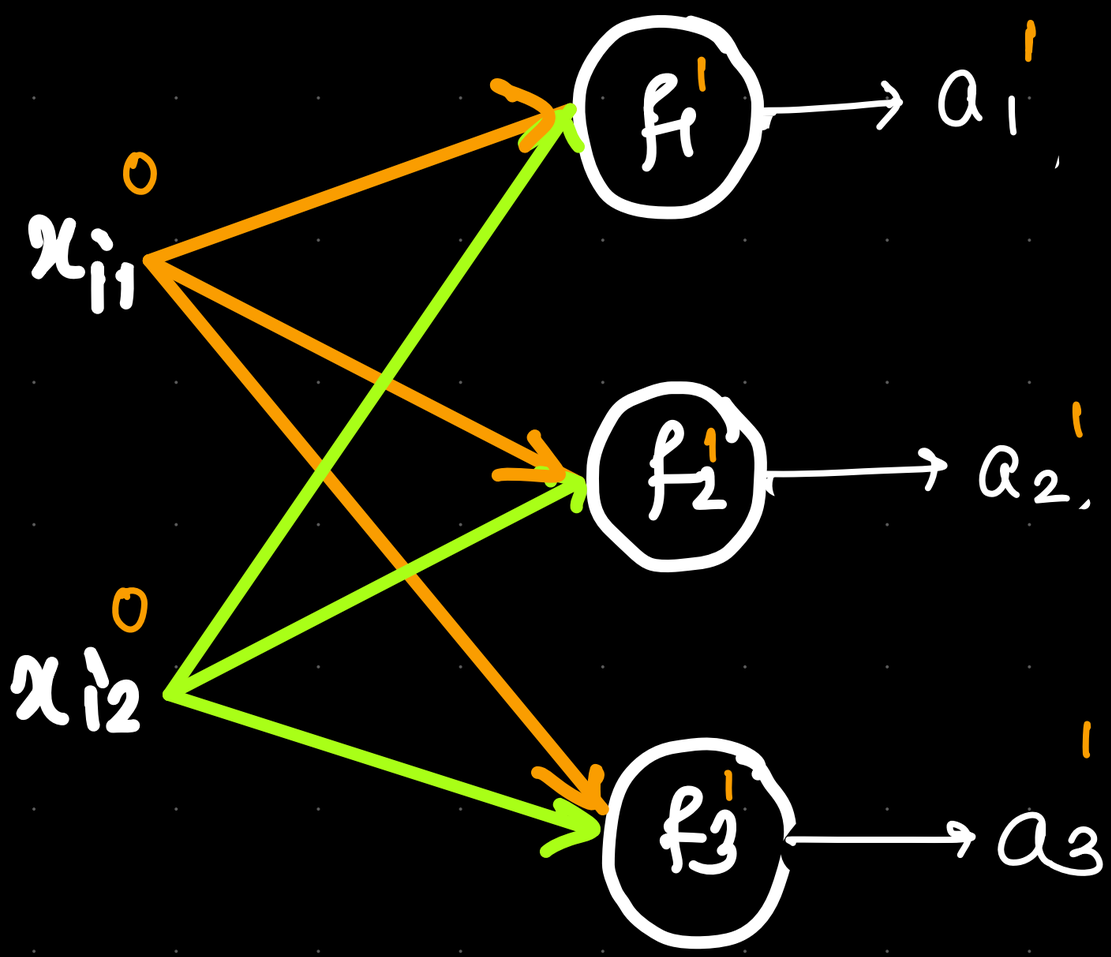
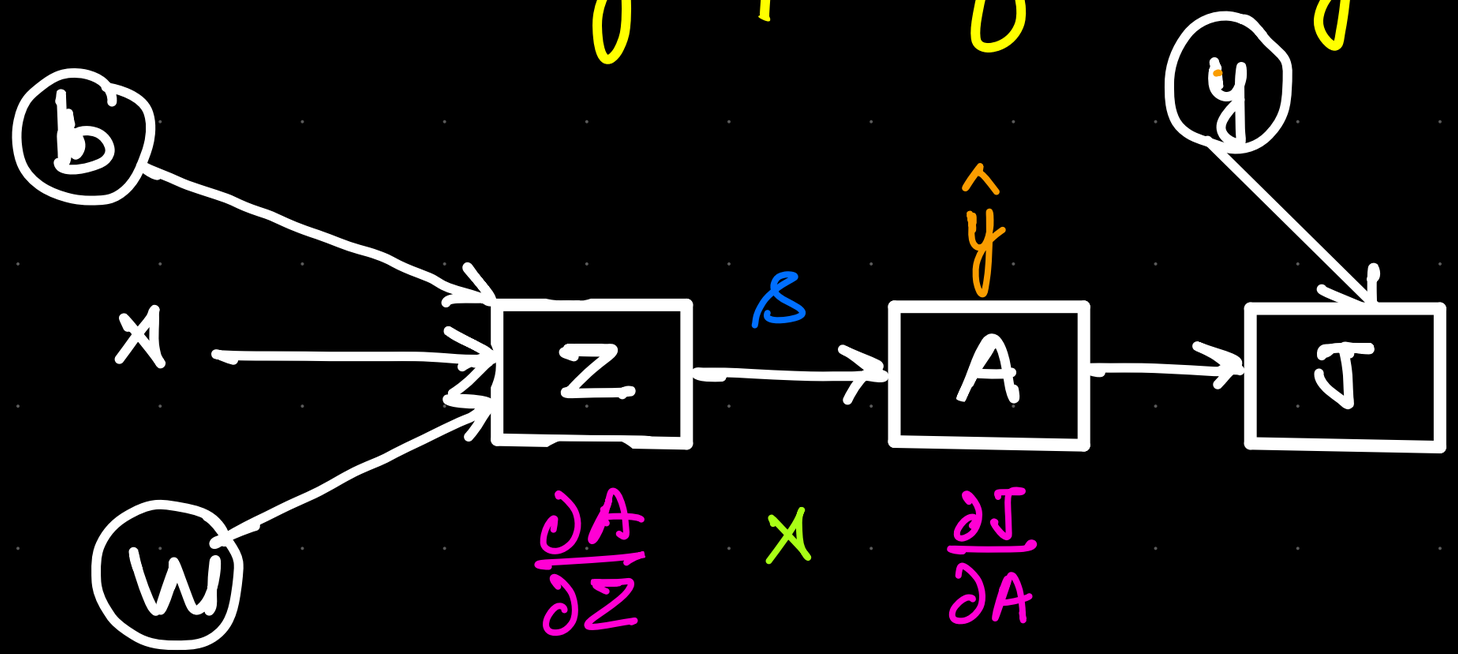
NN : lecture-5

