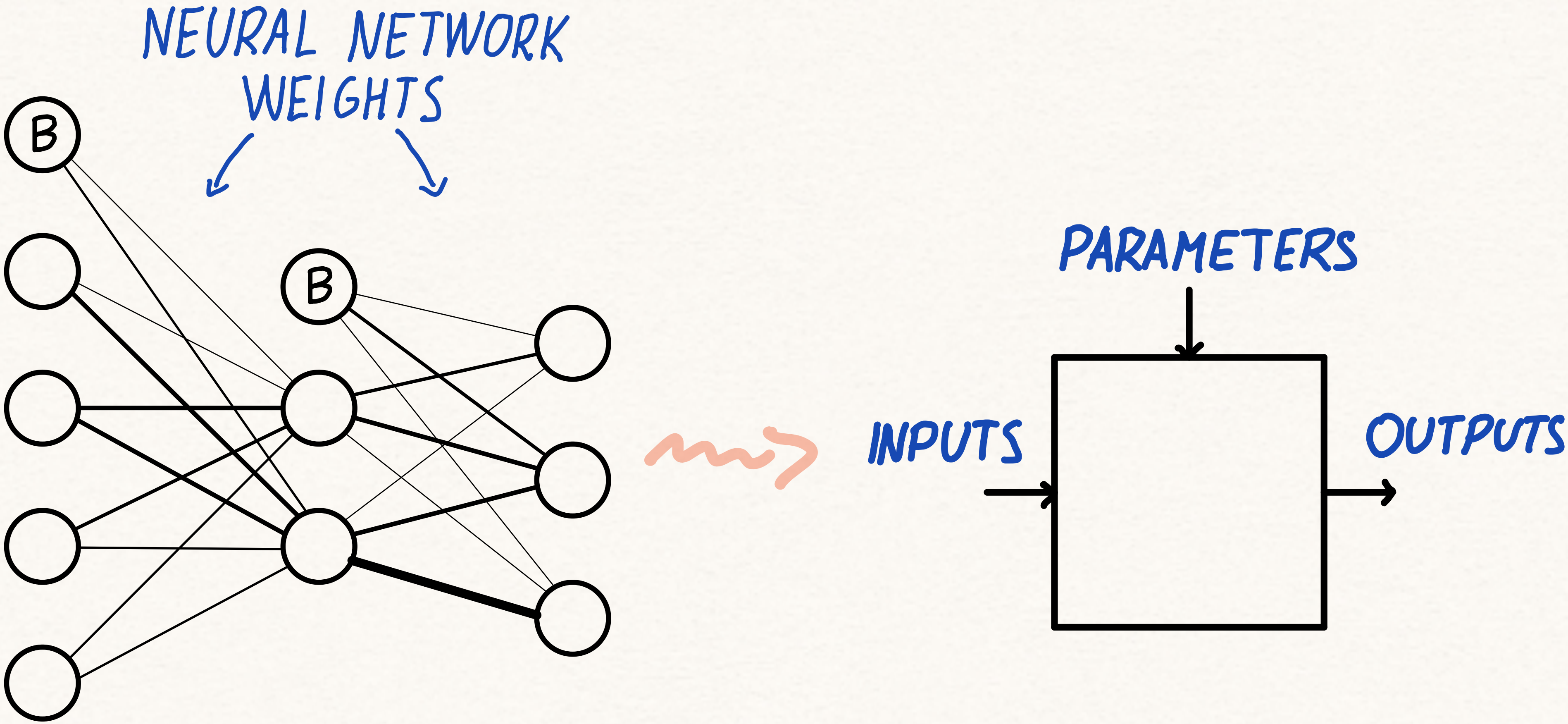
WE ALREADY HAVE A FRAMEWORK:

Categorical Foundations of Gradient-Based Learning

G.S.H. CRUTTWELL, Mount Allison University, Canada
BRUNO GAVRANOVIĆ and NEIL GHANI, University of Strathclyde, UK
PAUL WILSON and FABIO ZANASI, University College London, UK

We propose a categorical semantics of gradient-based machine learning algorithms in terms of lenses, parametrised maps, and reverse derivative categories. This foundation provides a powerful explanatory and unifying framework: it encompasses a variety of gradient descent algorithms such as ADAM, AdaGrad, and Nesterov momentum, as well as a variety of loss functions such as as MSE and Softmax cross-entropy, shedding new light on their similarities and differences. Our approach to gradient-based learning has examples generalising beyond the familiar continuous domains (modelled in categories of smooth maps) and can be realized in the discrete setting of boolean circuits. Finally, we demonstrate the practical significance of our framework with an implementation in Python.

PARA CONSTRUCTION



BASIC NN LAYER

$$f: \mathbb{R}^{k} \longrightarrow \mathbb{R}^{k'}$$
$$f(X) = \sigma(XW)$$

$$\begin{array}{cc} & k \\ 1 & \begin{bmatrix} \quad X \quad \end{bmatrix} \end{array} \quad \begin{array}{cc} & k' \\ k & \begin{bmatrix} \quad W \quad \\ \\ \end{bmatrix} \end{array}$$

