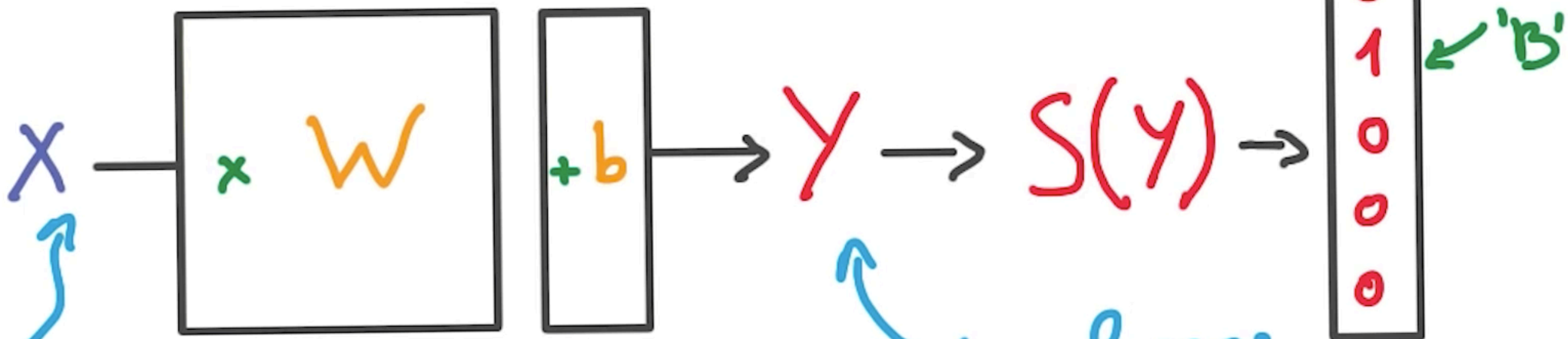
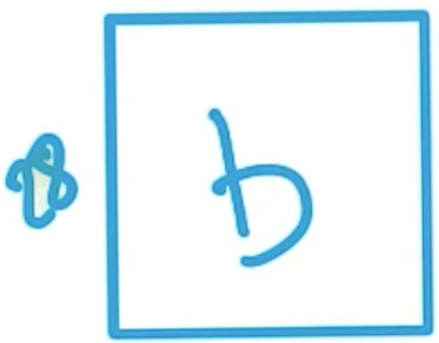
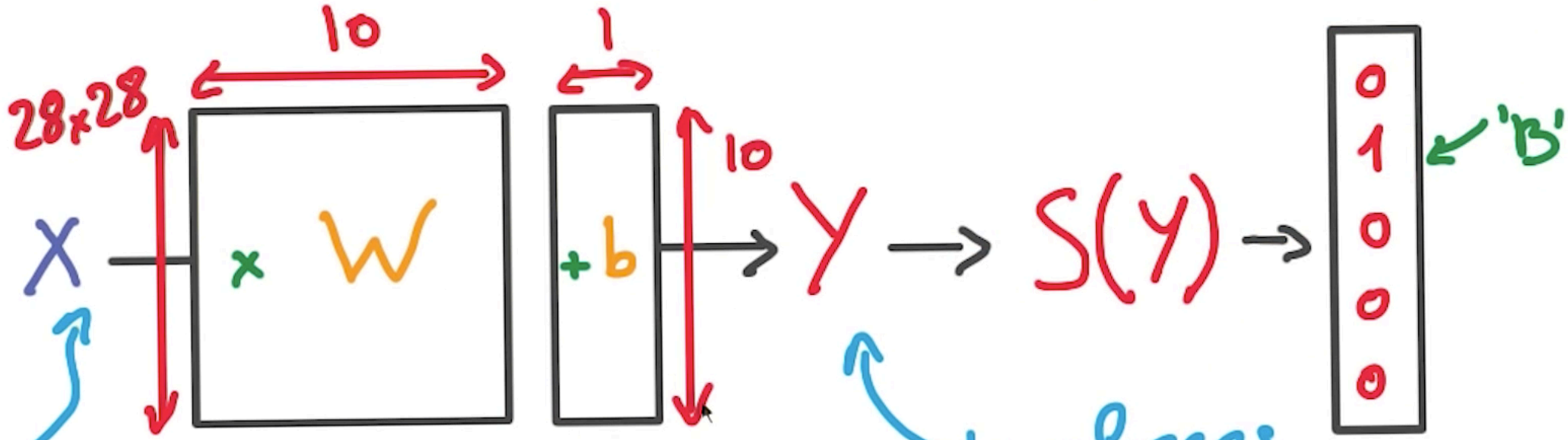
LINEAR MODEL COMPLEXITY



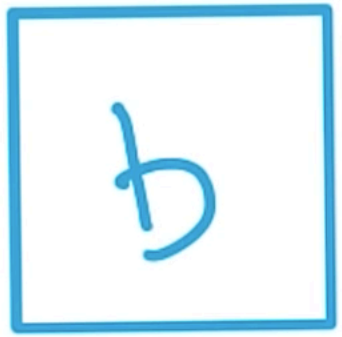


NUMBER OF PARAMETERS ?

LINEAR MODEL COMPLEXITY



28



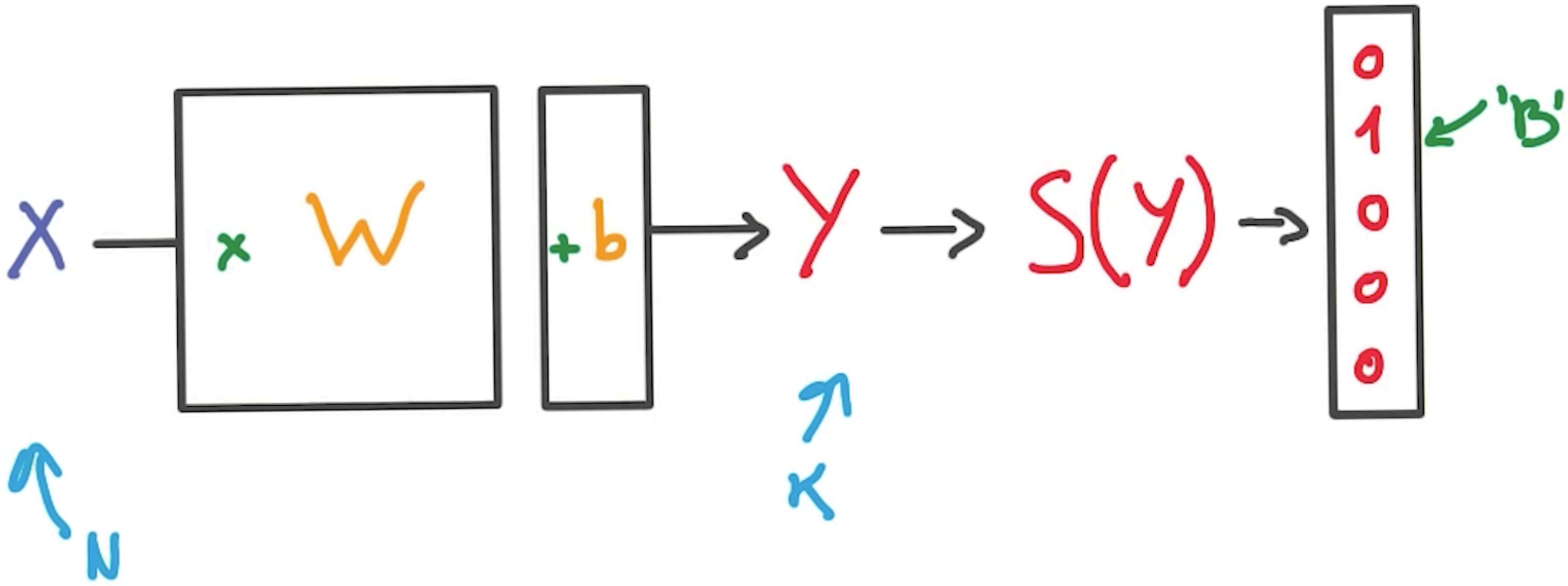
28

10 classes
'A', 'B', ... 'J'

NUMBER OF PARAMETERS?

7850 = 28x28x10+10

LINEAR MODEL COMPLEXITY



(N+1)K PARAMETERS

LINEAR MODELS ARE ... LINEAR!

$$Y = X_1 + X_2 \quad \checkmark$$

$$Y = X_1 \times X_2 \quad \times$$

LINEAR MODELS ARE ... STABLE?

$$Y = WX \rightarrow \Delta Y \sim |W| \Delta X$$

ΔY: SMALL

$|W|$: BOUNDED

ΔX: SMALL

LINEAR MODELS ARE ... STABLE?