Backprop for N-layer NN

1 - layer NN (Softmax Classifies)



Computational graph of 1-Layer NN



$J = \sum y \log \hat{y}$

$\frac{\partial J}{\partial W}$ $\frac{\partial J}{\partial b}$

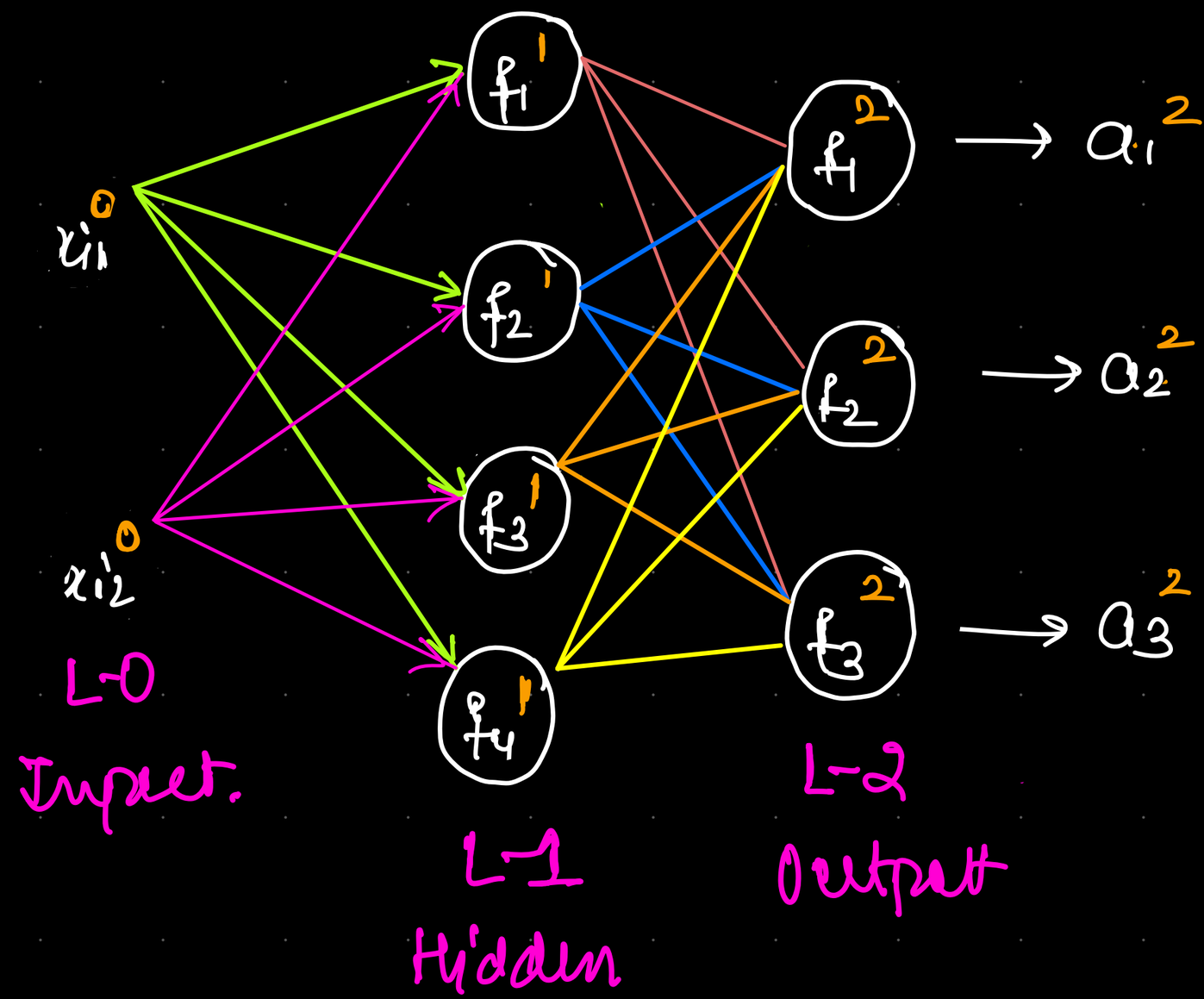
