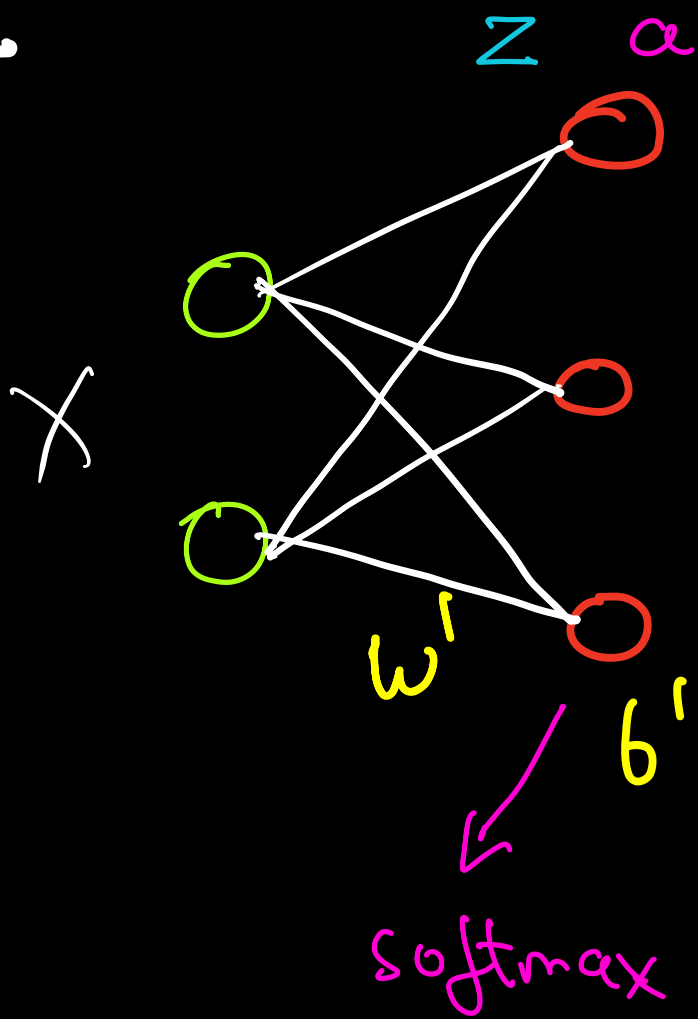
N-layer Networks
&
Activation Functions

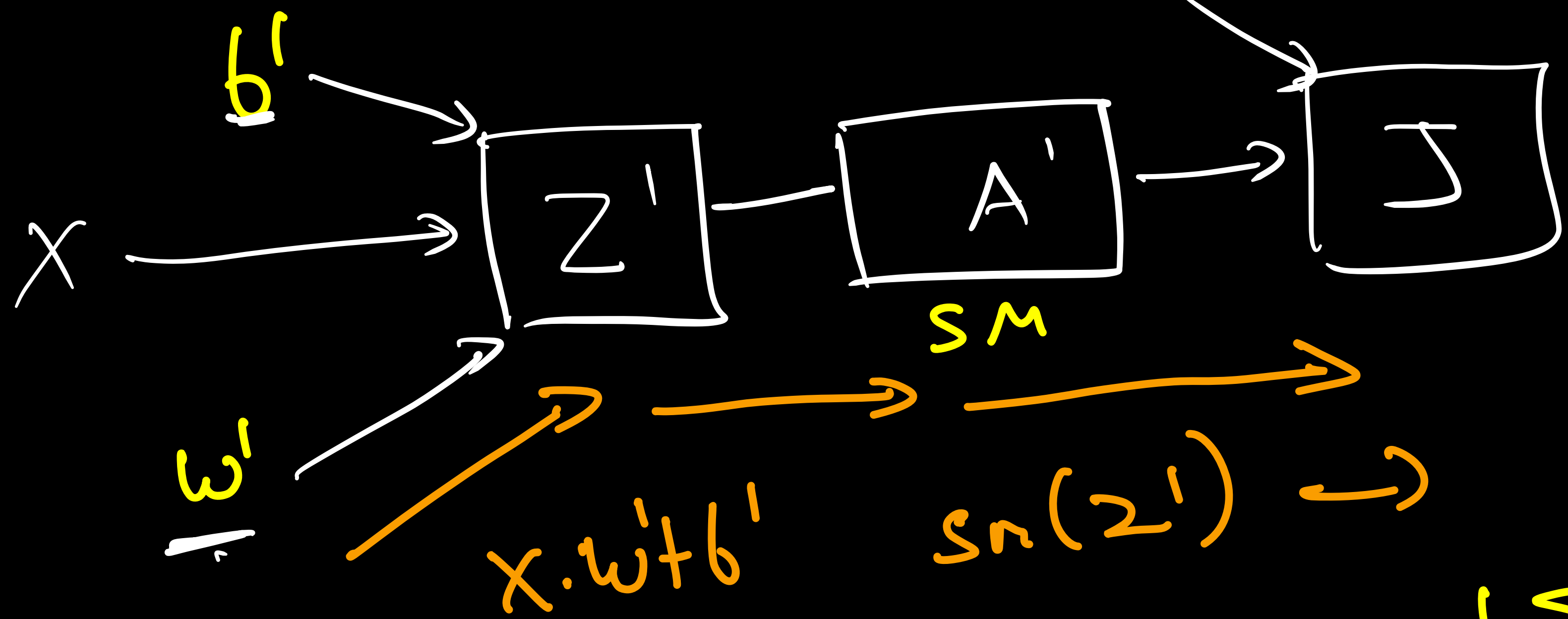
[Neural Networks]