$$Y = WX \longrightarrow \frac{\Delta Y}{\Delta X} = W^T$$

CONSTANTS

$$\frac{\Delta Y}{\Delta W} = X^T$$

LINEAR MODELS ARE ... HERE TO STAY

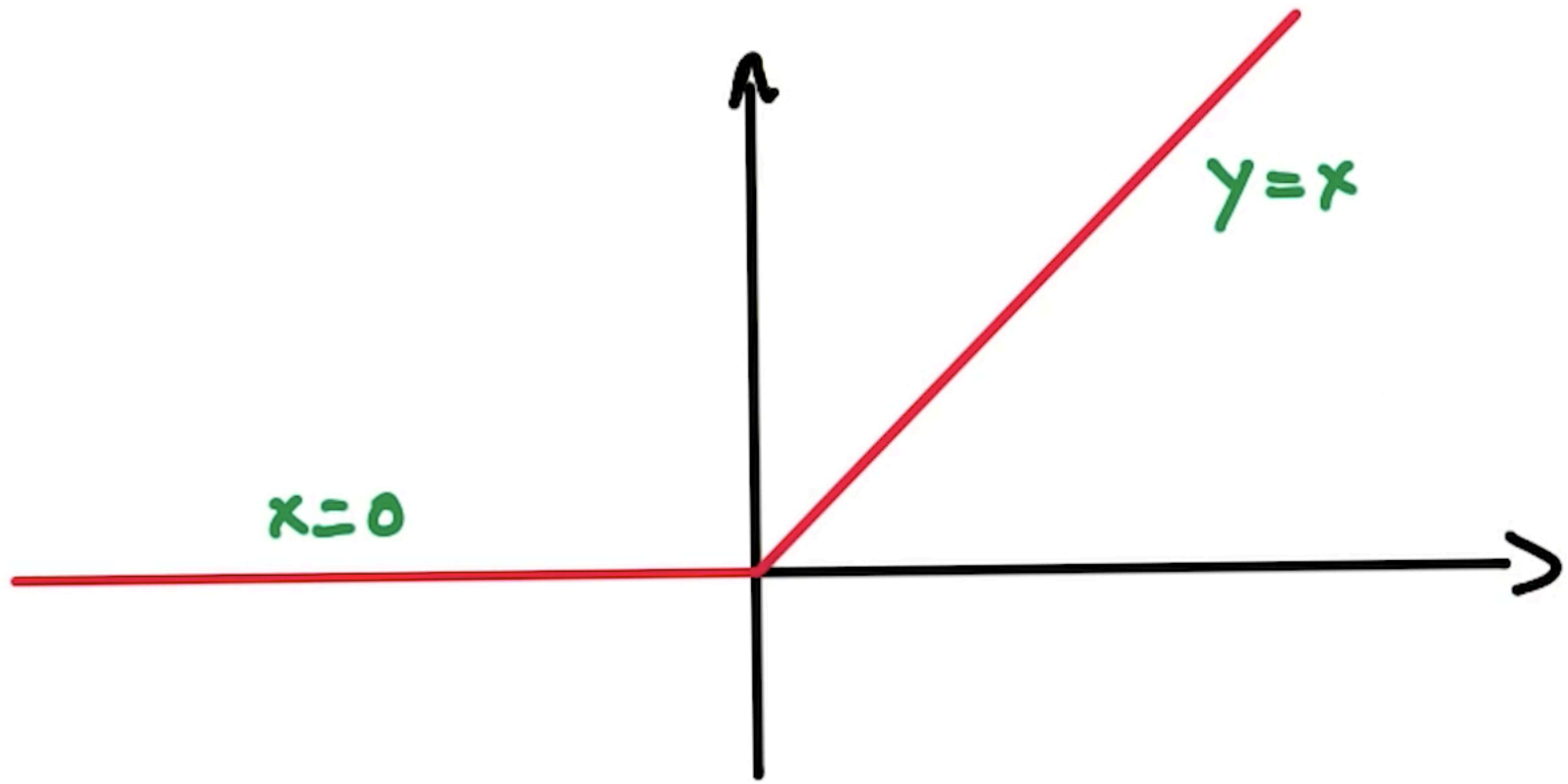
$$Y = W_1 W_2 W_3 X = W X$$

LINEAR MODELS ARE ... HERE TO STAY

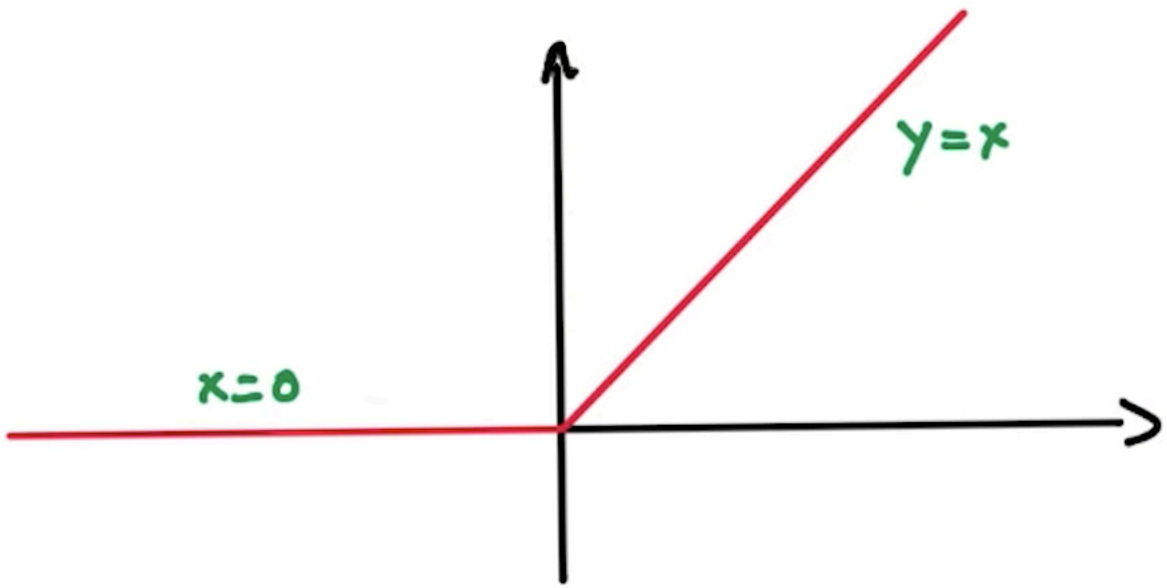
$$Y = W_1 W_2 W_3 X = \mathcal{W} X$$

NON-LINEARITIES

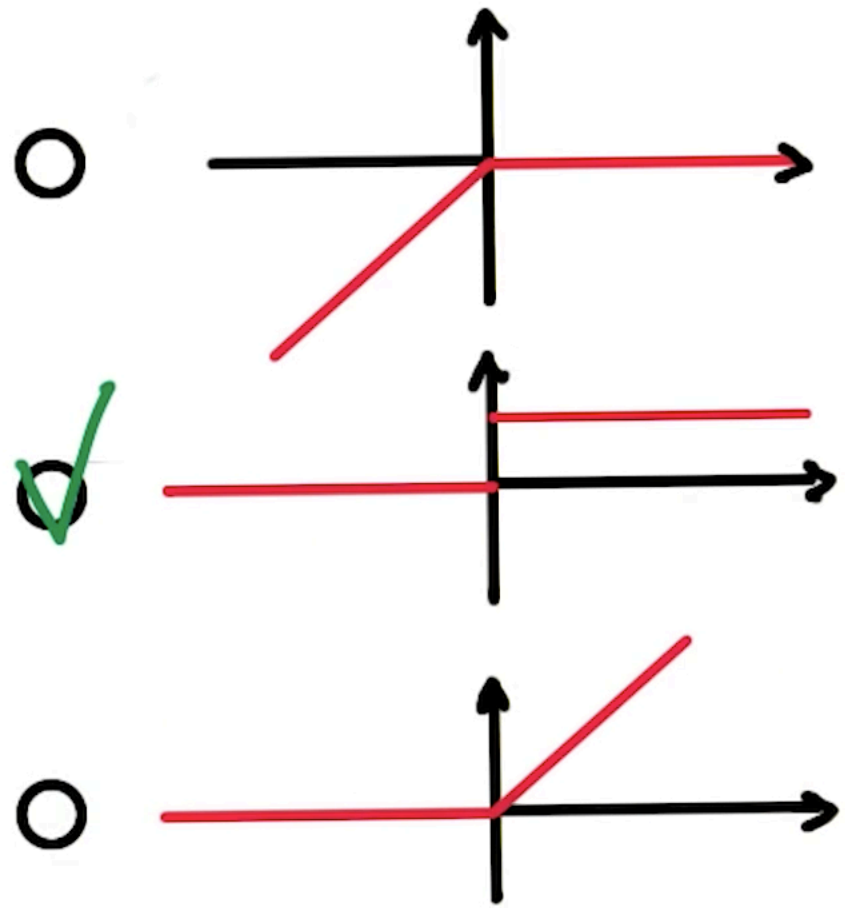
RECTIFIED LINEAR UNITS (RELU)



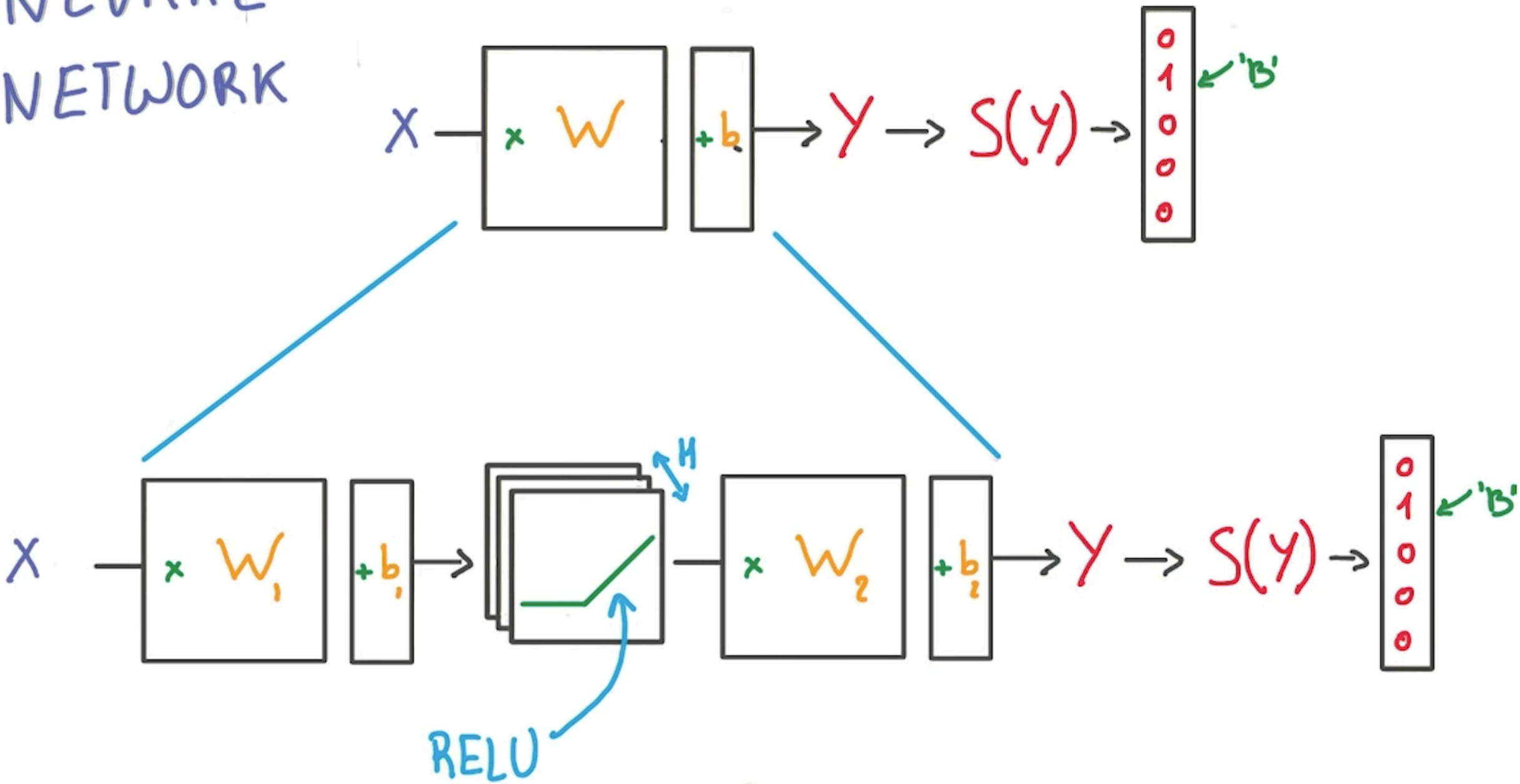
RECTIFIED LINEAR UNITS (RELU)



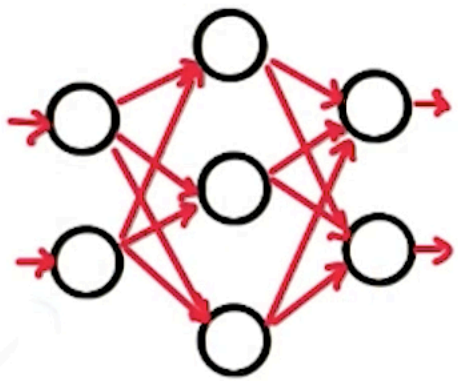
DERIVATIVE?