k - NUMBER OF INCOMING FEATURES

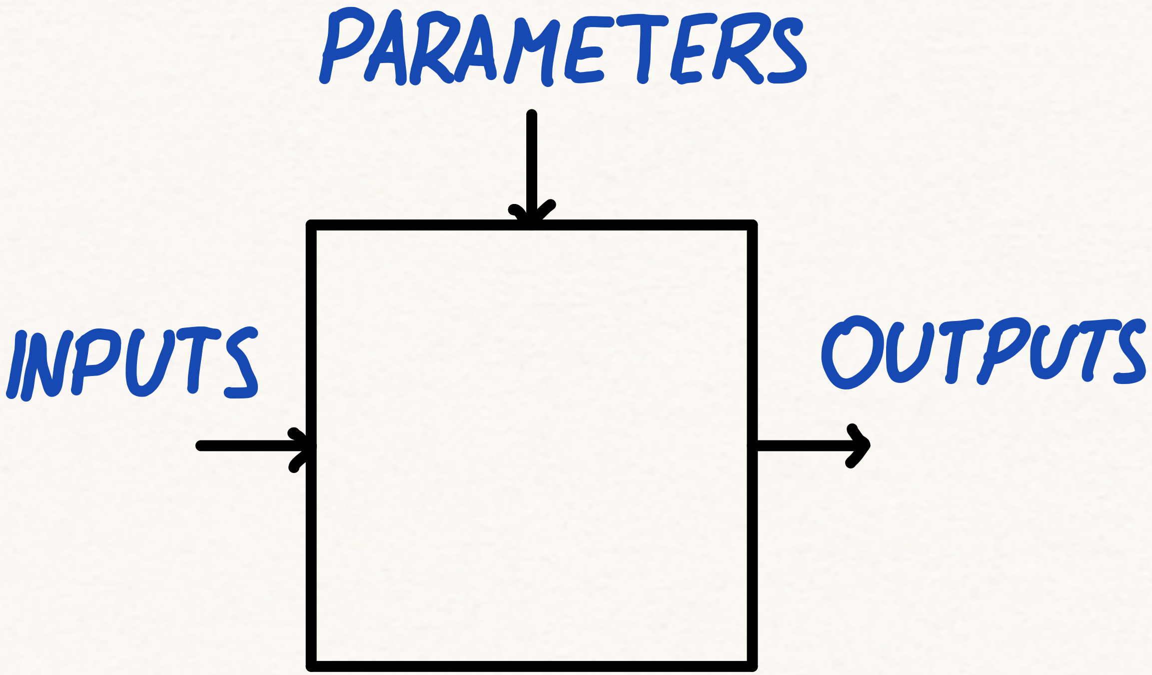
k' - NUMBER OF OUTGOING FEATURES

<u>EACH LAYER HAS ITS OWN WEIGHT MATRIX</u>

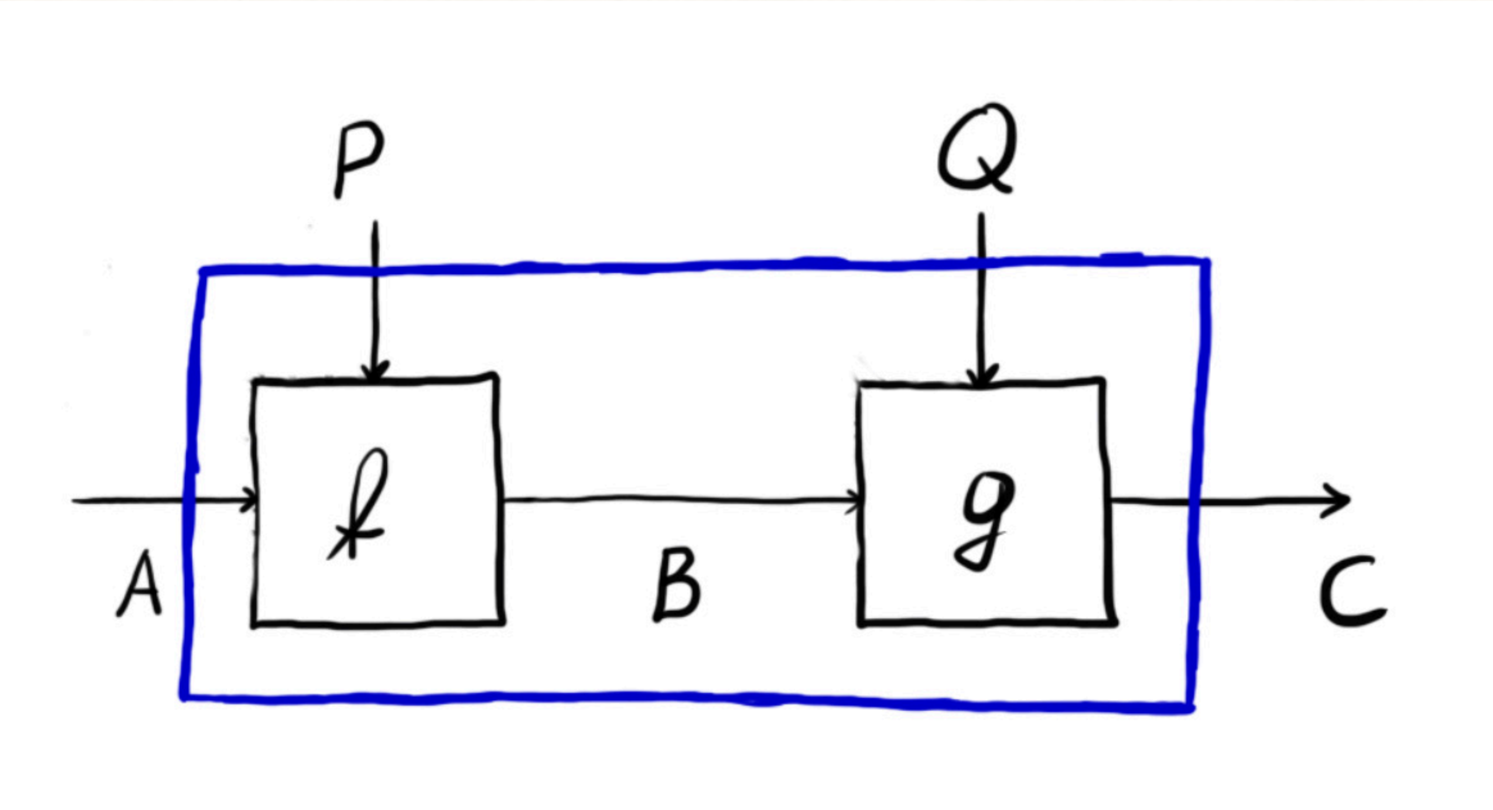
FIX A MONOIDAL CATEGORY $(\mathcal{C}, \otimes, I)$.

$\mathbf{Para}(\mathcal{C})$ IS A BICATEGORY WHERE

$$\mathbf{Para}(\mathcal{C})(A, B) := \sum_{P : \mathcal{C}} \mathcal{C}(P \otimes A, B)$$