Recap



$$\underset{\hat{y}}{\begin{bmatrix}0.1\\0.2\\0.1\end{bmatrix}} \sim \underset{y}{\begin{bmatrix}1\\0\\0\end{bmatrix}}$$

$$\Delta b' = \frac{\partial Z'}{\partial b'}$$



$$-\frac{1}{m}\sum y_i \log(A')$$

$$\Delta W' = \frac{\partial J}{\partial W'}$$

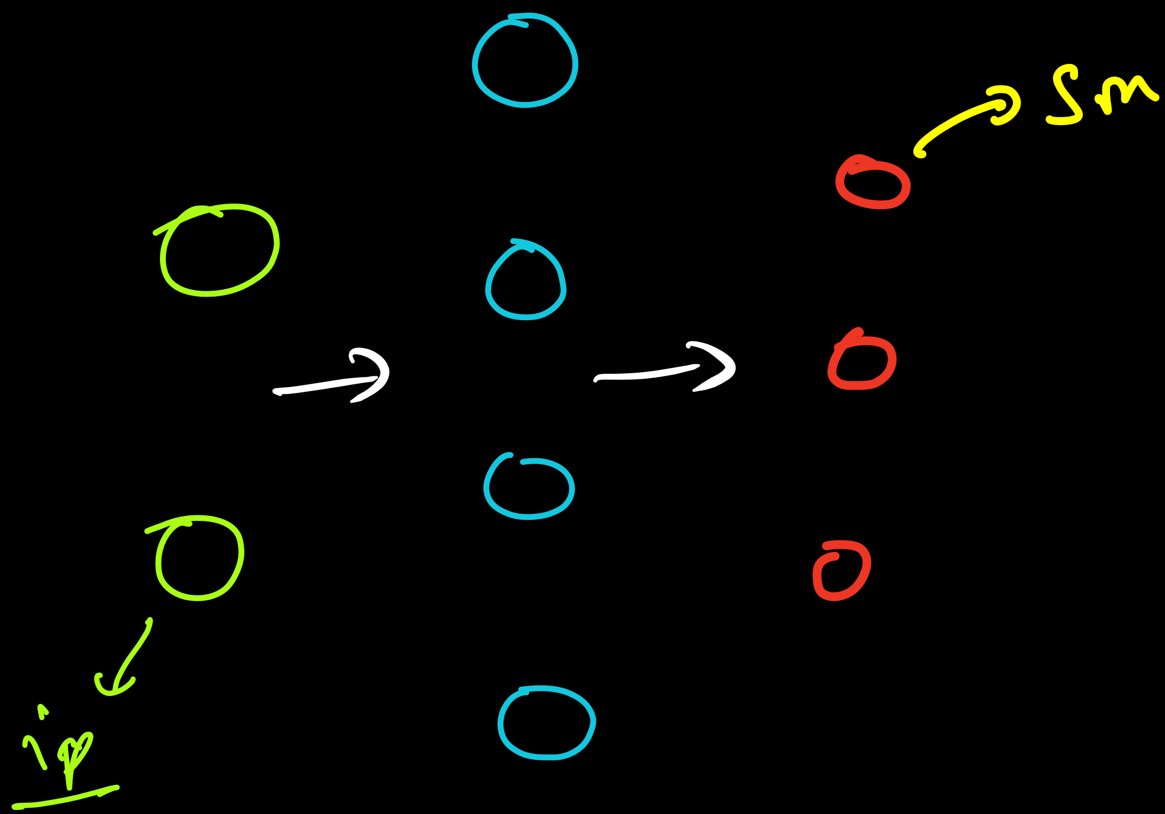
$$\frac{\partial Z'}{\partial W'} \quad \underbrace{\frac{\partial A'}{\partial Z'} \cdot \frac{\partial J}{\partial A'}}$$