NEURAL NETWORK

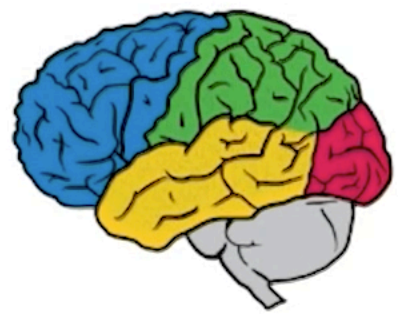


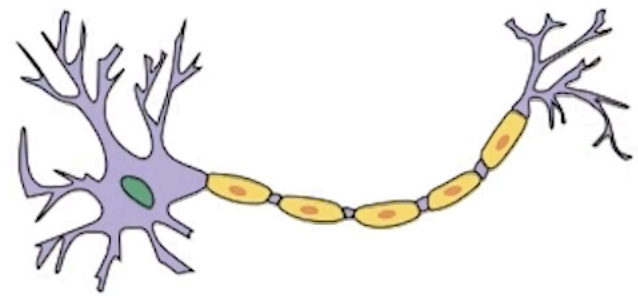
NEURAL NETWORK



NEURONS?

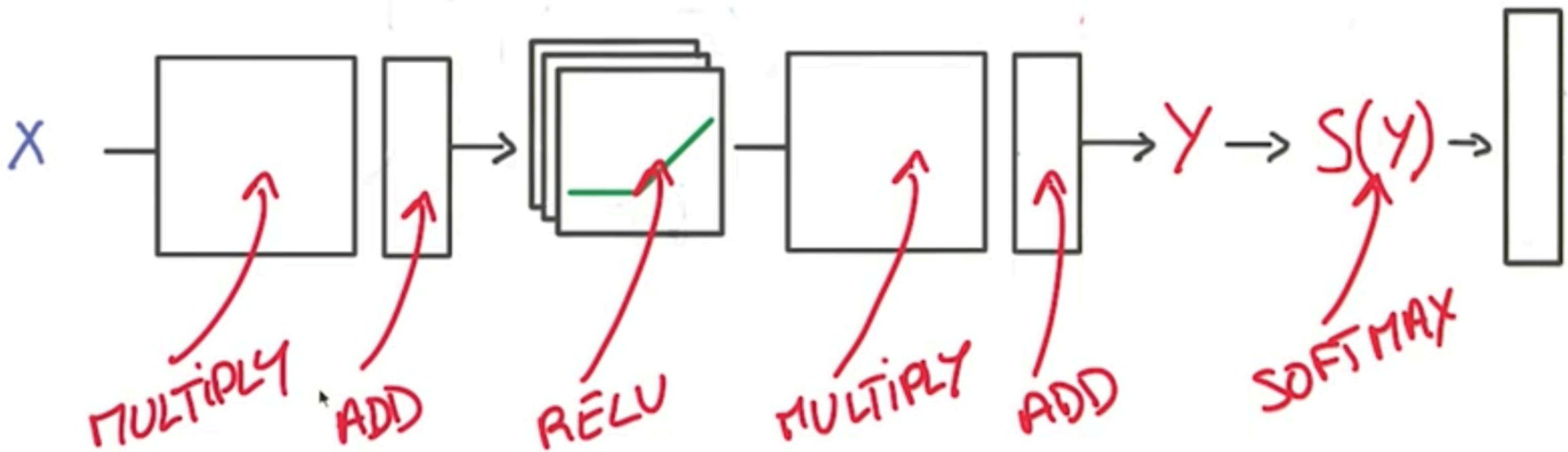
HOW THE BRAIN WORKS?





NEUROMORPHIC ENGINEERING?