$\frac{\partial A}{\partial Z}$ X $\frac{\partial J}{\partial A}$

$\frac{\partial Z}{\partial W} = X$ X $\frac{\partial J}{\partial Z} = P - Y = A - Y$

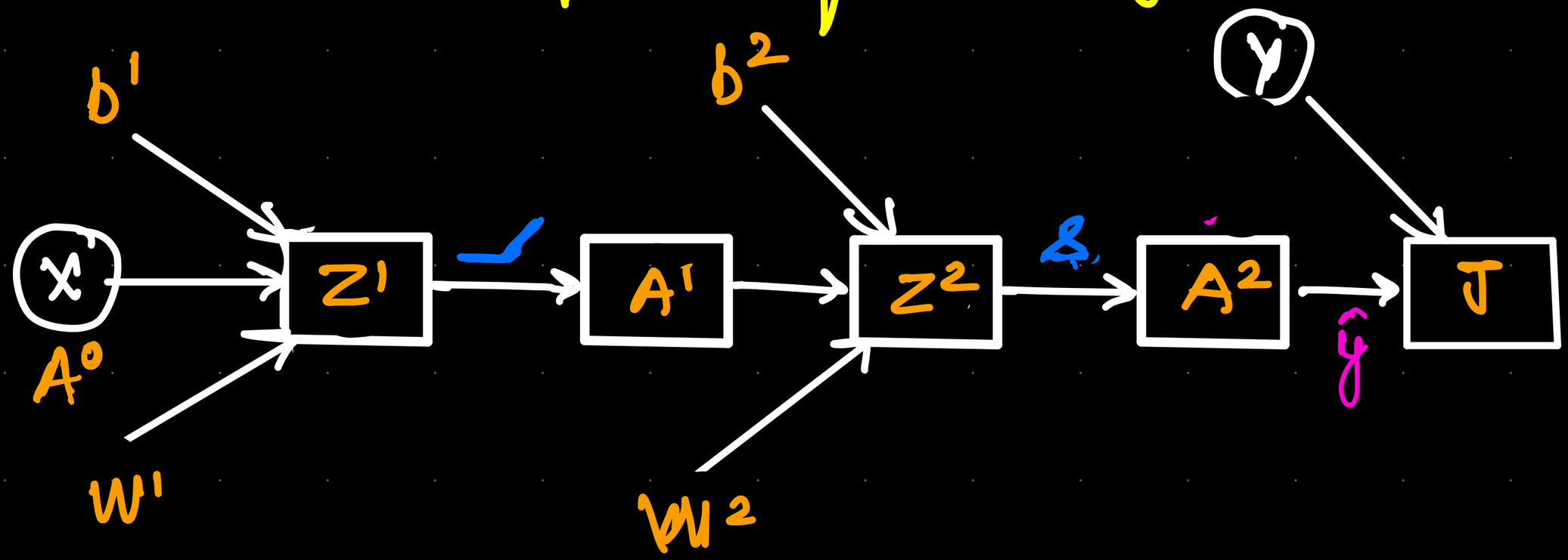
$\frac{\partial (XW+b)}{\partial W} = X$