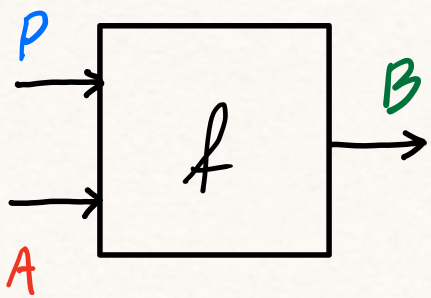
COMPOSITION TENSORS THE PARAMETERS

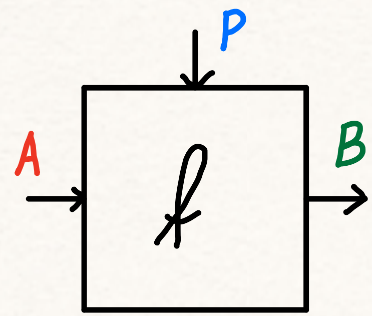


GRAPHICAL LANGUAGE

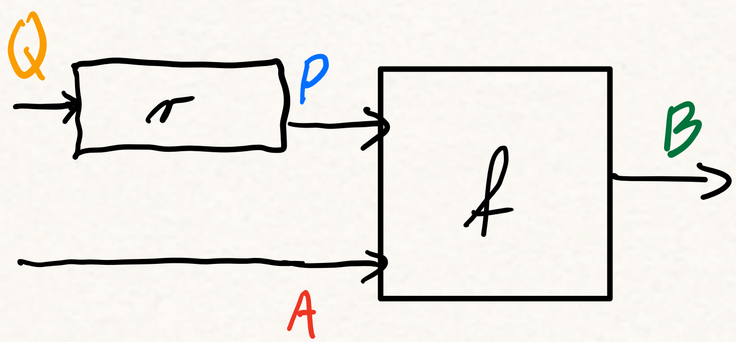
TEXTUAL NOTATION | **STANDARD STRING DIAGRAM** | **'2D' STRING DIAGRAM**

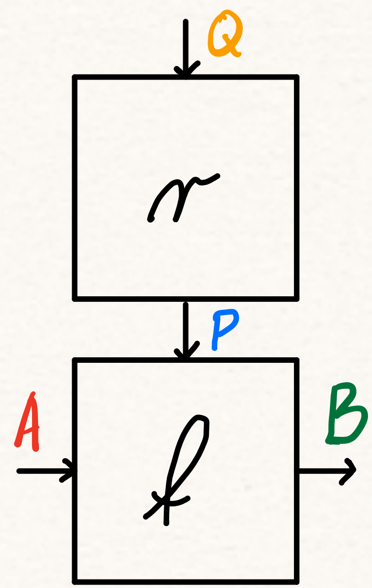
$$f : P \otimes A \longrightarrow B$$



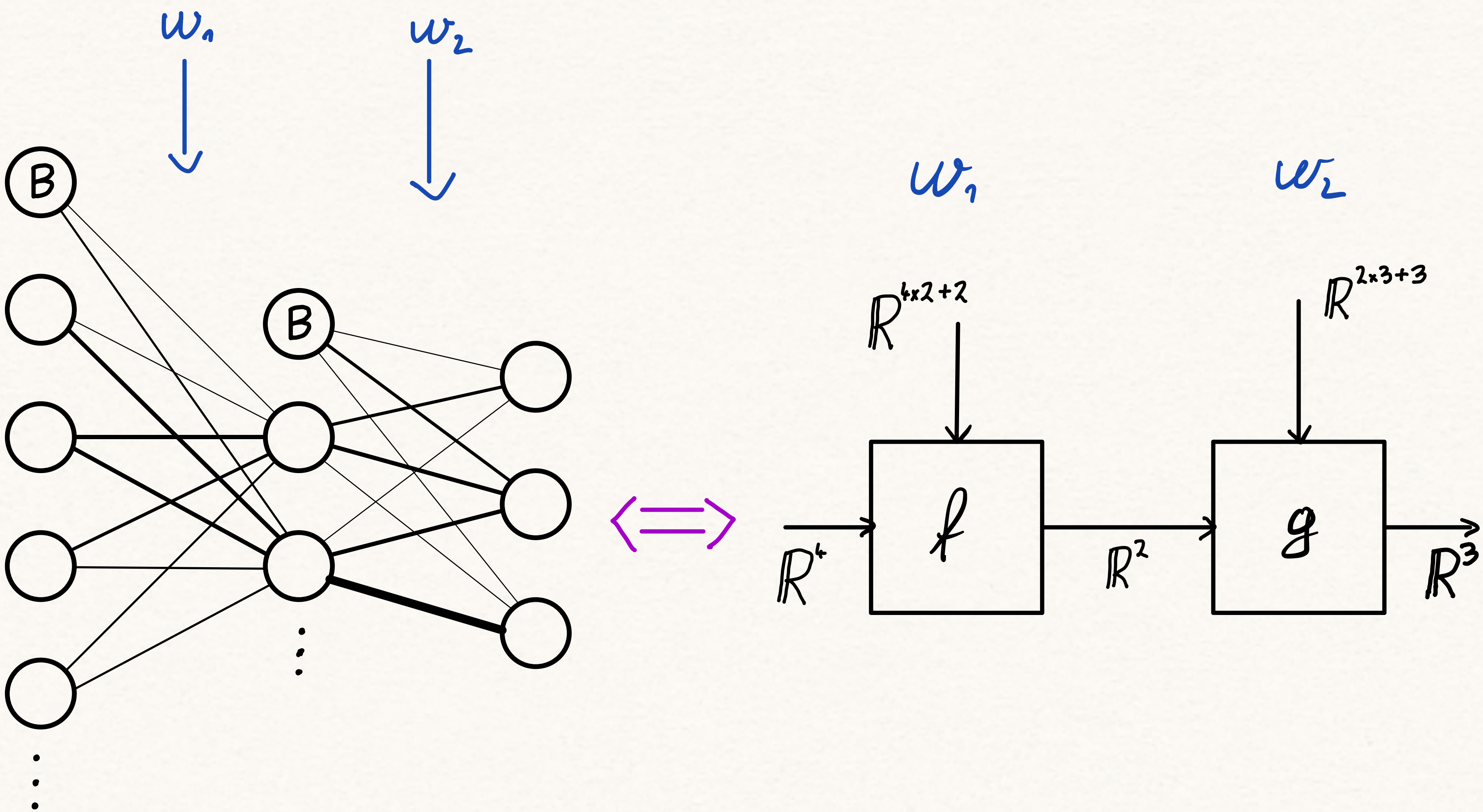


$$\begin{tikzcd} A \arrow[rr, bend left, "{(P, f)}"] \arrow[rr, bend right, "{(Q, g)}"'] \arrow[rr, phantom, "\Downarrow r"] & & B \end{tikzcd}$$