STACKING UP SIMPLE OPERATIONS



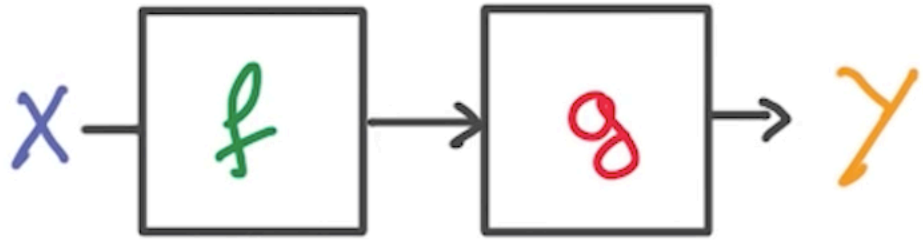
CHAIN RULE

$$[g(f(x))]' = g'(f(x)) \cdot f'(x)$$

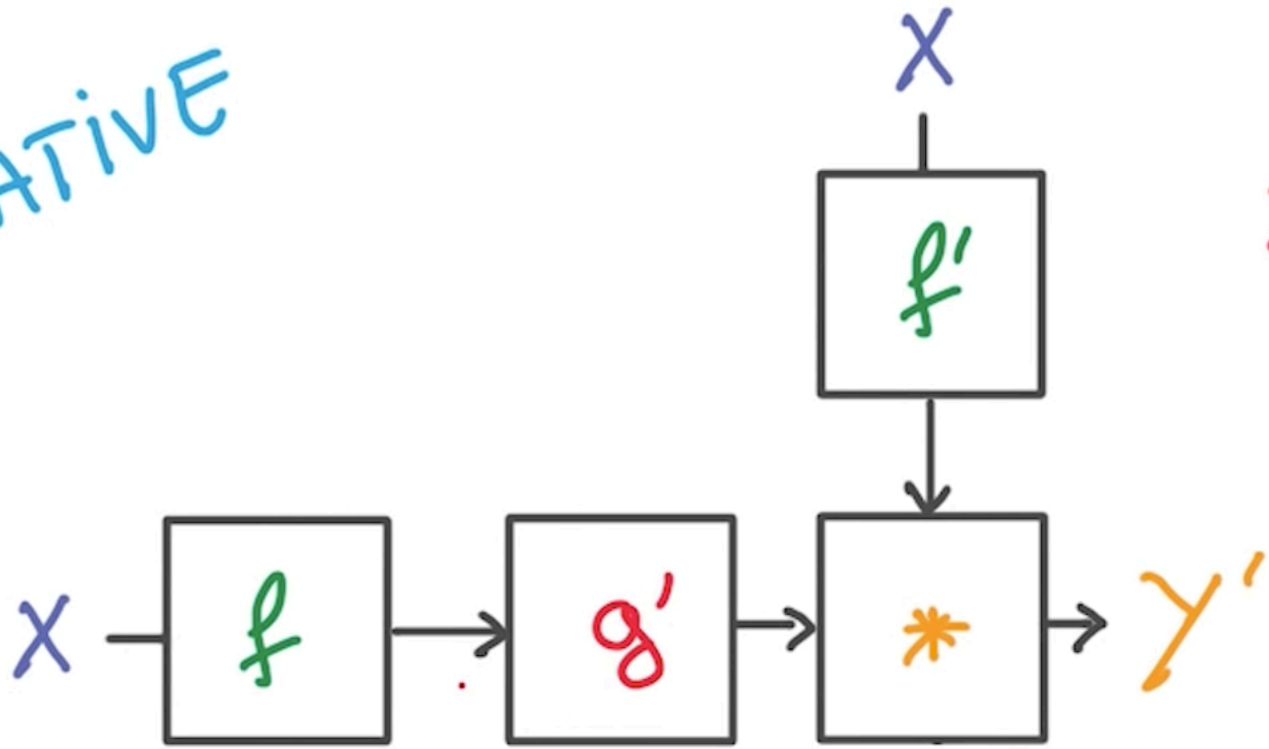
DERIVATIVE

PRODUCT

FUNCTION



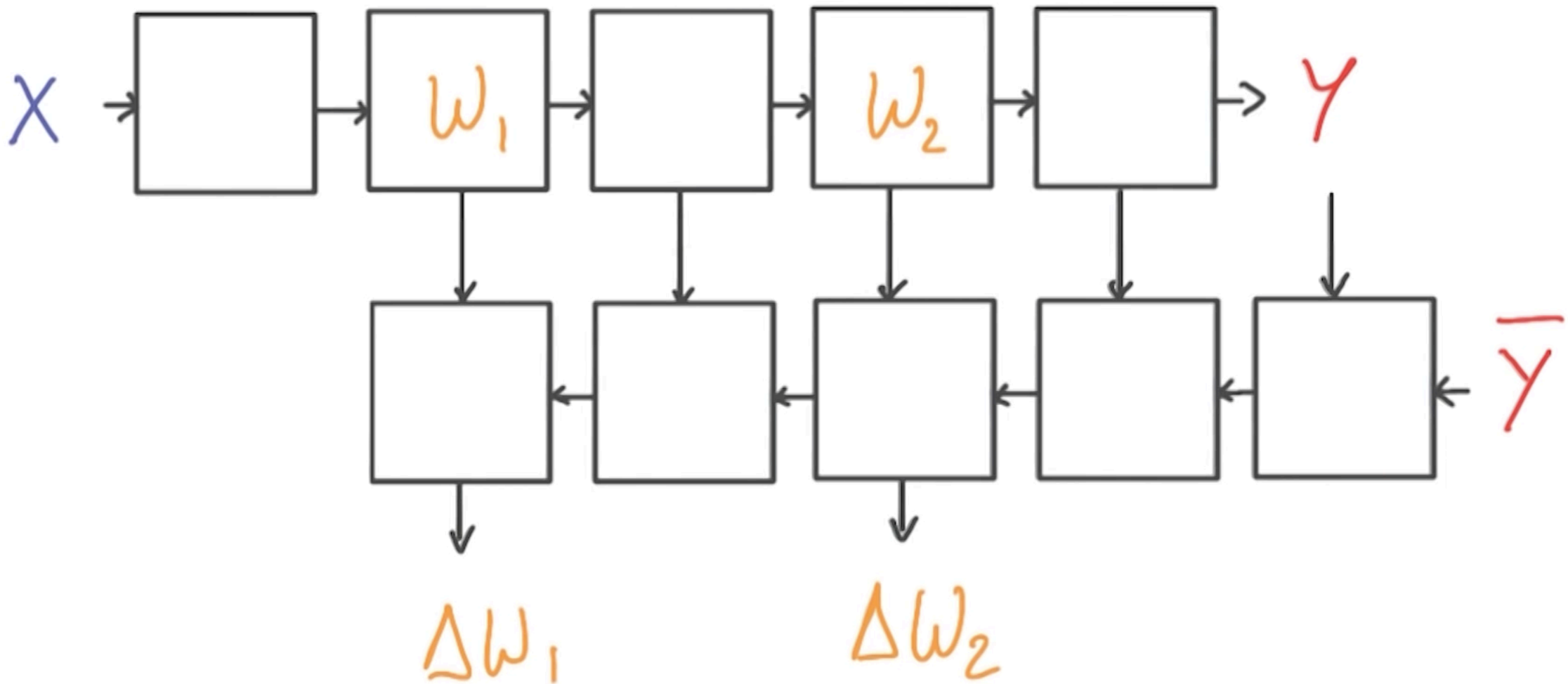
DERIVATIVE



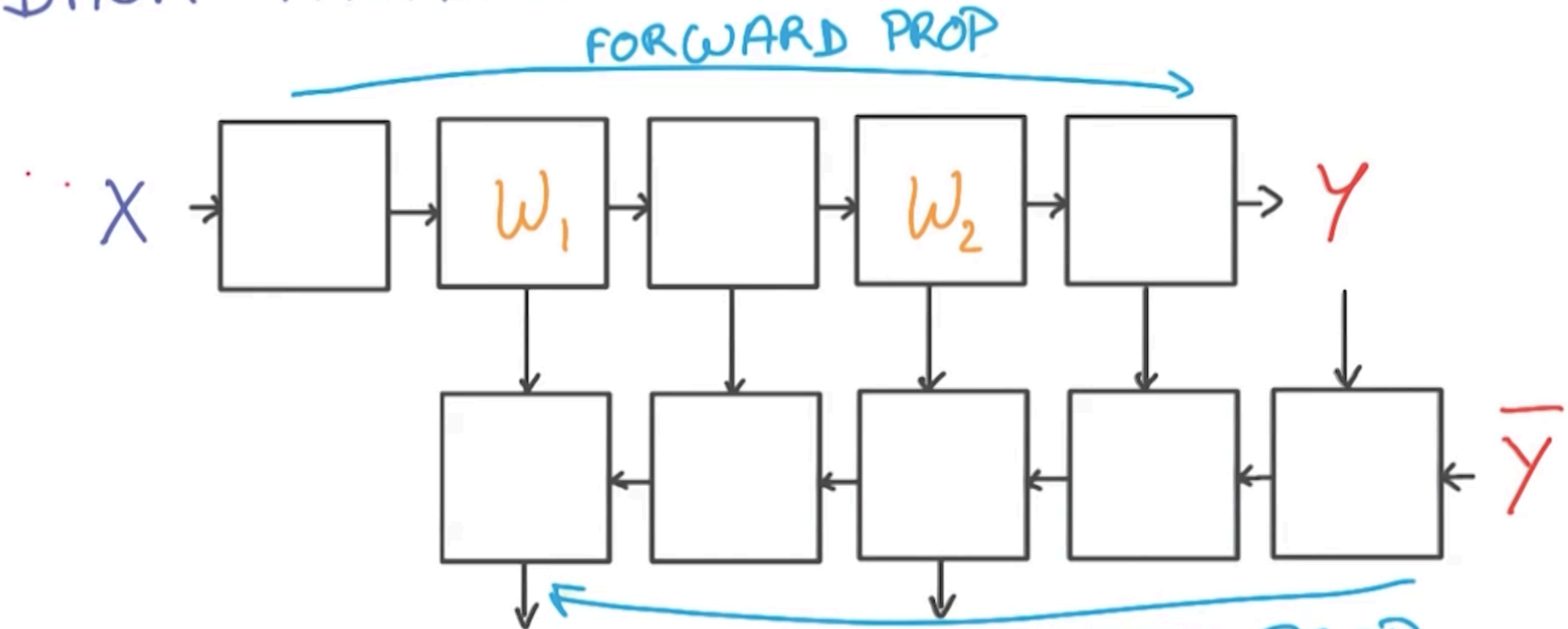
EFFICIENT DATA PIPELINE!

LOTS OF DATA REUSE!