$$\frac{\partial J}{\partial W} = (A-Y).X$$

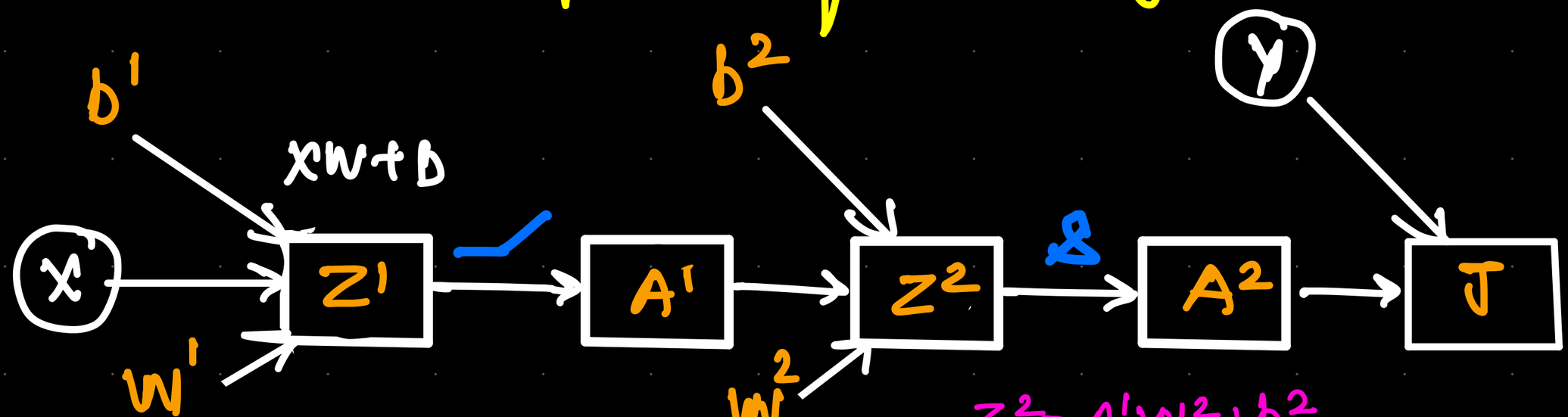
N-layer NN


f_1^1
f_2^1
f_3^1
f_4^1
f_1^2
f_2^2
f_3^2
x_1^0
x_2^0
$\rightarrow a_1^2$
$\rightarrow a_2^2$
$\rightarrow a_3^2$
L-0
Input.
L-1
Hidden
L-2
Output