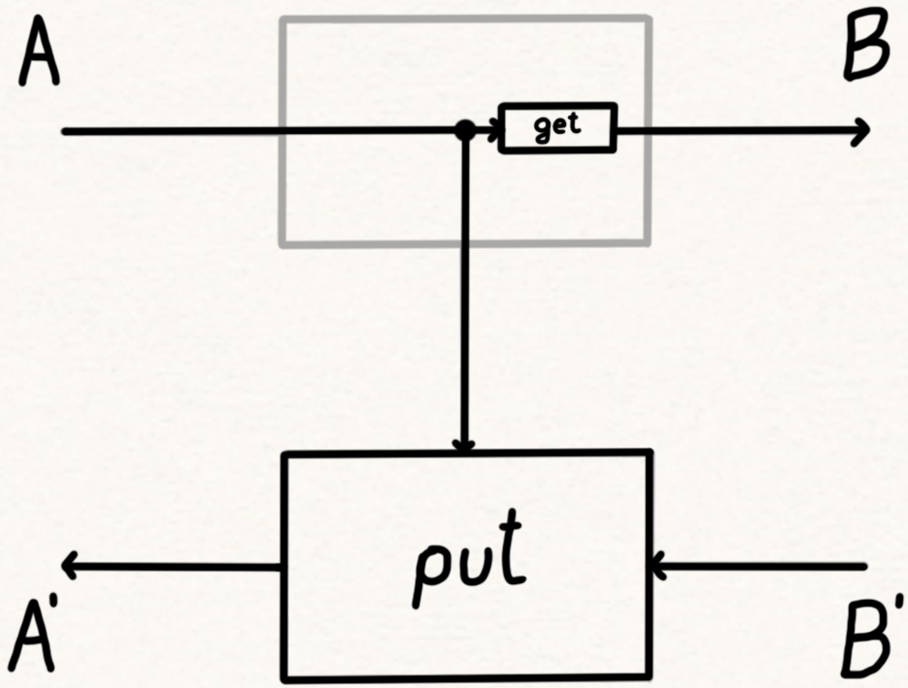




EXAMPLE

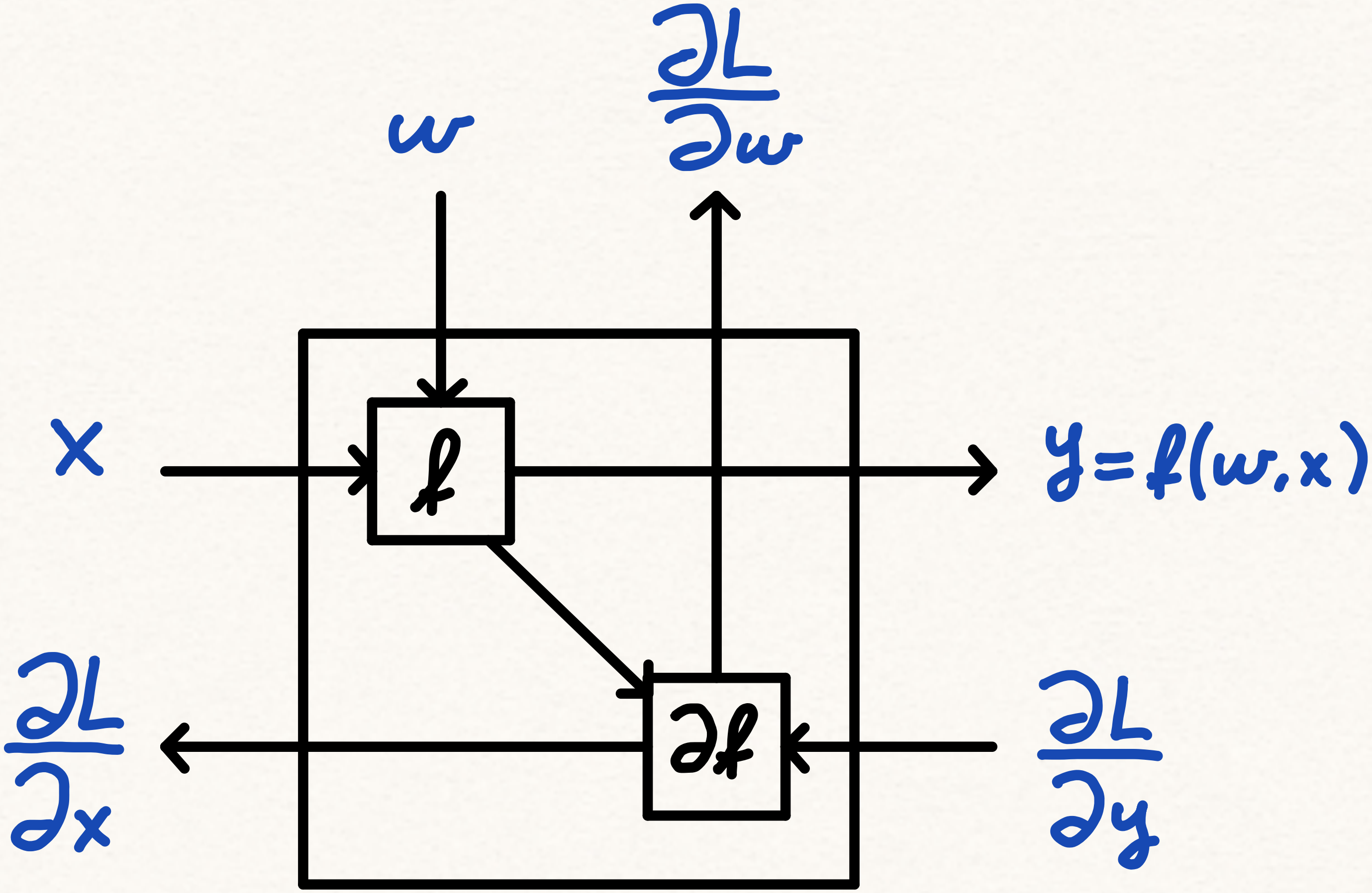

w_1
w_2
B
B
w_1
w_2
$\mathbb{R}^{4\times2+2}$
$\mathbb{R}^{2\times3+3}$
$\mathbb{R}^4$
f
$\mathbb{R}^2$
g
$\mathbb{R}^3$

LENS



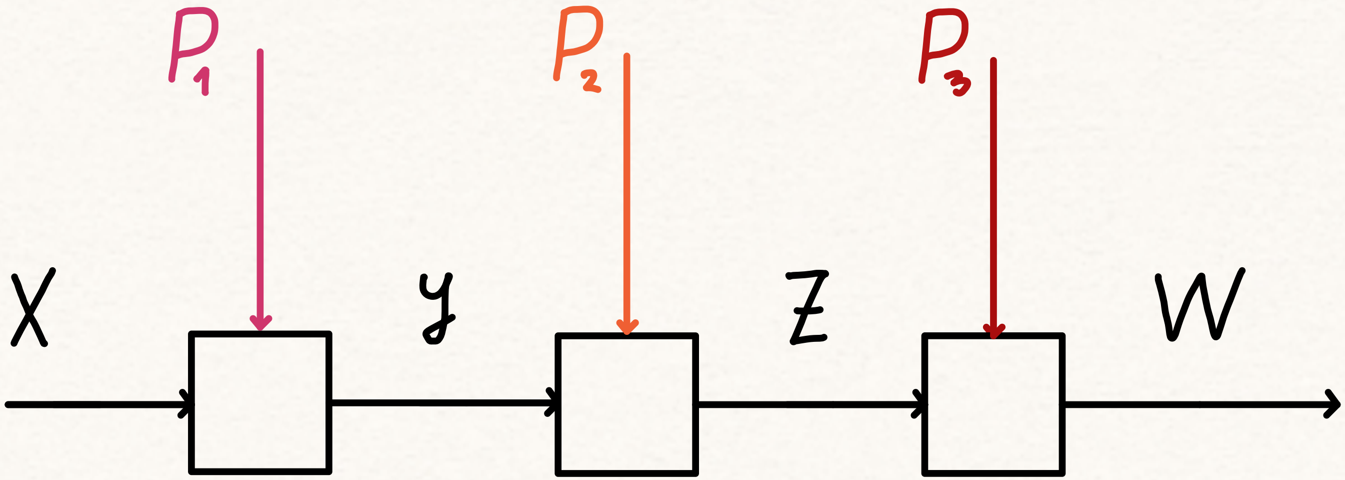
$$\mathcal{C} \xrightarrow{R} \mathrm{Lens}_A(\mathcal{C})$$

PARAMETRIC LENS



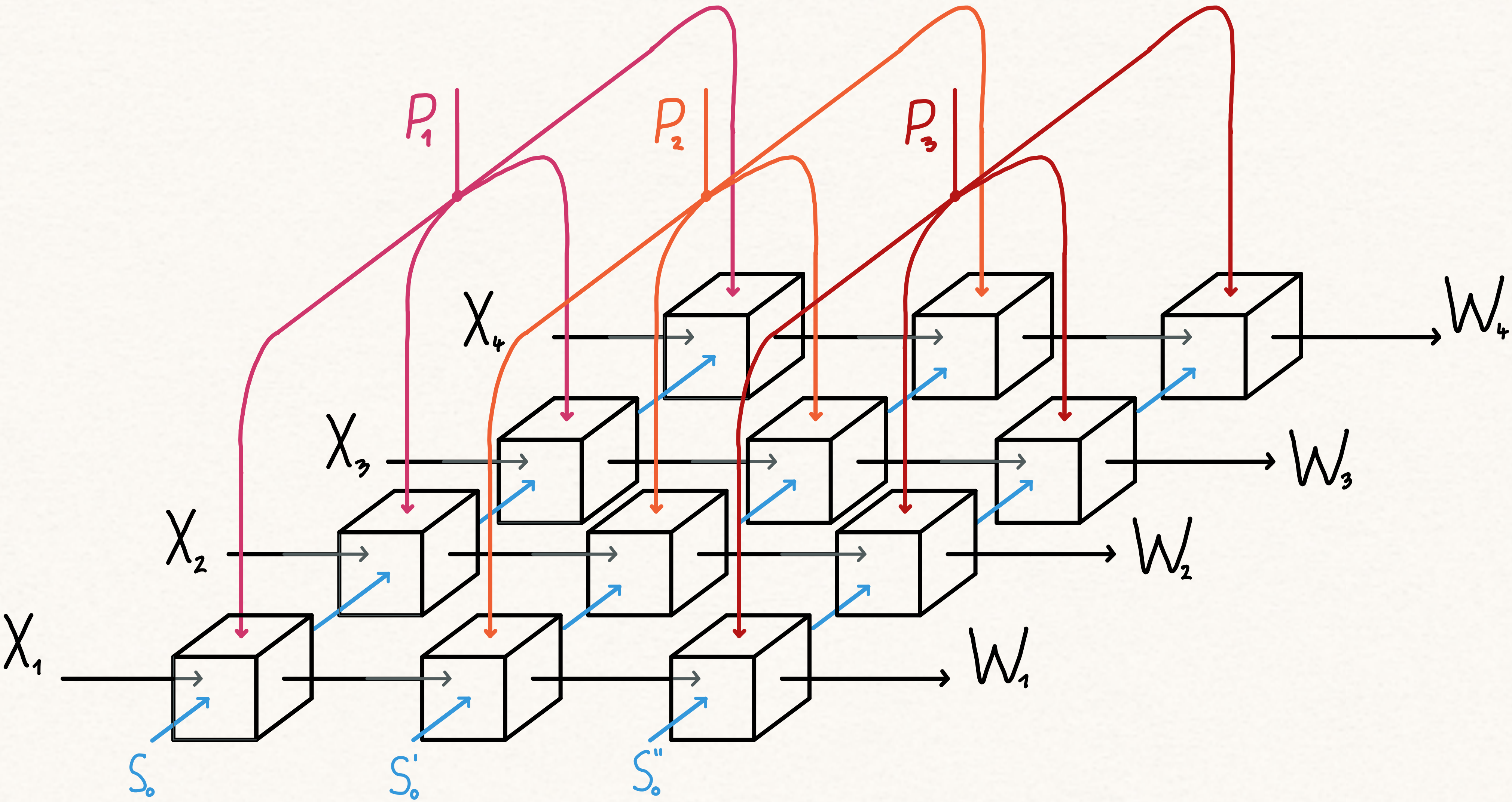
$$\mathrm{Para}(\mathcal{C}) \xrightarrow{\mathrm{Para}(R)} \mathrm{Para}(\mathrm{Lens}(\mathcal{C}))$$