BACK - PROPAGATION



BACK - PROPAGATION



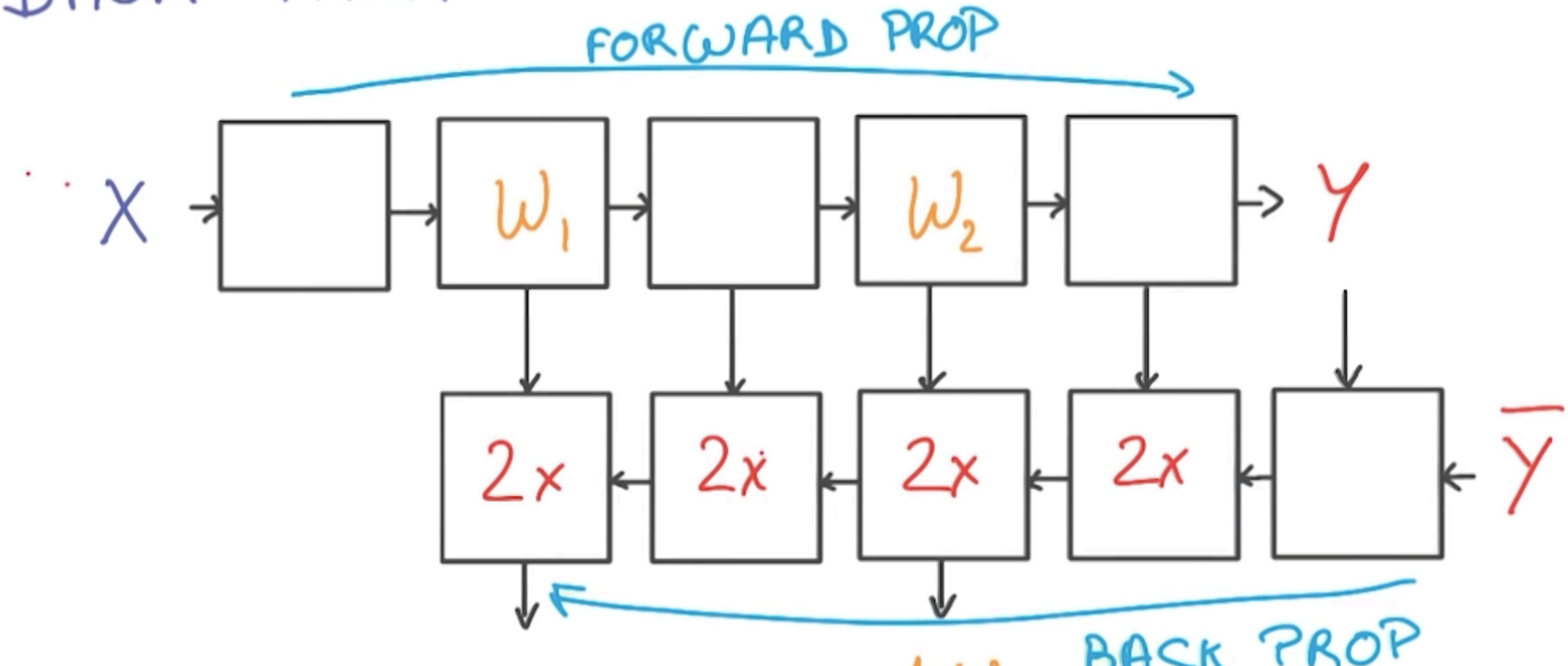
$$\Delta W_1$$

$$W_1 \leftarrow W_1 - \alpha \Delta W_1$$

$$\Delta W_2$$

$$W_2 \leftarrow W_2 - \alpha W_2$$

BACK - PROPAGATION



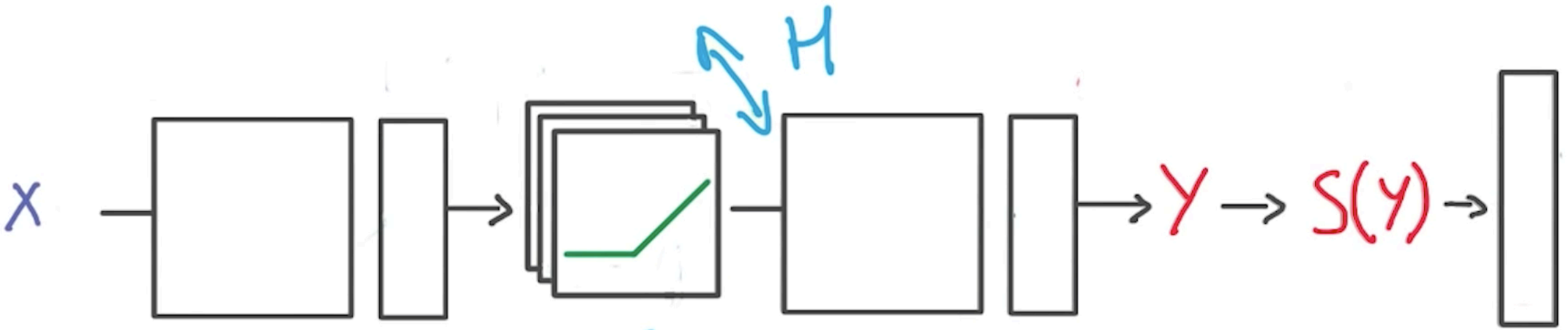
ΔW_1

$$W_1 \leftarrow W_1 - \alpha \Delta W_1$$

ΔW_2

$$W_2 \leftarrow W_2 - \alpha W_2$$

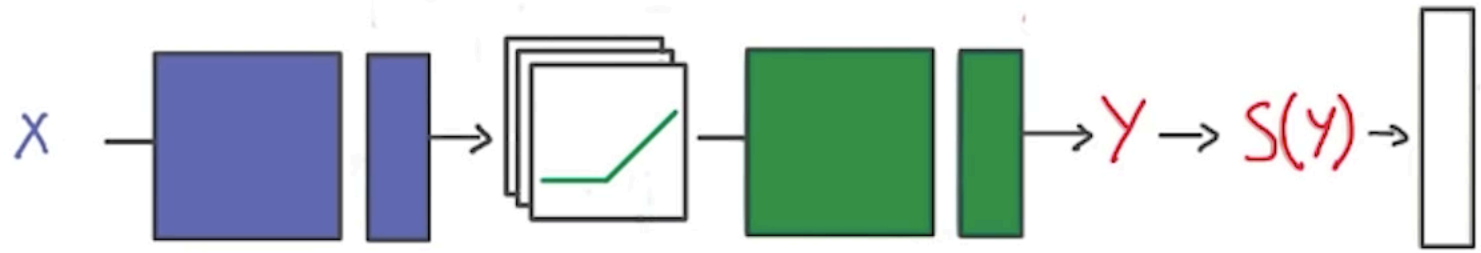
2-LAYER NEURAL NETWORK

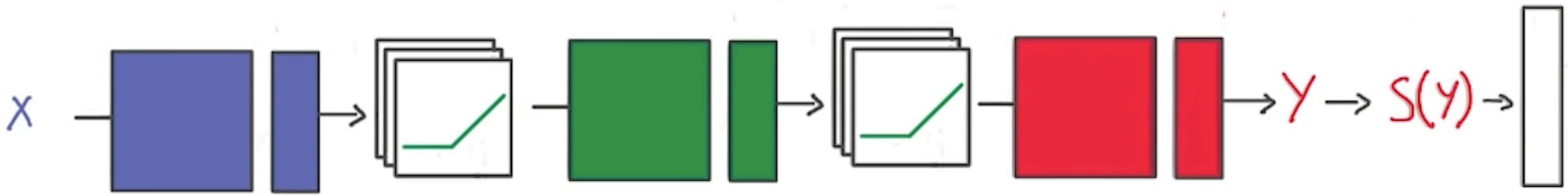


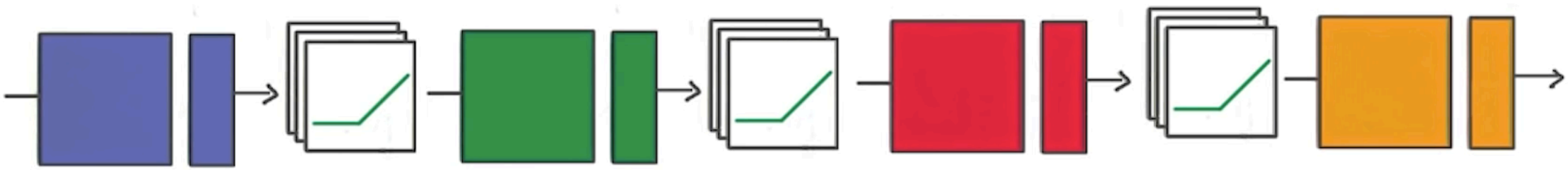
'HIDDEN LAYER'

NOT VERY DEEP YET ...