Computational graph for N-layer NN



Computational graph for N-layer NN



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

$W^1 \rightarrow (2, 4)$

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

$b^1 \rightarrow (1, 4)$

$A^1 \rightarrow (300, 4)$

$W^2 \rightarrow (4, 3)$

$Z^2 \rightarrow (300, 3)$

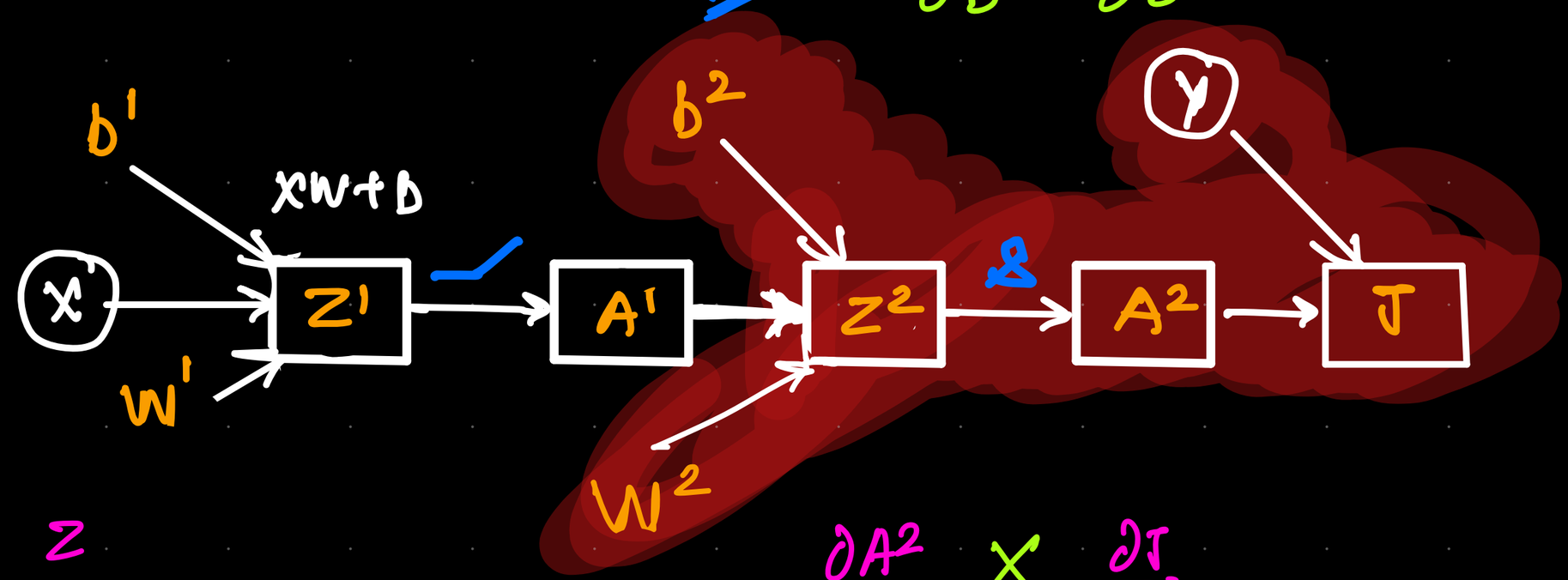
$b^2 \rightarrow (1, 3)$

$A^2 \rightarrow (300, 3)$

$y \rightarrow (300, 3)$

Backward Prop (∂W^2)

$\frac{\partial W^1}{\partial b^1}$ $\frac{\partial W^2}{\partial b^2}$



z

$\frac{\partial A^2}{\partial Z^2} \times \frac{\partial J}{\partial A^2}$

$\frac{\partial (A^1 W^2 + b^2)}{\partial W^2}$ ← $\frac{\partial Z^2}{\partial W^2}$ X $\frac{\partial J}{\partial Z^2}$ → $A2 - Y$

A^1

$\frac{\partial J}{\partial W^2} \longrightarrow \partial W^2$

Backward Prop (dW^2)