- FRAMEWORK OF PARAMETRIC LENSES IS INCREDIBLY FLEXIBLE

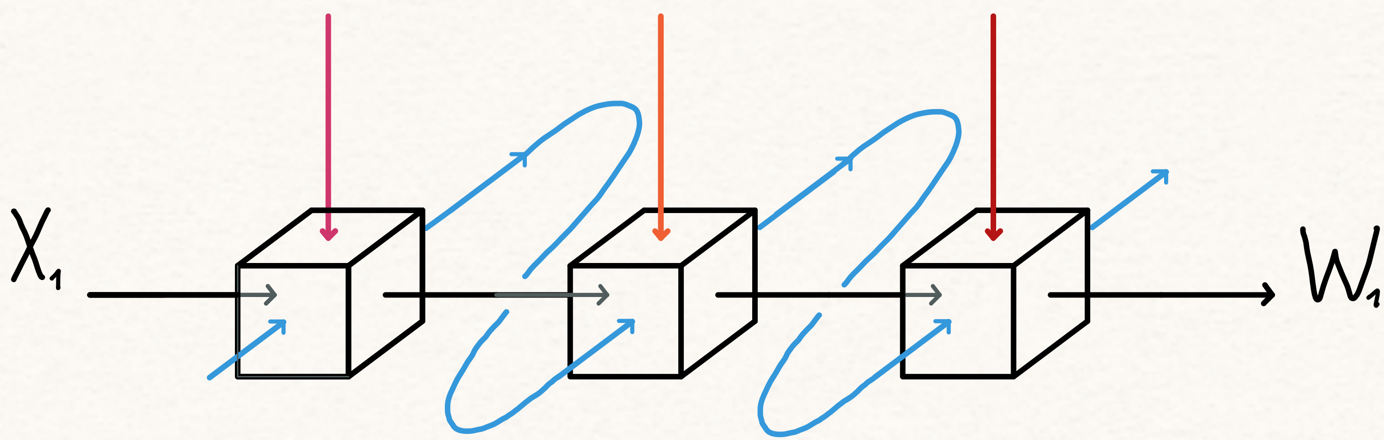


- IT DOESN'T MAKE ANY ASSUMPTIONS ABOUT THE ARCHITECTURE, THUS MODELLING...