$$\frac{\partial J}{\partial z} = (a - y)$$

$$\frac{\partial J}{\partial w} = (a - y) \cdot x$$

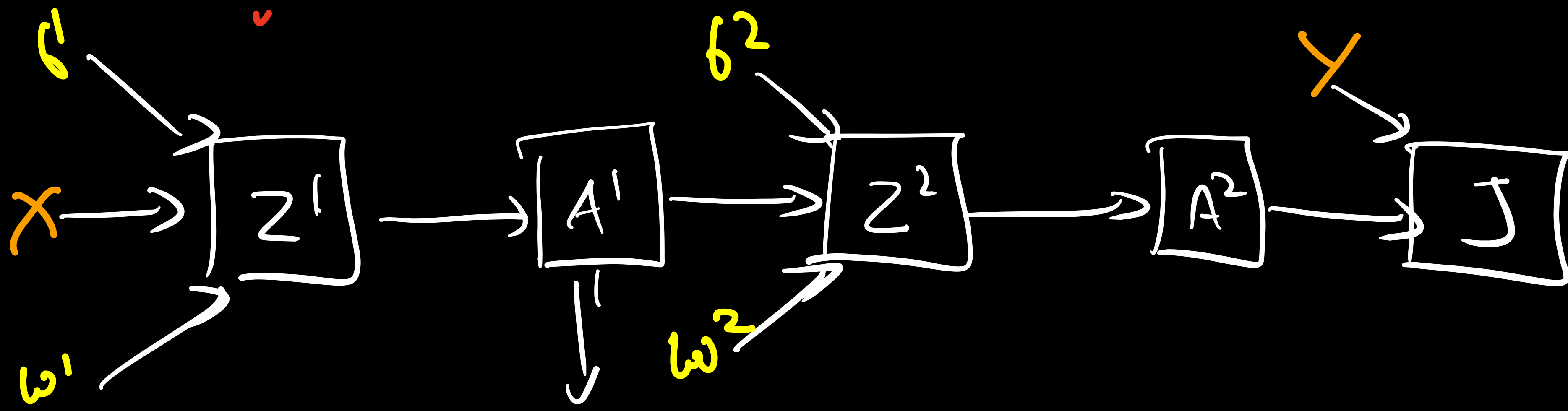
Linear DB ✓

→ We need to add one more layer with non-linear activation



$$\underbrace{2 \times 4 + 4}_{w^1 + b^1} \qquad \underbrace{4 \times 3 + 3}_{w^2 + b^2}$$

params $\rightarrow$ $12 + 15 = 27$



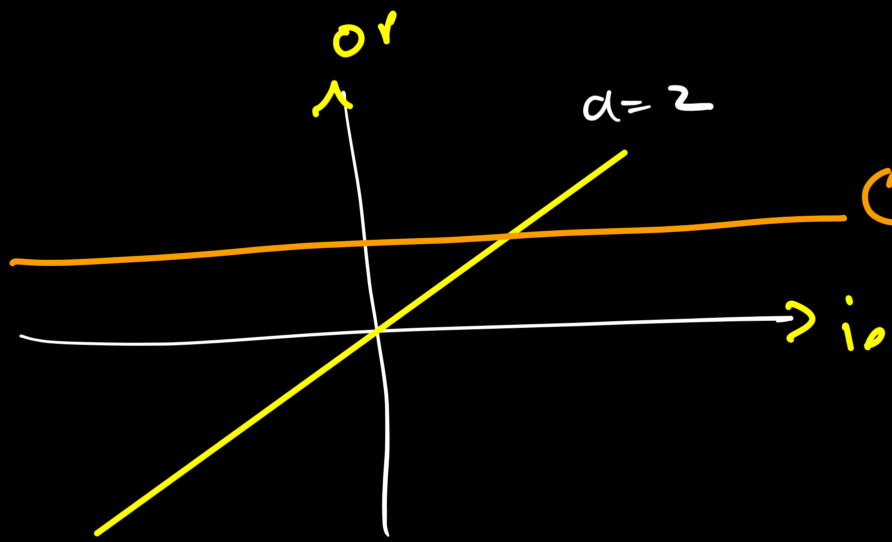
$$A^2\left(w^2 A^1(w^1 x + b^1) + b^2\right)$$