$$A^1, \partial Z^2 = \partial W^2$$

$A^1 \longrightarrow (300, 4)$

$A^2 \longrightarrow (300, 3)$

$\partial W^2 \longleftrightarrow W^2 \longrightarrow (4, 3)$

// to $\partial W =$ $X^T . \partial Z$

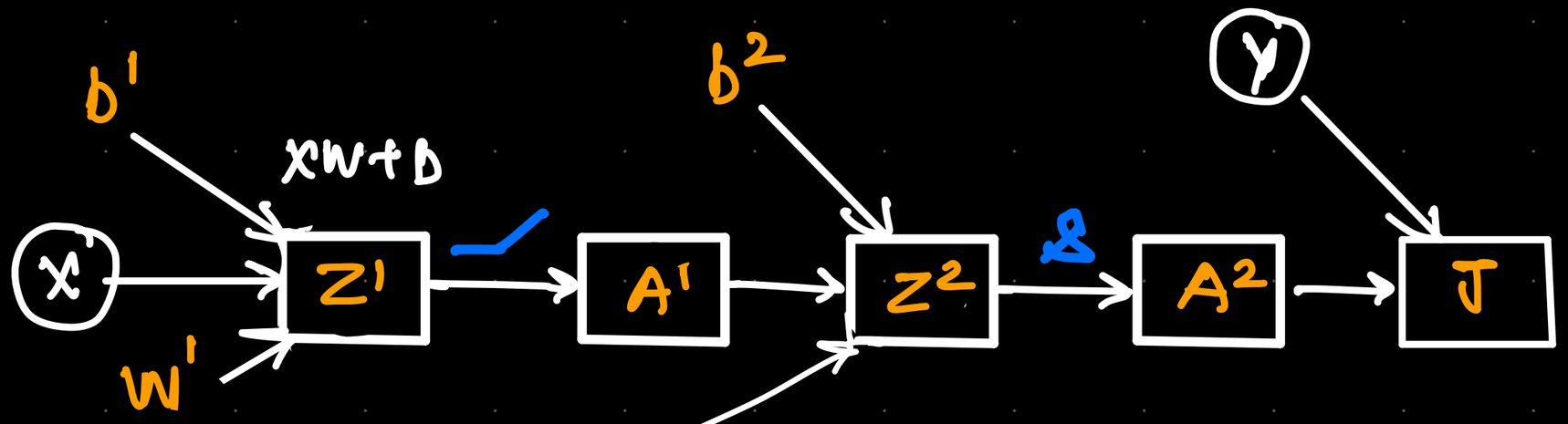
$$\partial W^2 = A^{1T} . \partial Z^2 \quad / m$$

But why are we dividing by m ?

$$w \longrightarrow w - \alpha \frac{1}{m} \sum_{i}^{m} \frac{\partial L}{\partial w}$$

∂w

Backward Prop (∂b^2)



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

$$\frac{\partial J}{\partial Z^2} \rightarrow A^2 - y$$

$$1$$

$$\frac{\partial J}{\partial b^2} \overset{***}{=} 1, A^2 - y$$

Backward Propagation (db^2)

