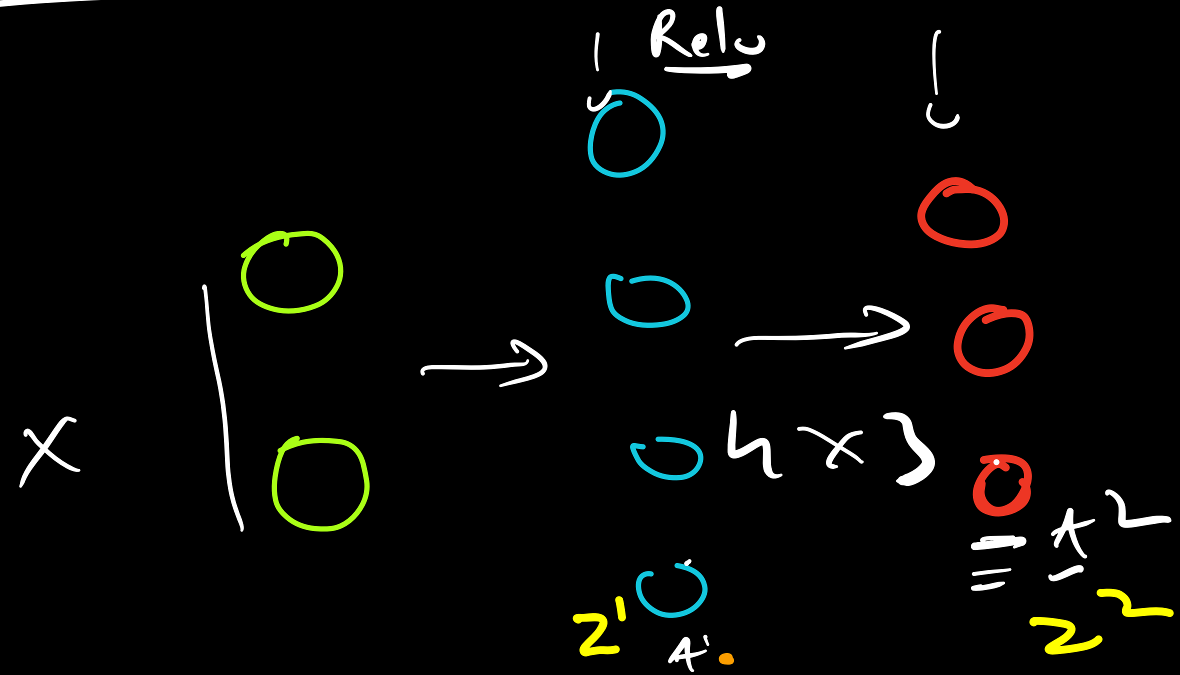
N-layer Networks

[Neural Networks]