activation $\rightarrow$ We need to introduce small amount of non-linearity

Activation Functions

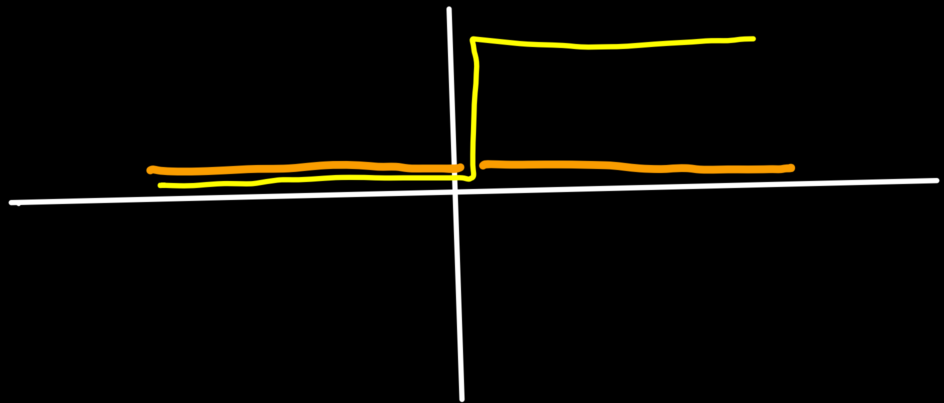
1) Linear = $A(z) = Z$

↳ may be used in last layer in Regression



① $da = 1$

2 Step



→ kills information

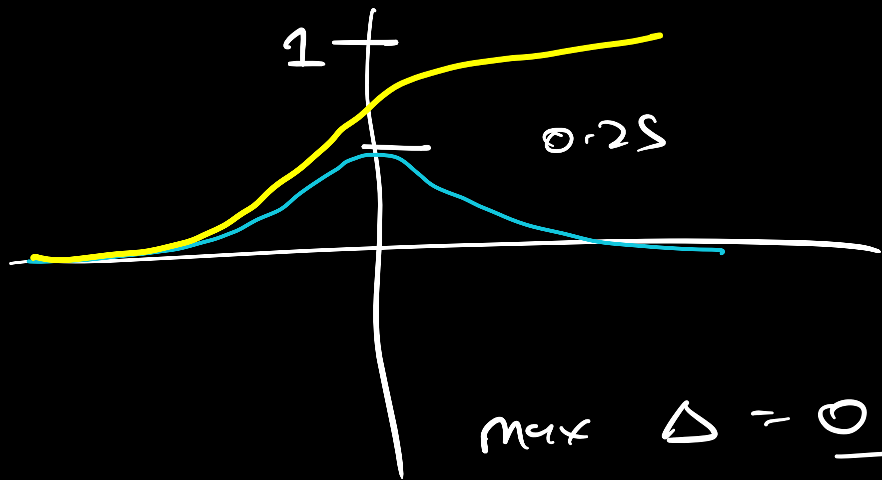
$$A(z) = \begin{cases} 1; & z > T \\ 0; & z < T \end{cases}$$

$da = 0$ (undefined at $z = T$)