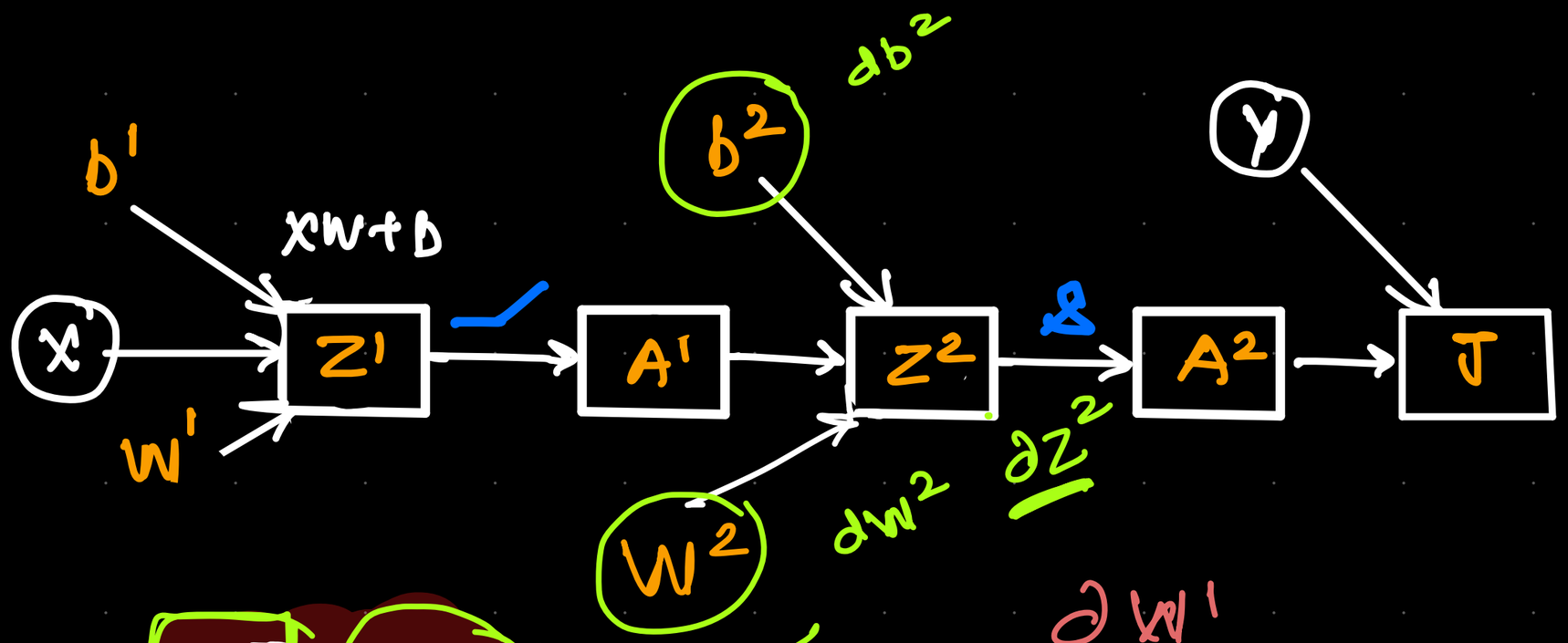
$A^2 - Y \rightarrow (300, 3)$

$db^2 \rightarrow (1, 3)$

b^2

$$db^2 = np.sum(A^2 - Y, axis = 0, keepdims = True)/m$$

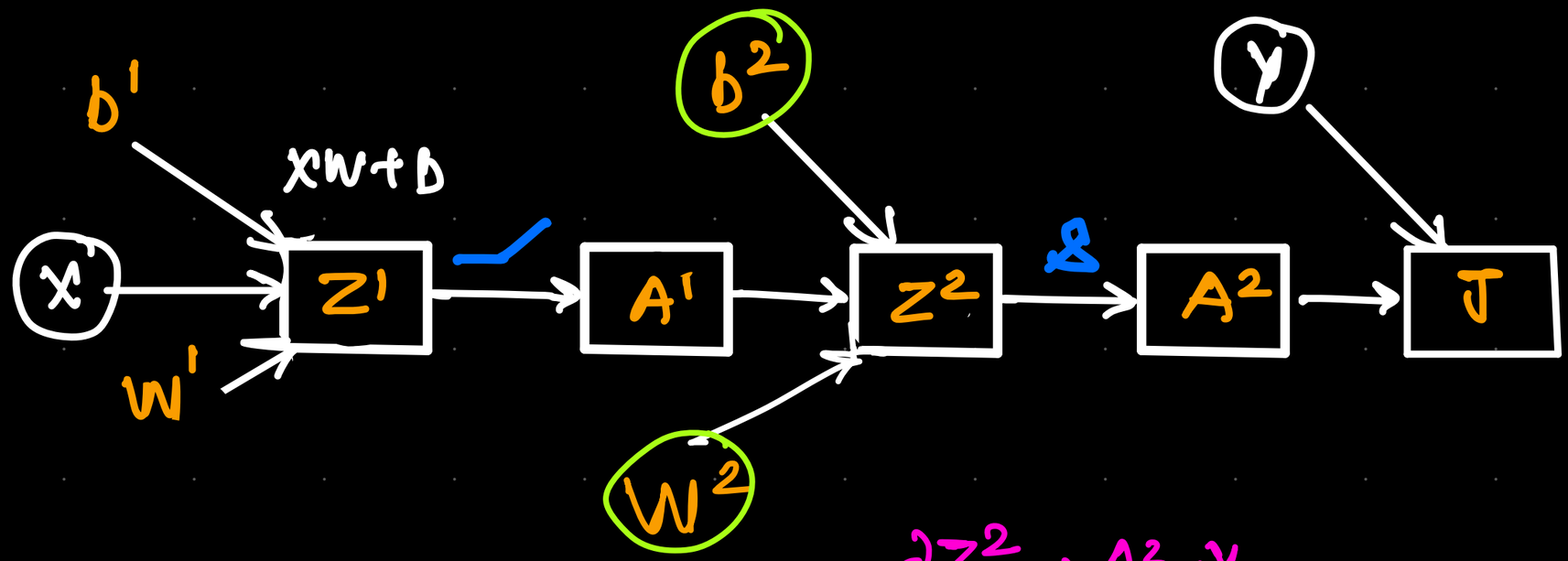
Are we done here ?