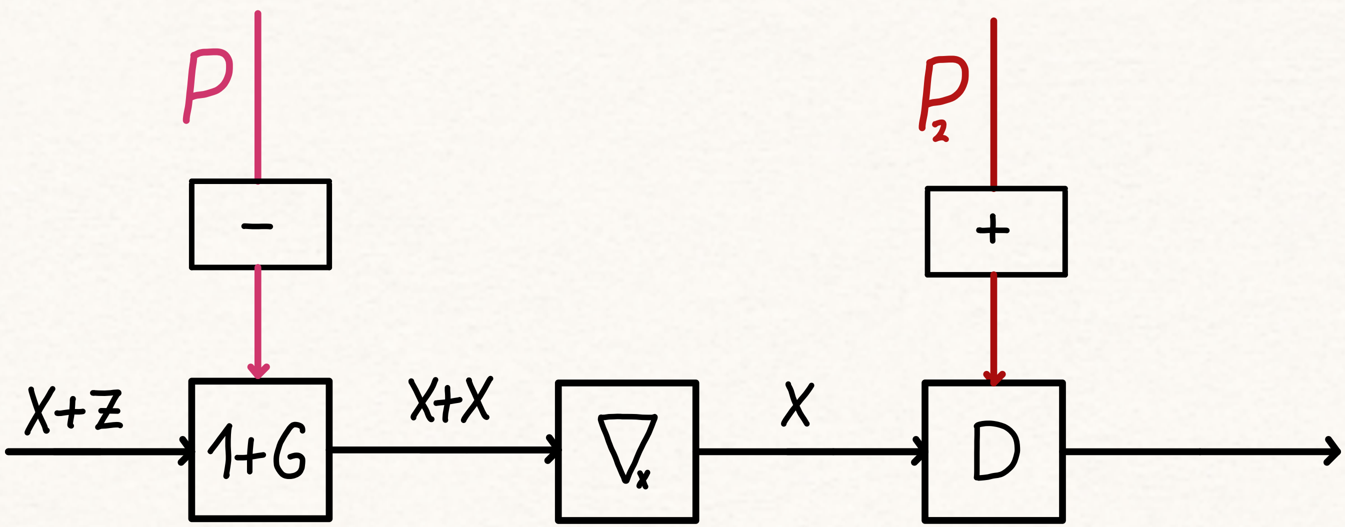
RECURRENT NEURAL NETWORKS



AUTOREGRESSIVE NEURAL NETWORKS



GENERATIVE ADVERSARIAL NETWORKS



- CONVOLUTIONAL NEURAL NETWORKS
- RECURSIVE NEURAL NETWORKS
- GRAPH NEURAL NETWORKS

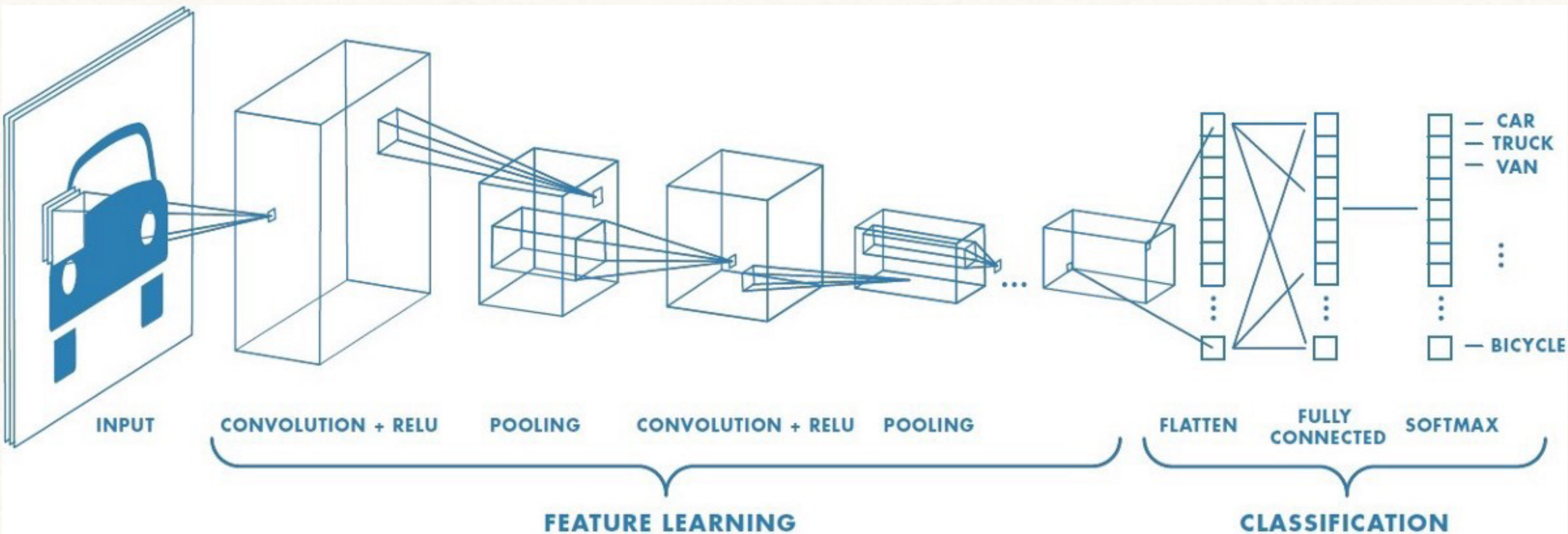
...

ALL OF THESE ARE JUST PARAMETRIC LENSES!

HOW CAN WE MAKE THIS STRUCTURE
VISIBLE IN CATEGORY THEORY?

CONVOLUTIONAL NEURAL NETWORKS

· COMMONLY APPLIED TO PROCESSING IMAGE DATA



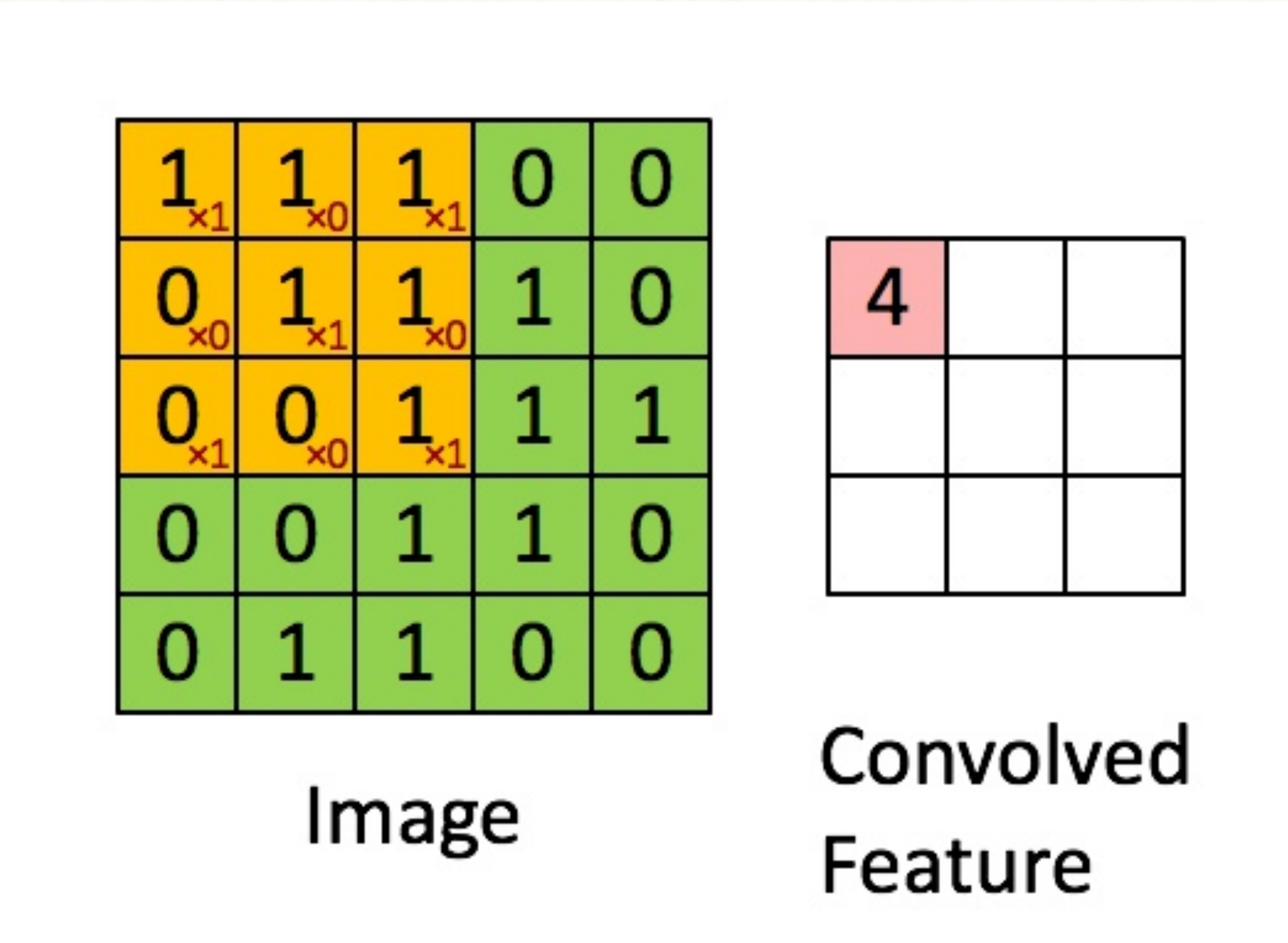
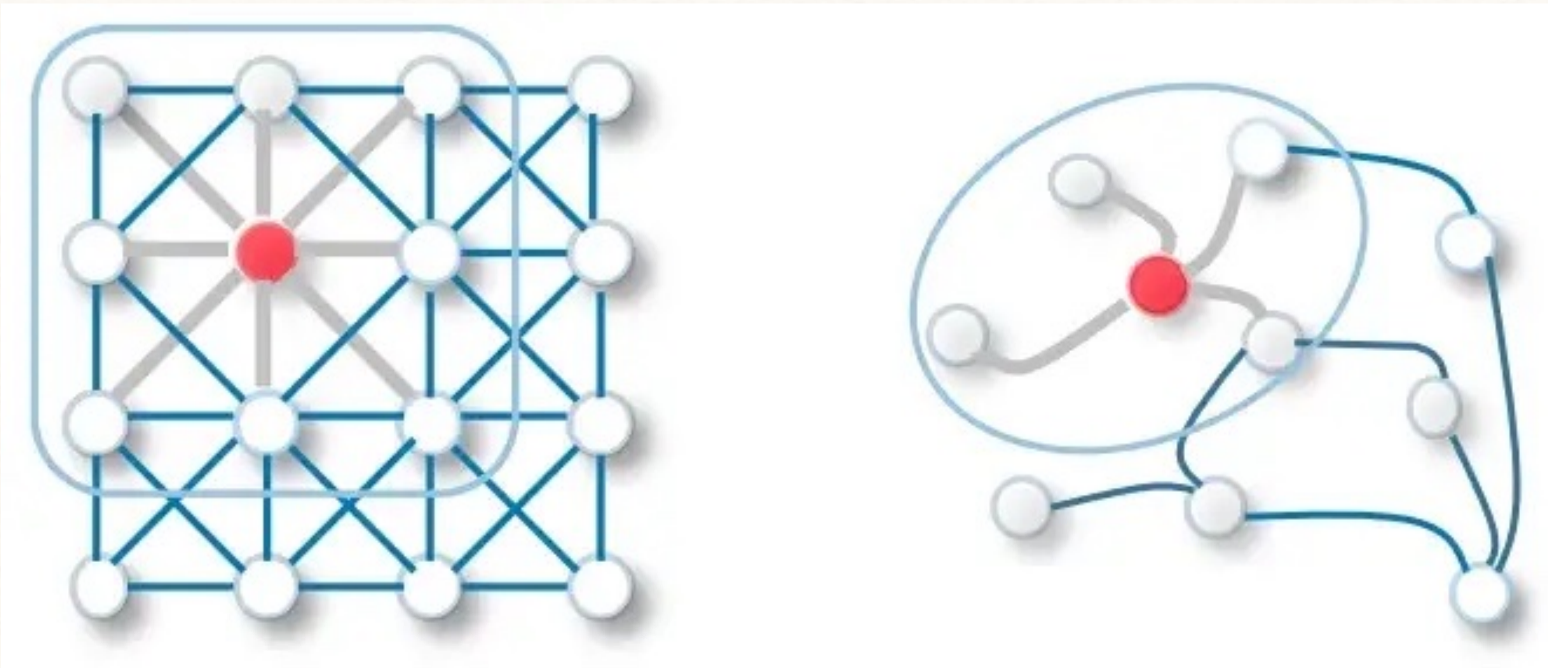