↓

approx $= \underline{\underline{0}}$

3) Sigmoid



$$A(z) = \frac{1}{1+e^{-z}}$$

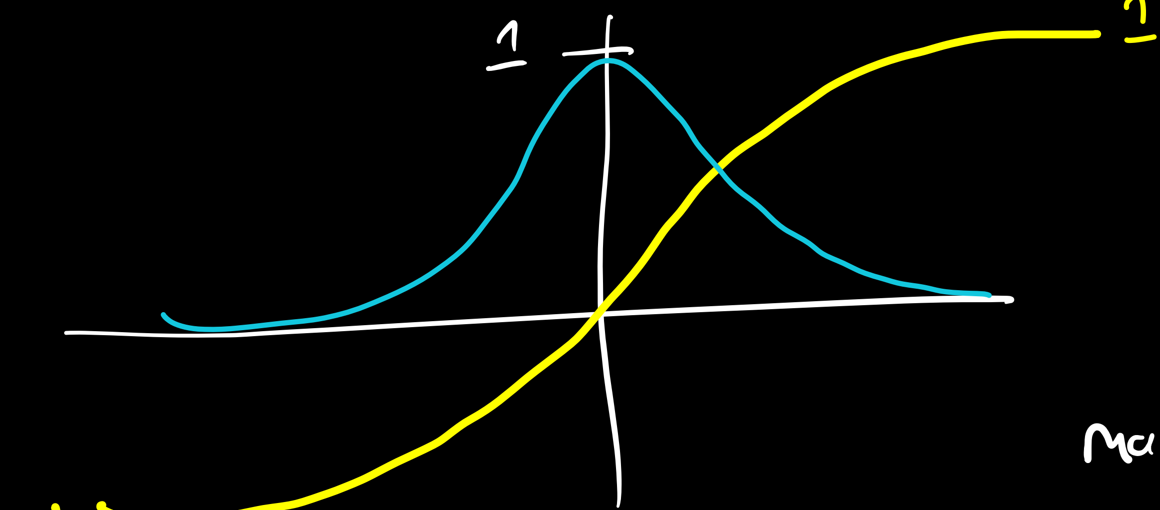
$$da = a(1-a)$$

max $\Delta = \underline{0.25}$

4) Softmax

$$A(z) = \frac{e^{z_i}}{\sum e^{z_j}}$$

5) Tanh



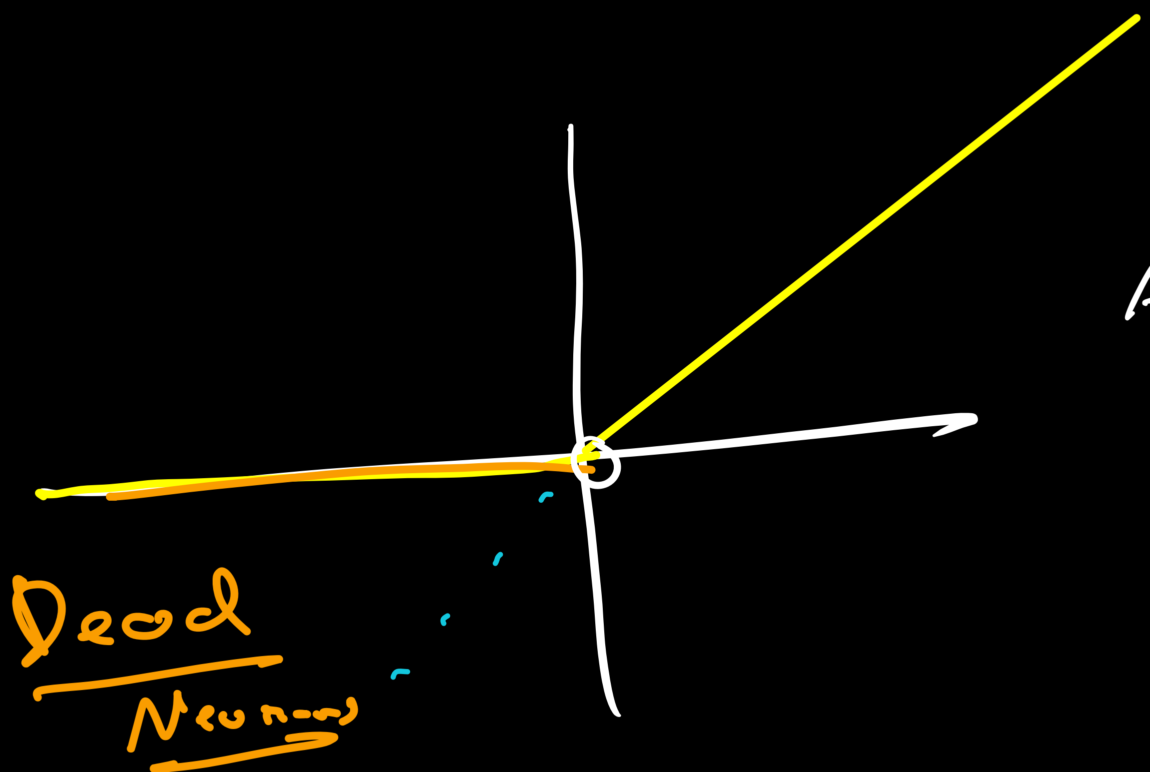
$$A(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