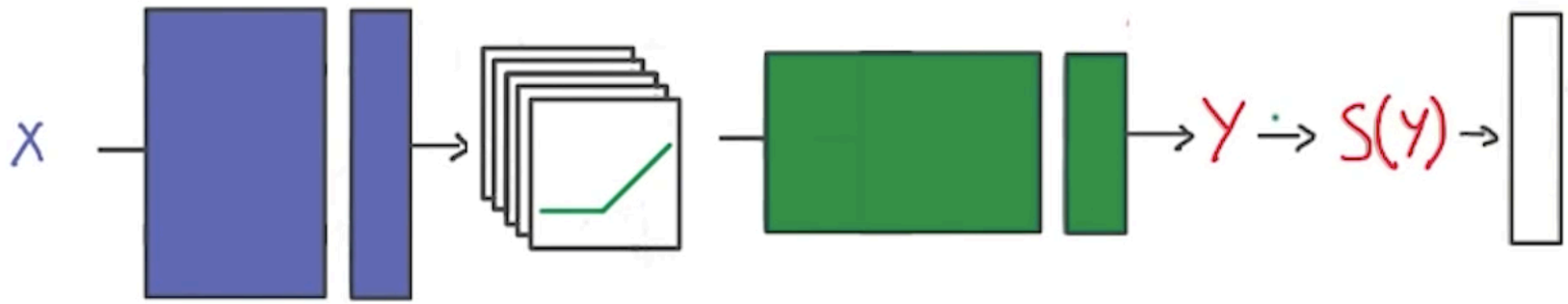
DEEP NETWORKS



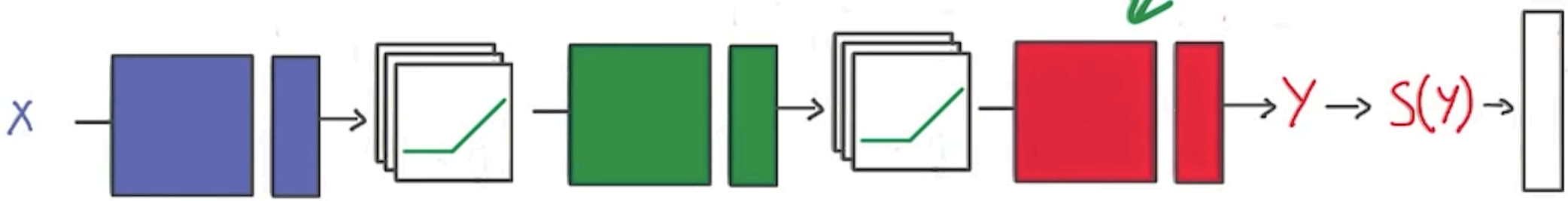




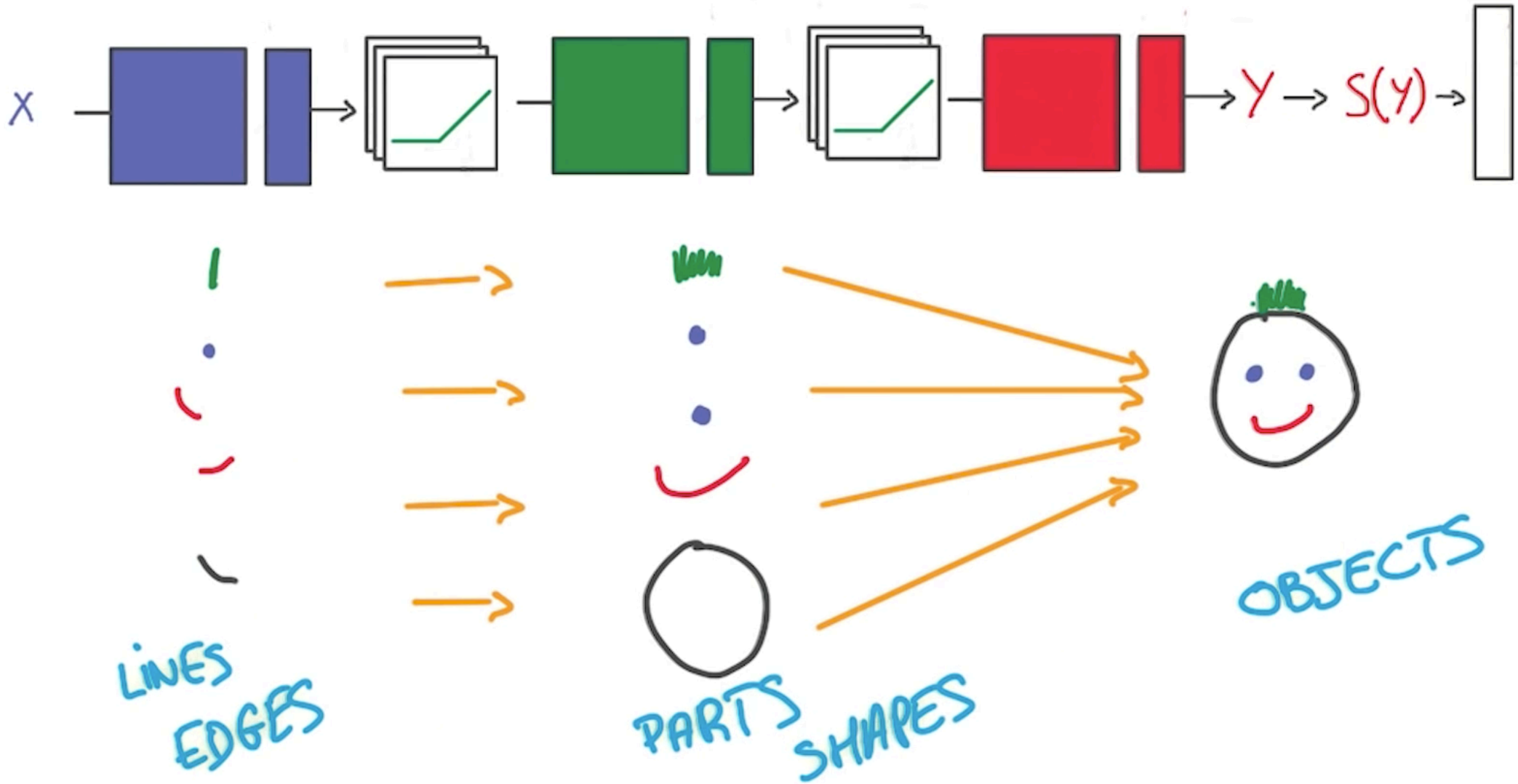
DEEP NETWORKS



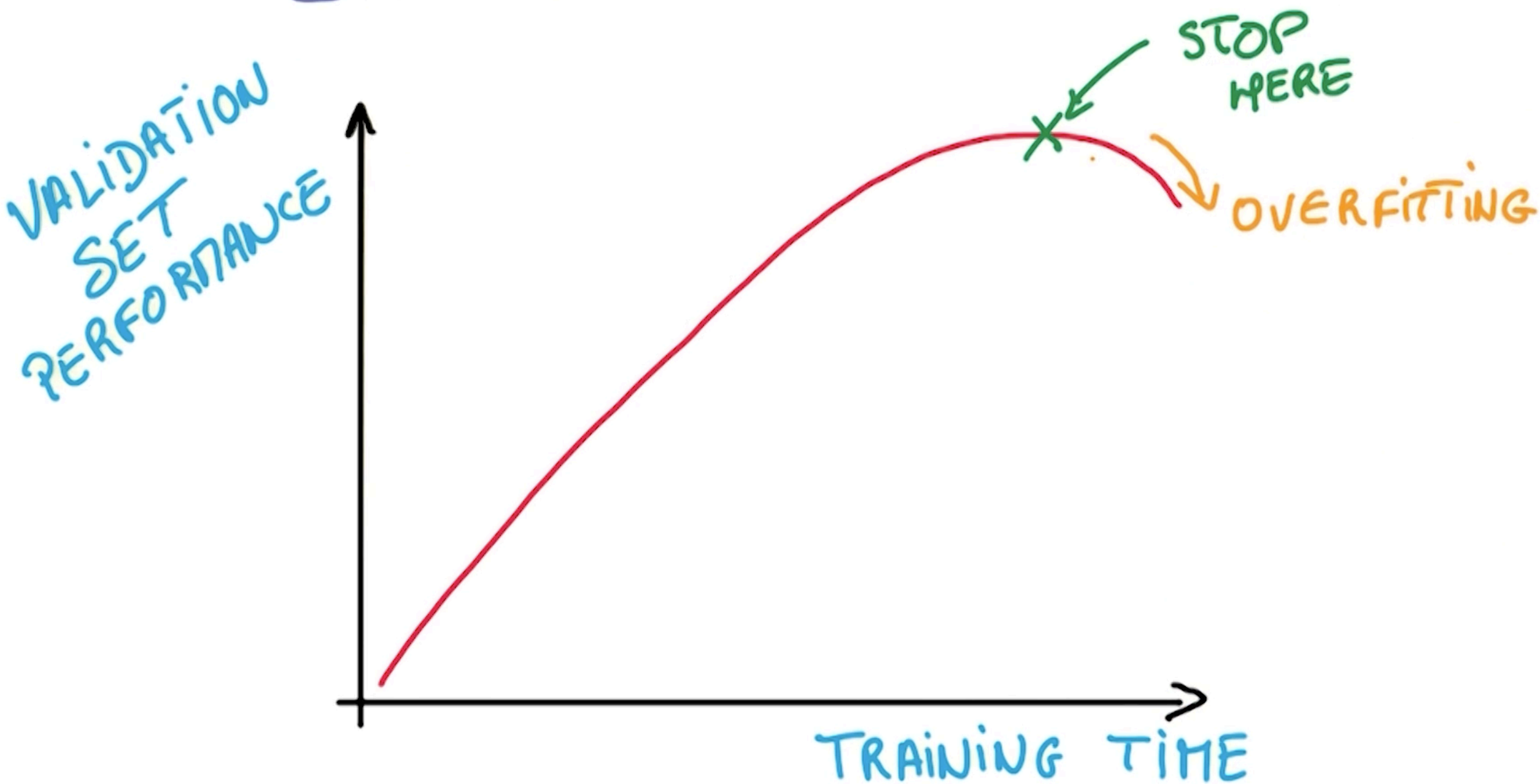
WIDER VS. DEEPER



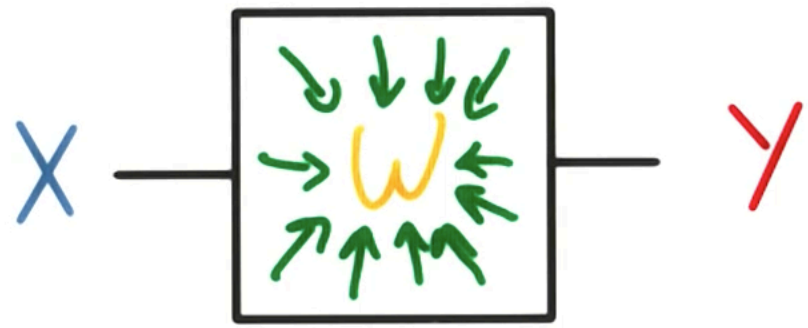
DEEP NETWORKS

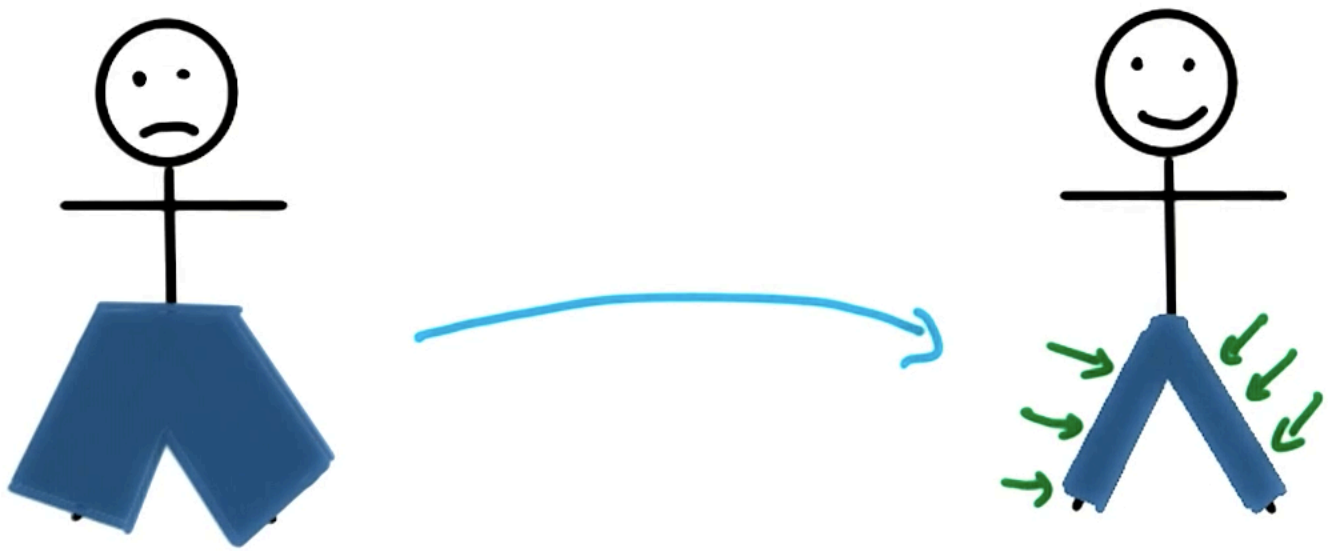


EARLY TERMINATION



REGULARIZATION



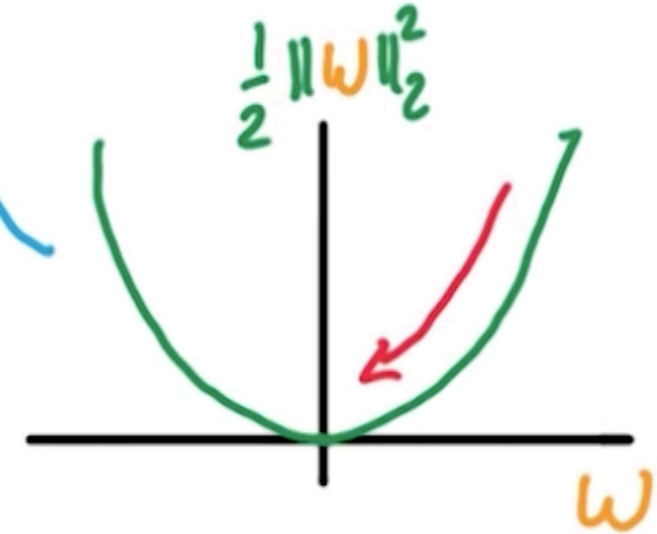