Image
Convolved Feature

GRAPH CONVOLUTIONAL NEURAL NETWORKS

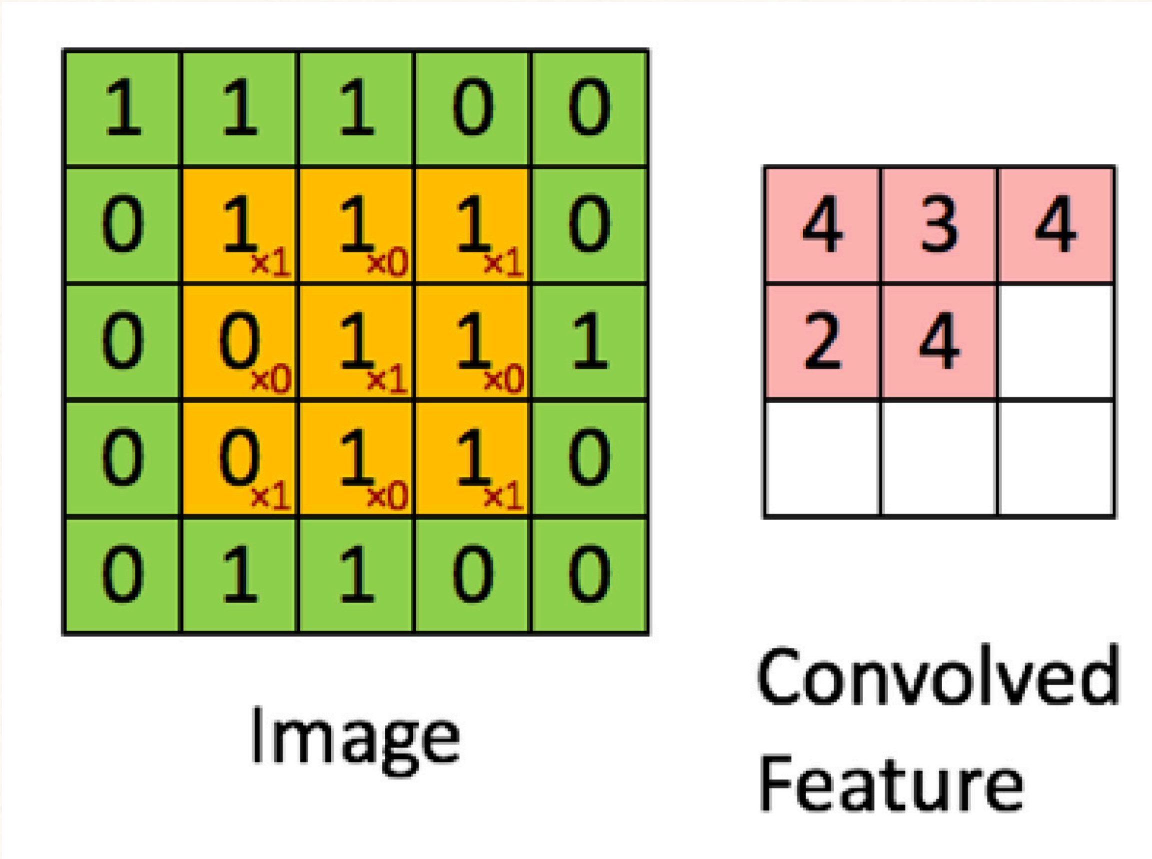
IN ONE SLIDE:

- THEY GENERALISE CONVOLUTIONAL NN's ...



... TO WORK FOR AN ARBITRARY GRAPH.

• IDEA: THINK ABOUT EACH NODE AS RECEIVING MESSAGES FROM ITS NEIGHBOURS!



'PROCESS MANY KINDS OF DATAPOINTS AT ONCE'

BASIC GCNN LAYER

$$f: \mathbb{R}^{n \times k} \longrightarrow \mathbb{R}^{n \times k'}$$

$$f(X) = \sigma(AXW)$$

$$\begin{array}{cc} & n \\ n & \begin{bmatrix} \quad A \quad \end{bmatrix} \end{array} \begin{array}{cc} & k \\ n & \begin{bmatrix} \quad X \quad \end{bmatrix} \end{array} \begin{array}{cc} & k' \\ k & \begin{bmatrix} \quad W \quad \end{bmatrix} \end{array}$$

n – NUMBER OF NODES IN A GRAPH

k – NUMBER OF INCOMING FEATURES

k' – NUMBER OF OUTGOING FEATURES

EACH LAYER HAS ITS OWN WEIGHT MATRIX

EACH LAYER SHARES THE ADJACENCY MATRIX

WE NEED SOMETHING LIKE Para?

THE COREADER COMONAD

LET $\mathcal{C}$ BE A CARTESIAN CATEGORY.