variation of sigmoid
with range $(-1, 1)$

max $\Delta = 2$

6) Relu

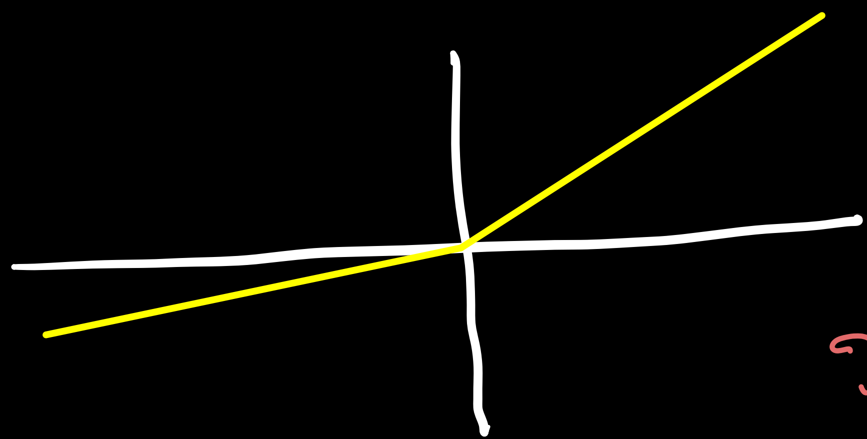
Rectified Linear Unit



$$A(z) = \begin{cases} z \,; & z > T \\ 0 \,; & z \leq T \end{cases}$$

$$dA = \begin{cases} 1 \,; & z > T \\ 0 \,; & z \leq T \end{cases}$$

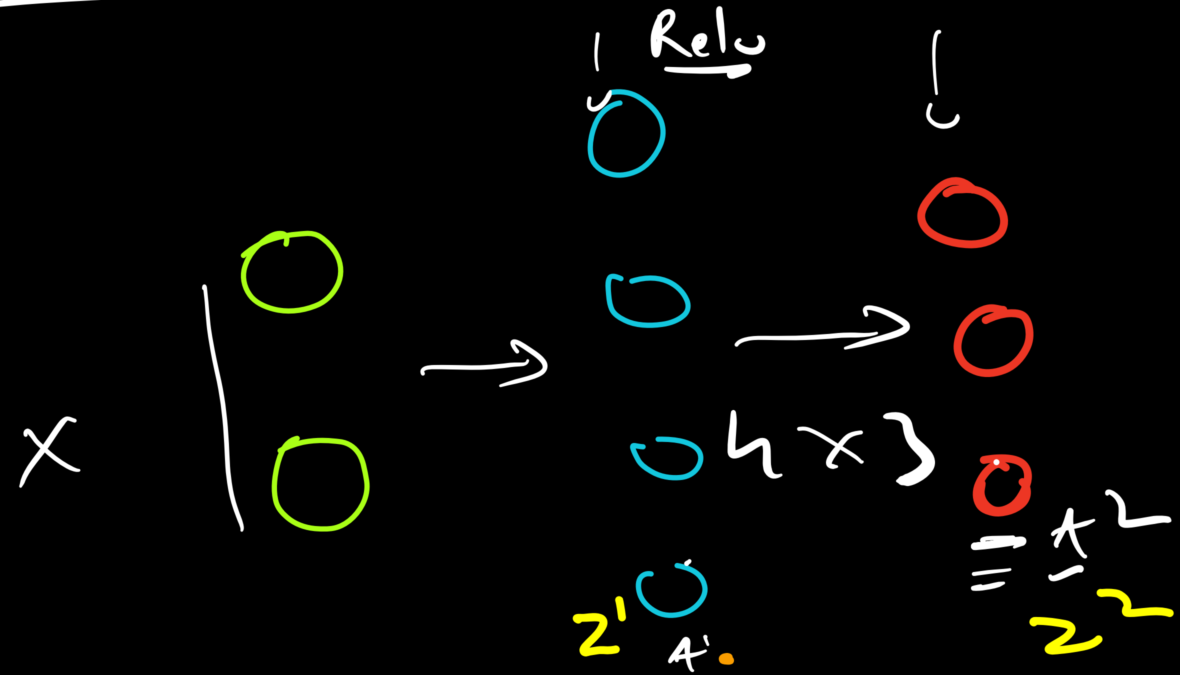
7) Leaky Relu



$$A = \begin{cases} Z\,; & Z > T \\ 0.1\,Z\,; & Z \leq T \end{cases}$$