L_2 REGULARIZATION

NEW LOSS

$$\mathcal{L}' = \mathcal{L} + \beta \frac{1}{2} \| w \|_2^2$$

LOSS



L_2 REGULARIZATION

$$\mathcal{L}' = \mathcal{L} + \beta \underbrace{\frac{1}{2}\|w\|_2^2}$$

$\frac{1}{2}(w_1^2 + w_2^2 + \dots + w_N^2)$

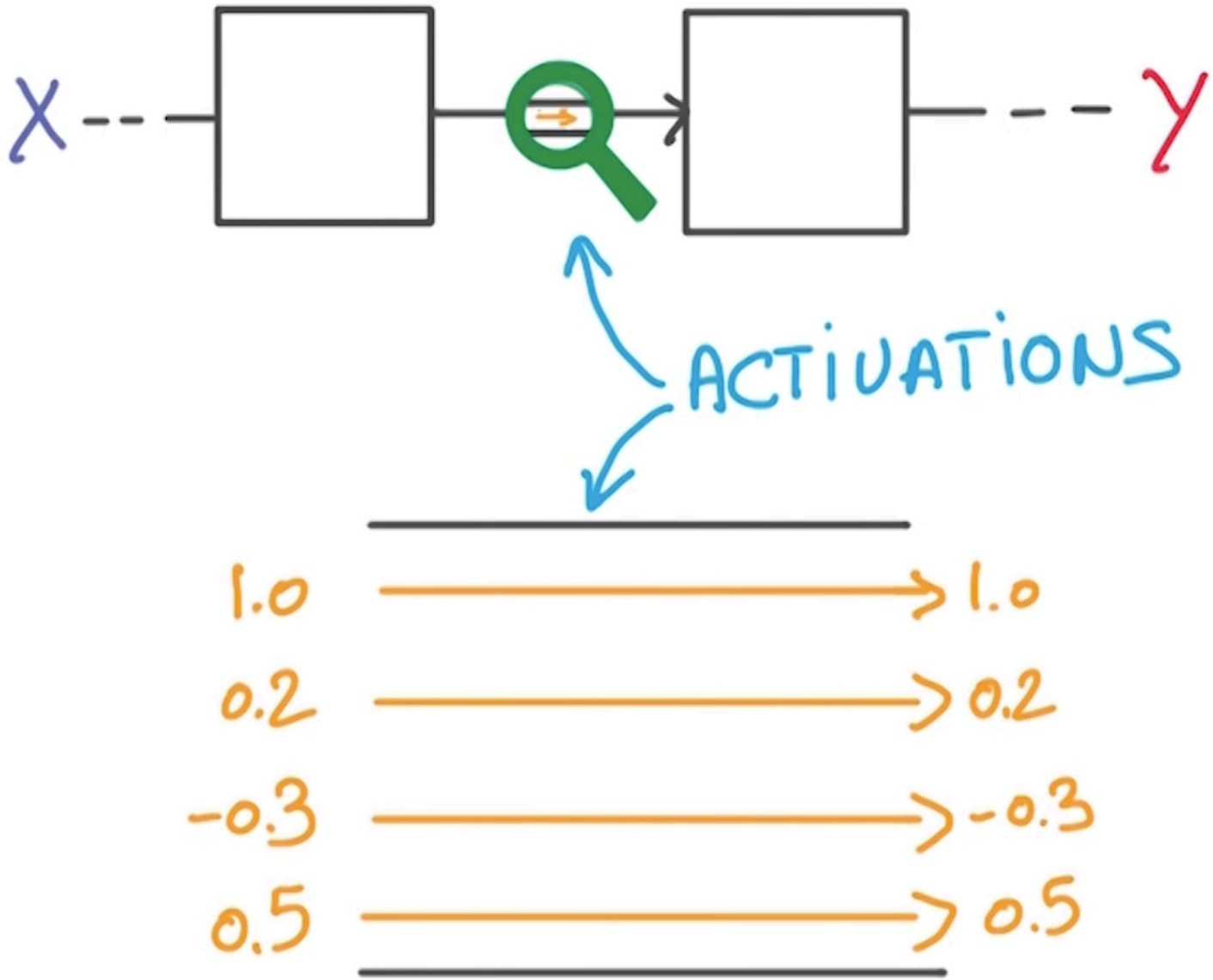
$$\left(\frac{1}{2}x^2\right)' = x$$

$$\left(\frac{1}{2}w_i\right)' = w_i$$

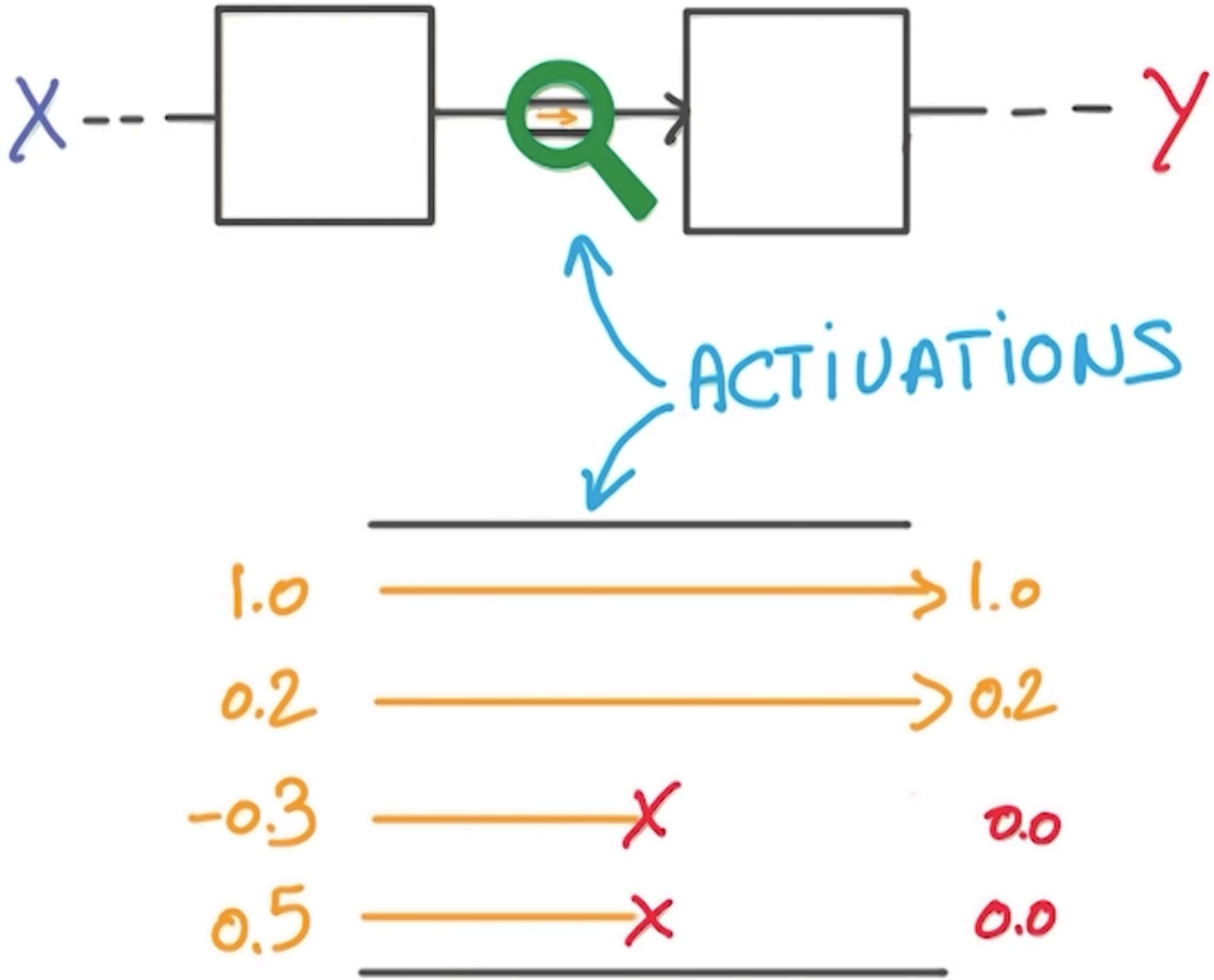
DERIVATIVE ?