FIX $A : \mathcal{C}$.

$$\mathcal{C} \xrightarrow{A \times -} \mathcal{C}$$

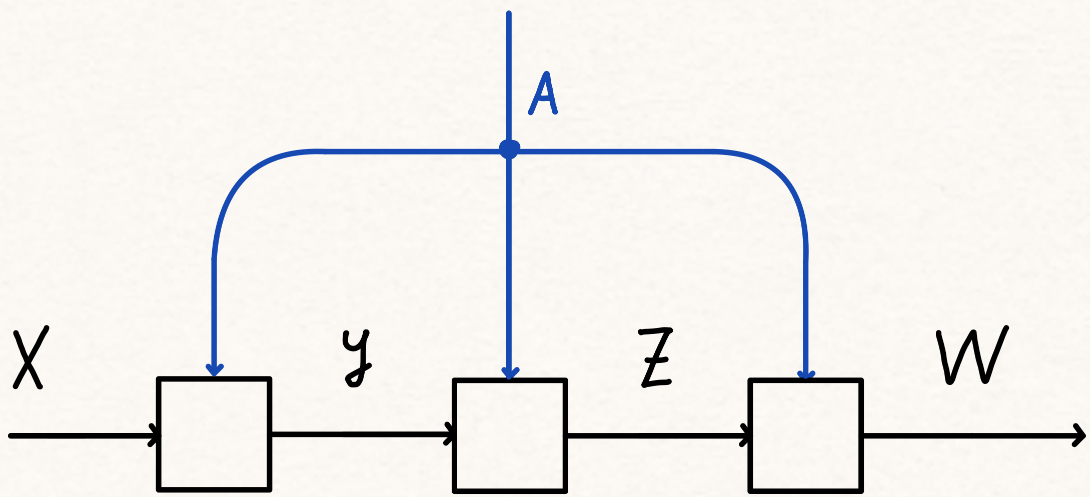
$$\delta_X : A \times X \longrightarrow A \times A \times X$$

$$\varepsilon_X : A \times X \longrightarrow X$$

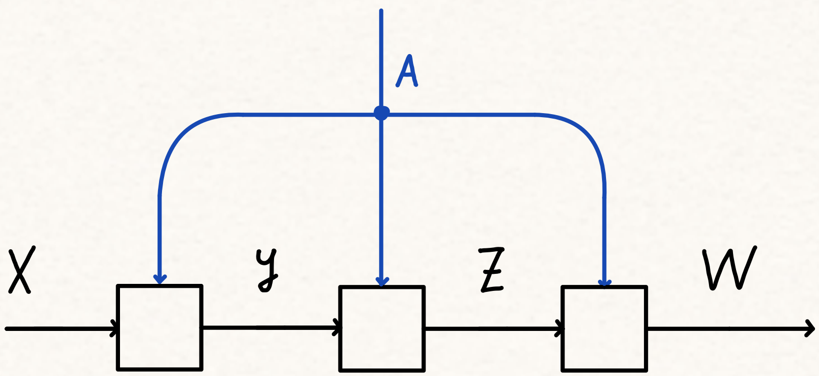
$\mathrm{CoKl}(A\times -)$

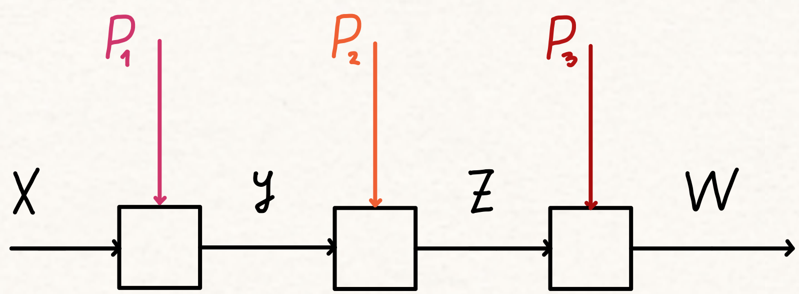
- CATEGORY WITH THE SAME OBJECTS AS $\mathcal{C}$
- $\mathrm{CoKl}(A\times -)(X, Y) := \mathcal{C}(A\times X, Y)$

COMPOSITION SHARES A



$\mathrm{CoKl}(A \times -)$	$\mathrm{Para}(\mathcal{C})$
GLOBAL	LOCAL
CATEGORY	BICATEGORY
COMONAD	GRADED COMONAD





„Para IS THE LOCAL VERSION OF $\mathrm{CoKl}(- \times =)$"

CAN WE USE $\mathrm{CoK\ell}(A\times -)$ AS THE BASE CATEGORY FOR LEARNING?

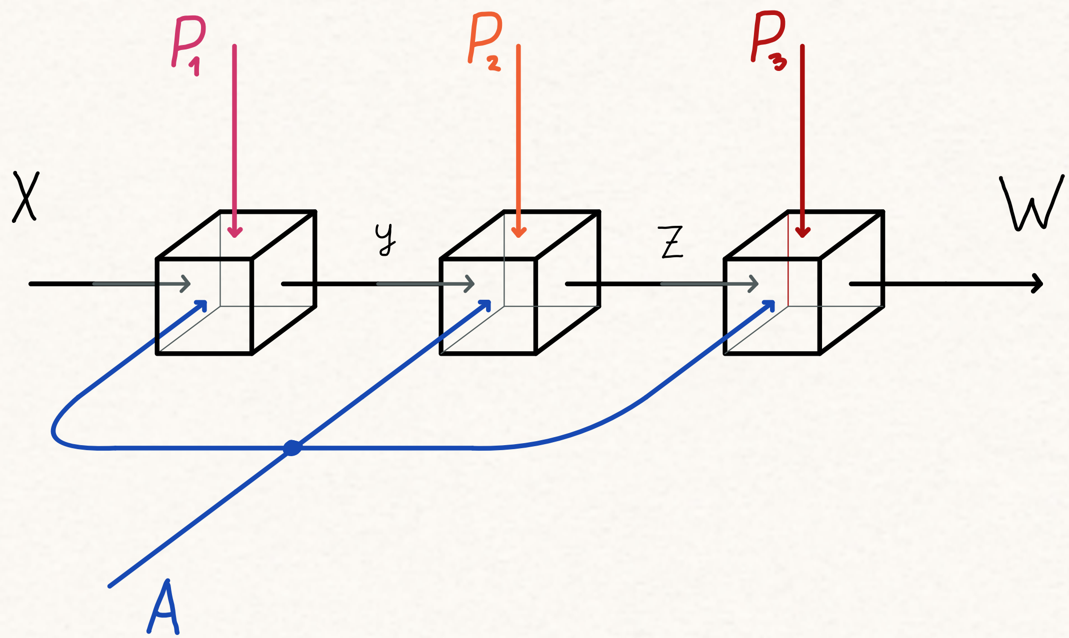
- $\mathrm{Para}(\mathrm{CoK\ell}(\mathbb{R}^{n\times n}))$?

- $\mathrm{CoK\ell}(\mathbb{R}^{n\times n}\times -) \longrightarrow \mathrm{Lens}(\mathrm{CoK\ell}(\mathbb{R}^{n\times n}\times -))$?

YES!

- $\mathrm{CoKl}(A\times -)$ IS A $\mathcal{C}$-ACTEGORY

- $\mathrm{CoKl}(A\times -) \longrightarrow \mathrm{Lens}_A(\mathrm{CoKl}(A\times -))$ EXISTS WHEN
  $\mathcal{C} \longrightarrow \mathrm{Lens}_A(\mathcal{C})$ DOES



$$\mathrm{GCNN}_n \xrightarrow{K_n} \mathbf{Para}(\mathbf{CoKl}(\mathbb{R}^{n \times n} \times -))$$

$$
\begin{tikzcd}
\mathbf{GCNN}_n \arrow[d, "\mathcal{K}_n"] \\
\mathbf{Para}(\mathbf{CoKl}(\mathbb{R}^{n \times n} \times -)) \arrow[d, "\mathbf{Para}(\mathbf{R})"] \\
\mathbf{Para}(\mathbf{Lens}(\mathbf{CoKl}(\mathbb{R}^{n \times n} \times -)))
\end{tikzcd}
$$

FUTURE WORK

- GENERAL THEORY OF ARCHITECTURES?
- USE (CO)ALGEBRAS TO DESCRIBE THEIR OFTEN STRUCTURALLY RECURSIVE NATURE?