This mitigates dead neuron

N-layered Networks



$A^0 \to X \to (300, 2)$

$Z^1 \to (300, 4)$

$w^1 \to (2, 4)$

$b^1 -$

$A^1 = 300, 4$

$Z^2 = 300, 3$

$w^2 \to (4, 3)$

$b^2 \to (1, 3)$

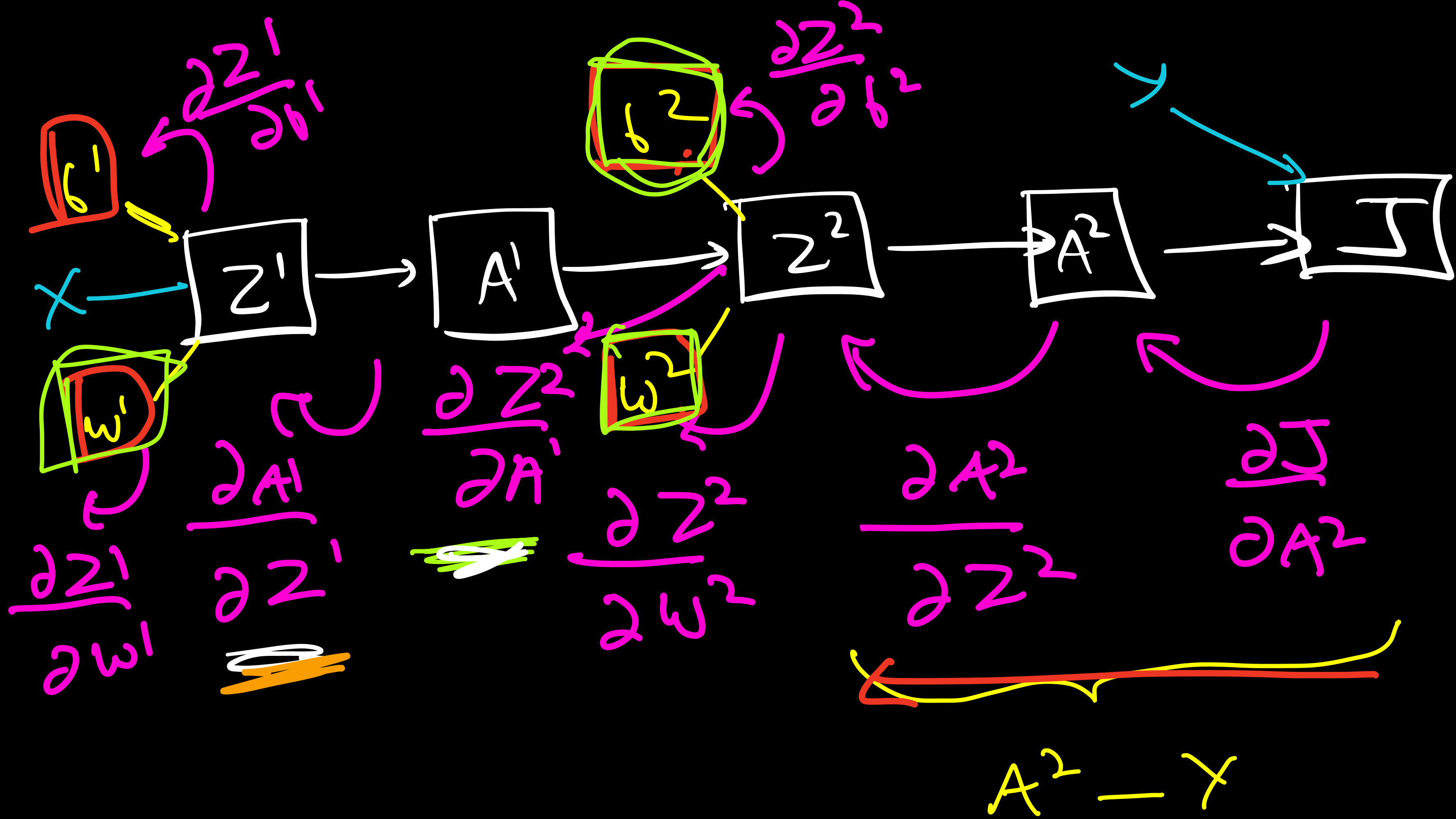
$\hat{y} = A^2 = 300, 3$

Fwd Prop

$$\begin{cases} Z^1 = w^1 x + b \\ A^1 = \max(0, Z^1) \\ Z^2 = w^2 A^1 + b^2 \\ A^2 = sm(z^2) \end{cases}$$