- ○ $\frac{1}{3}\|w\|_2^3$
- ○ $w^T w$
- ✓ w

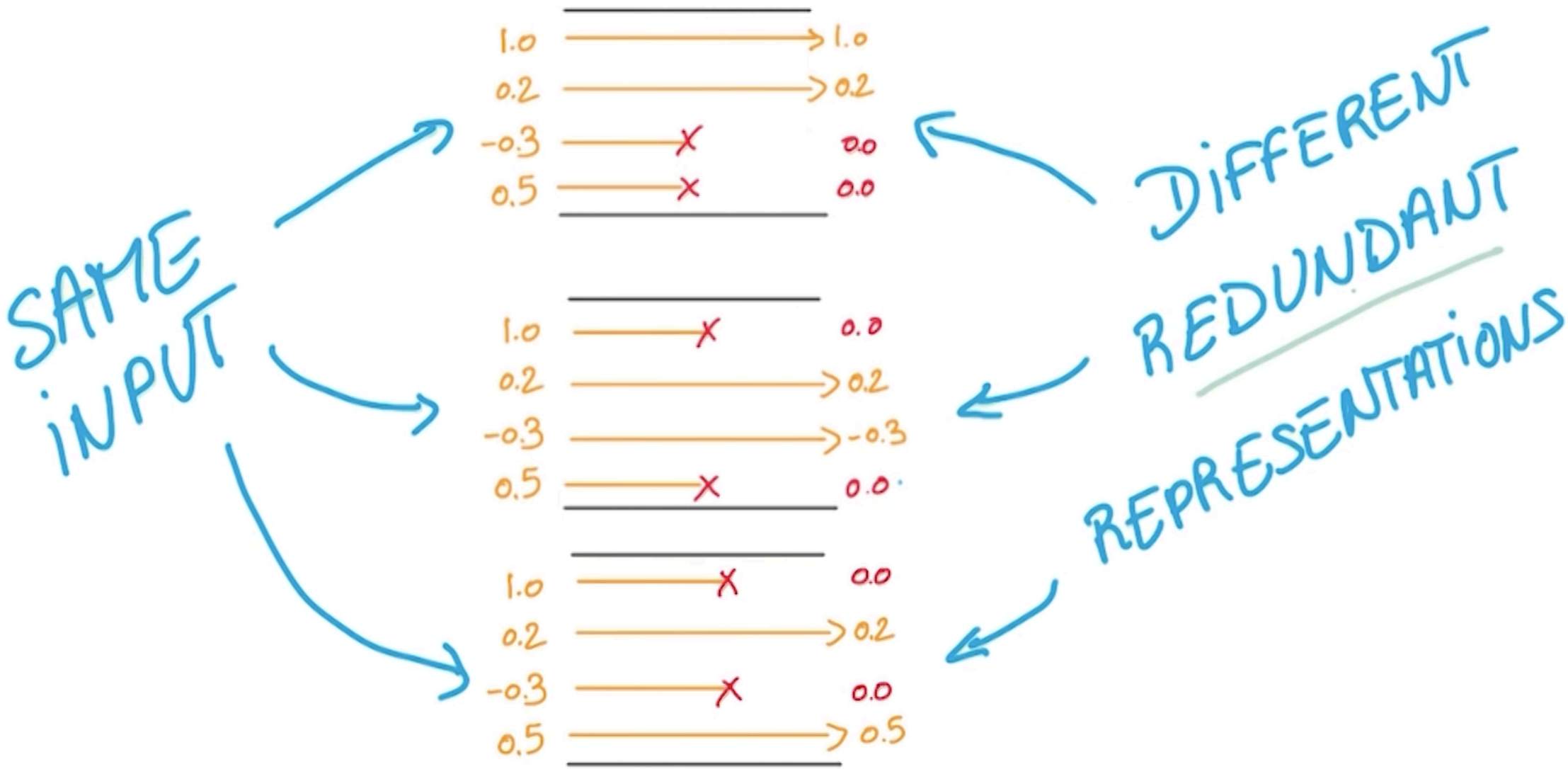
DROPOUT



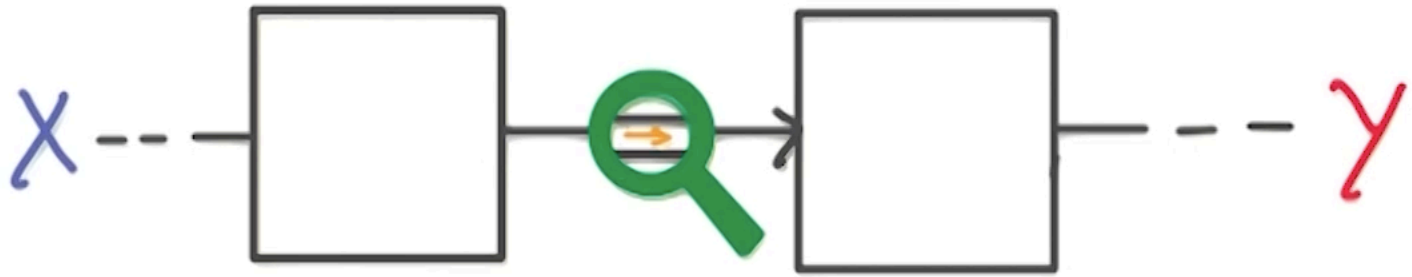
DROPOUT



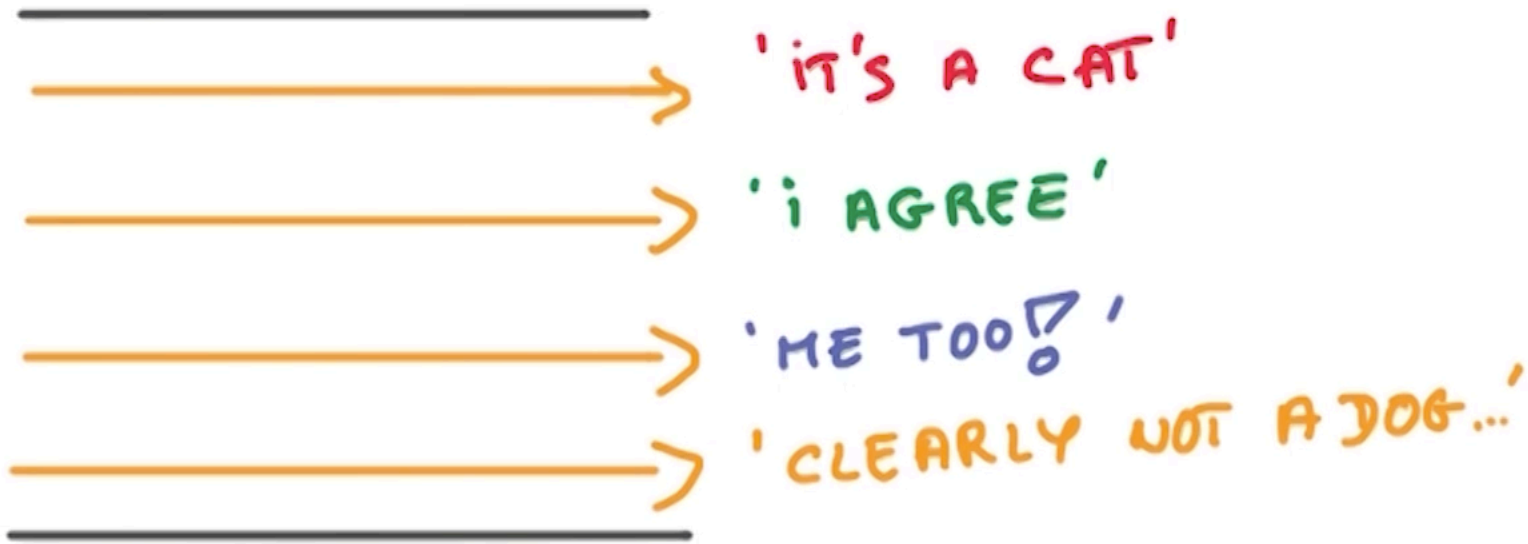
DROPOUT


SAME INPUT
1.0
0.2
-0.3
0.5
1.0
0.2
0.0
0.0
1.0
0.2
-0.3
0.5
0.0
0.2
-0.3
0.0
1.0
0.2
-0.3
0.5
0.0
0.2
0.0
0.5
DIFFERENT REDUNDANT REPRESENTATIONS

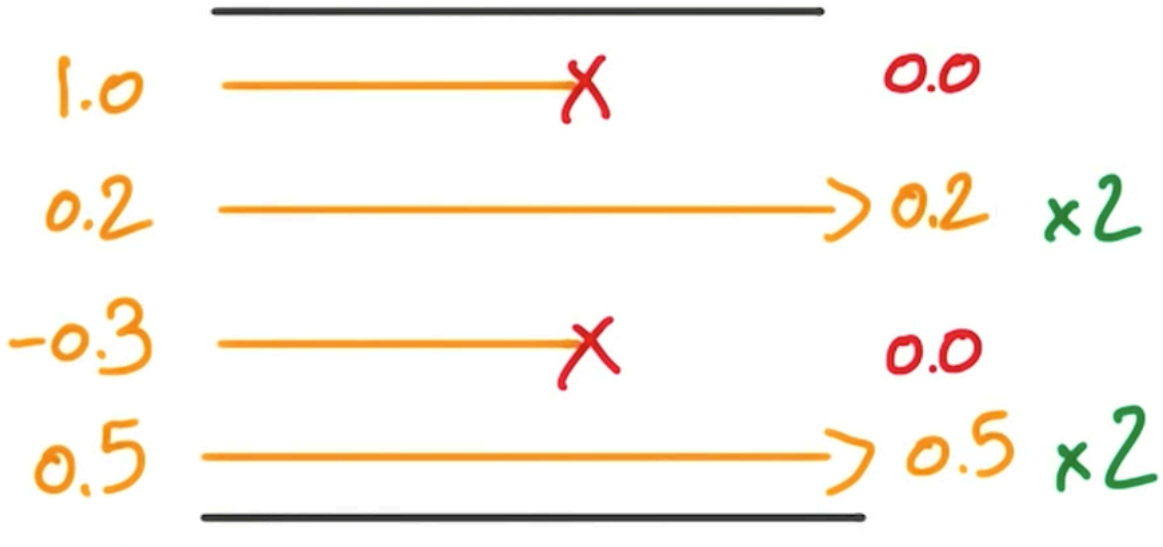
DROPOUT



REDUNDANT REPRESENTATIONS



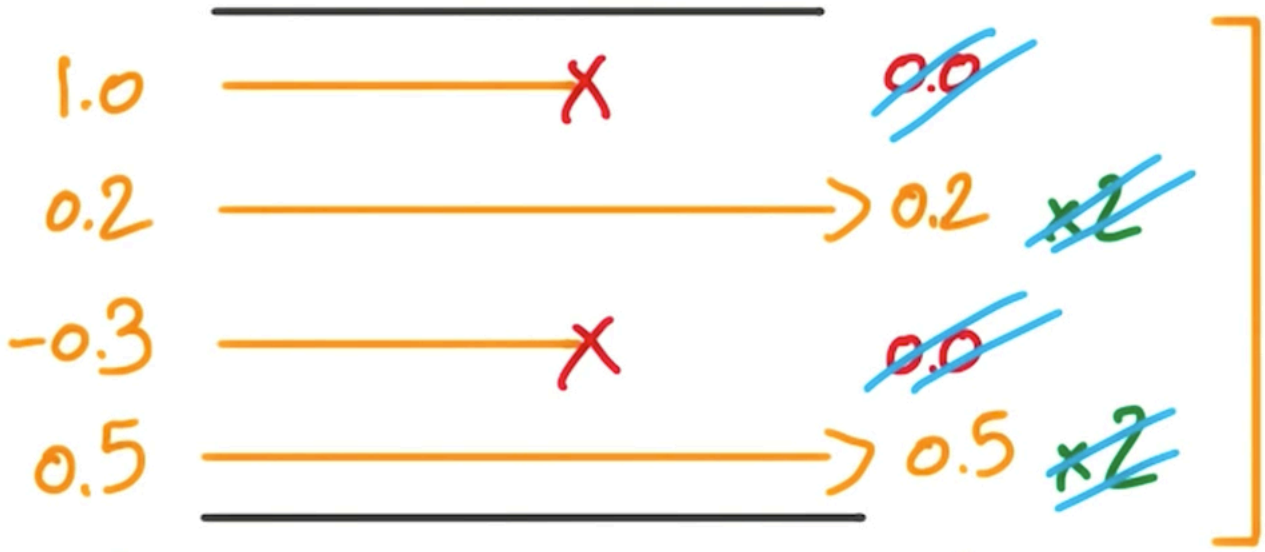
TRAINING



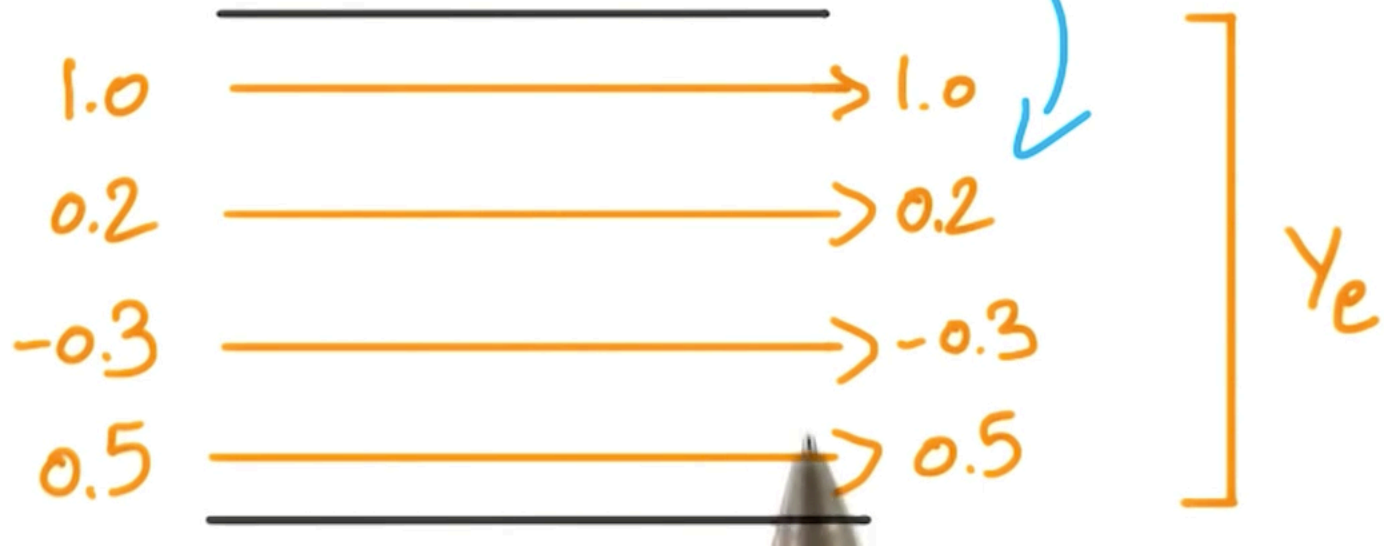
EVALUATION

$$Y_e \sim E(Y_t)$$

TRAINING



EVALUATION



$$Y_e \sim E(Y_t)$$