N-layered Networks



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

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

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

$b^1 \to$

$A^1 = 300, 4$

$Z^2 = 300, 3$

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

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

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

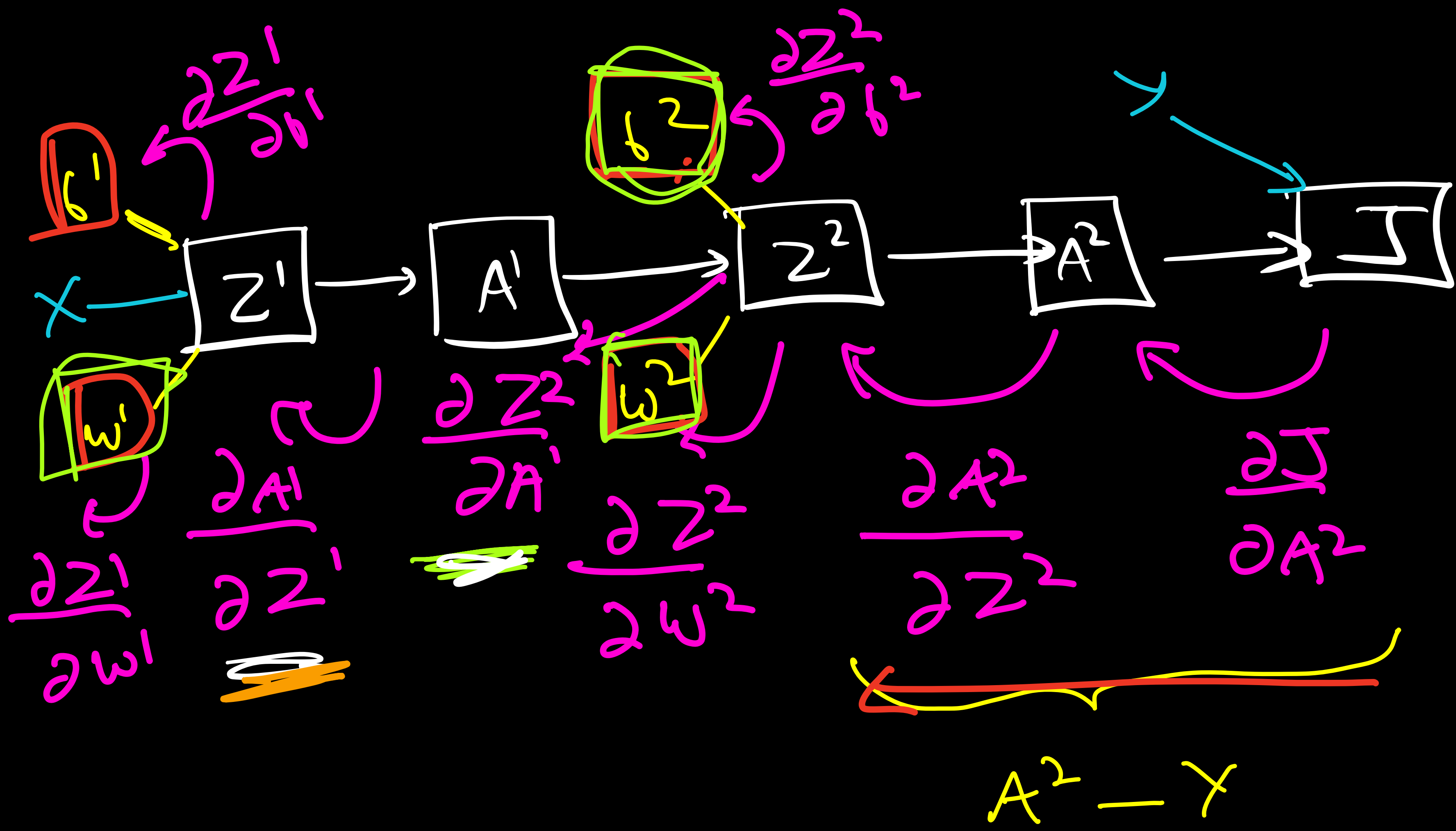
Fwd Prop

$$\left\{\begin{aligned} Z^1 &= w^1 x + b \\ A^1 &= \max(0, Z^1) \\ Z^2 &= w^2 A^1 + b^2 \\ A^2 &= sm(z^2) \end{aligned}\right.$$

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

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

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

$$= A^1$$

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

$\cancel{306}$ $\cancel{\text{any}}$ $\cancel{(4 \times 1)(1 \times 3)}$ $= (4 \times 300)(300 \times 3)$

$\boxed{4 \times 3}$

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

$$(1, 3) \qquad (A^2 - Y) \cdot 1$$

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

mean(axis = 1)

Remaining in next class

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

$$= W^2$$

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

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

4) $\frac{\partial A^1}{\partial Z^1} = \text{Relu} \rightarrow R' = \begin{cases} 1 ; & Z^1 > 0 \\ 0 ; & Z^1 \leq 0 \end{cases}$

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

$$\Delta Z^1 = \Delta A^1$$

$$\Rightarrow \Delta Z^1 [Z^1 < 0] = 0 \quad \text{(override)}$$

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

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

$$/\underline{m}$$

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 $\leftarrow$ $\Delta Z^1 \cdot 1$