loss $\left[J = -\frac{1}{N} \sum_{i=1}^{n} y_i \log(a_i^2) \right.$

Back Prop


X
Y
Z¹
A¹
Z²
A²
J
b¹
W¹
b²
W²
∂Z¹ / ∂b¹
∂Z² / ∂b²
∂Z¹ / ∂W¹
∂A¹ / ∂Z¹
∂Z² / ∂A¹
∂Z² / ∂W²
∂A² / ∂Z²
∂J / ∂A²
A² − Y

1) $\Delta W^2 \rightarrow \frac{\partial J}{\partial W^2} = \underbrace{\frac{\partial J}{\partial A^2} \cdot \frac{\partial A^2}{\partial Z^2}} \cdot \frac{\partial Z^2}{\partial W^2}$

$4 \times 3 \longrightarrow \underbrace{(A^2 - Y)}_{(300 \times 3)} \cdot \underbrace{A^1}_{(300 \times 4)}$

$$\boxed{\frac{\partial Z^2}{\partial W^2} = \frac{\partial (W^2 \cdot A^1 + b^2)}{\partial W^2} = A^1}$$

$\Delta W^2 \Rightarrow (A^1.T)(A^2 - Y)/m$

$\frac{300}{avg}$ $\frac{(4 \times 1)(1 \times 3)}{} = (4 \times 300)(300 \times 3)$

$\boxed{4 \times 3}$

1) $\Delta b^2 \rightarrow \dfrac{\partial J}{\partial b^2} = \underbrace{\dfrac{\partial J}{\partial A^2} \cdot \dfrac{\partial A^2}{\partial Z^2}} \cdot \dfrac{\partial Z^2}{\partial b^2}$

$(1, 3)$

$(A^2 - Y) \cdot 1$

$(300 \times 3) \cdot 1$?!

mean(axis = 1)

Remaining in next class

1) $\dfrac{\partial Z^2}{\partial A^1} = \dfrac{\partial (W^2 A^1 + b^2)}{\partial A^1}$

$= W^2$

$\Delta A^1 = \dfrac{\partial J}{\partial A^2} \cdot \dfrac{\partial A^2}{\partial Z^2} \cdot \dfrac{\partial Z^2}{\partial A^1}$

$(A^2 - Y) \cdot W^2$

4) $\dfrac{\partial A^1}{\partial Z^1} = \underline{\text{Relu}}$

$\hookrightarrow R^1 = \begin{cases} 1 \; ; & Z^1 > 0 \\ 0 \; ; & Z^1 \leq 0 \end{cases}$

$$\frac{\partial J}{\partial Z^1} = \underbrace{\frac{\partial J}{\partial A^2} \cdot \frac{\partial A^2}{\partial Z^2} \cdot \frac{\partial Z^2}{\partial A^1}} \cdot \frac{\partial A^1}{\partial Z^1}$$

ΔA^1 ;

$\Delta Z^1 = \Delta A^1$ ✓

$\Rightarrow \Delta Z^1 [Z^1 < 0] = 0$ (override)

5) $$\Delta w^1 = \frac{\partial J}{\partial w^1} = \underbrace{\underbrace{\frac{\partial J}{\partial A^2} \cdot \frac{\partial A^2}{\partial Z^2} \cdot \frac{\partial Z^2}{\partial A^1}}_{\Delta A^1} \cdot \frac{\partial A^1}{\partial Z^1}}_{\Delta Z^1} \cdot \underbrace{\frac{\partial Z^1}{\partial w^1}}_{X}$$

$$\Delta Z^1 \cdot X$$

$/m$

6) $$\Delta b^1 = \frac{\partial J}{\partial b^1} = \underbrace{\underbrace{\frac{\partial J}{\partial A^2} \cdot \frac{\partial A^2}{\partial Z^2} \cdot \frac{\partial Z^2}{\partial A^1}} \cdot \frac{\partial A^1}{\partial Z^1}} \cdot \frac{\partial Z^1}{\partial b^1}$$

means $\longleftarrow \quad \Delta Z^1 \cdot 1$