$$\partial W^2 = A^{1T} \cdot \partial Z^2 / m$$

$$\partial b^2 = np.sum(\partial Z^2) / m$$

Backward Prop (dW^1)



$\frac{\partial Z^1}{\partial W^1}$: 3 : X

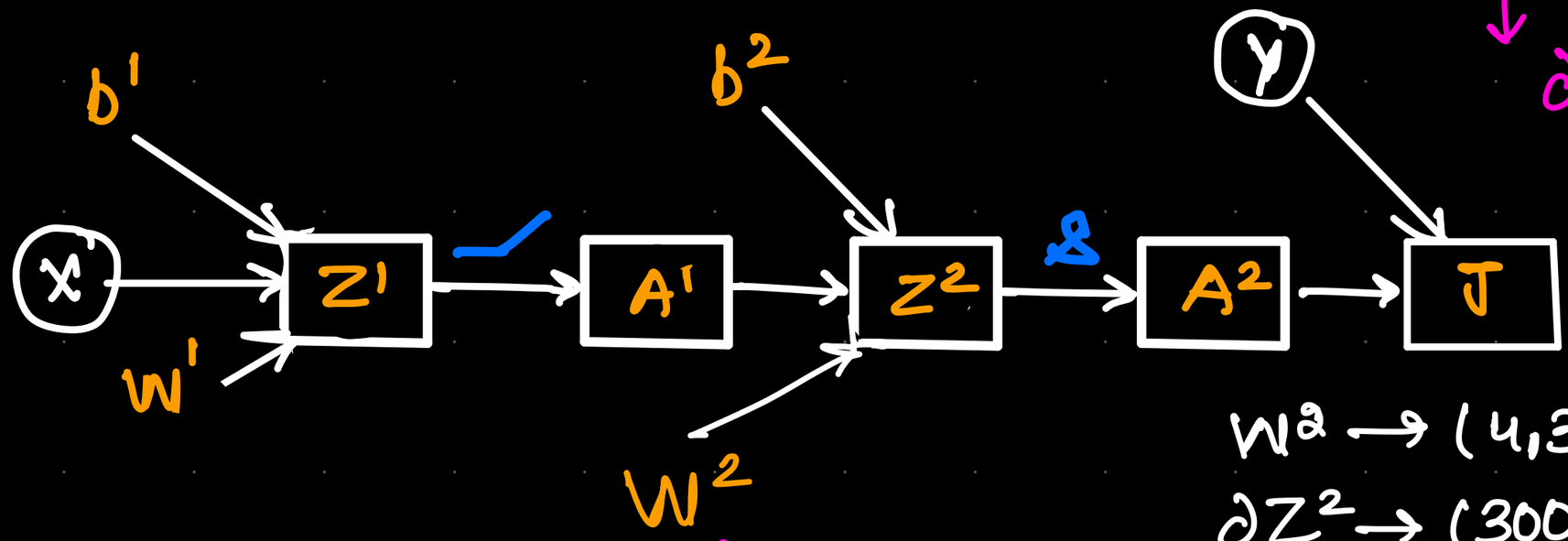
$\frac{\partial A^1}{\partial Z^1}$: 2 :

$\frac{\partial Z^2}{\partial A^1}$: 1 :

$\partial Z^2 \rightarrow A^2 - Y$: 0

Backward Prop (∂A^1)

$\frac{\partial Z^1}{\partial W^1}$ = 3 $\quad$ $\frac{\partial A^1}{\partial Z^1}$ = 2 $\quad$ $\frac{\partial Z^2}{\partial A^1}$ = 1 $\quad$ dZ^2 = 0 $\rightarrow \partial A^1$



$W^2 \rightarrow (4,3)$

$\partial Z^2 \rightarrow (300,3)$

$\partial A^1 \rightarrow (300,4)$

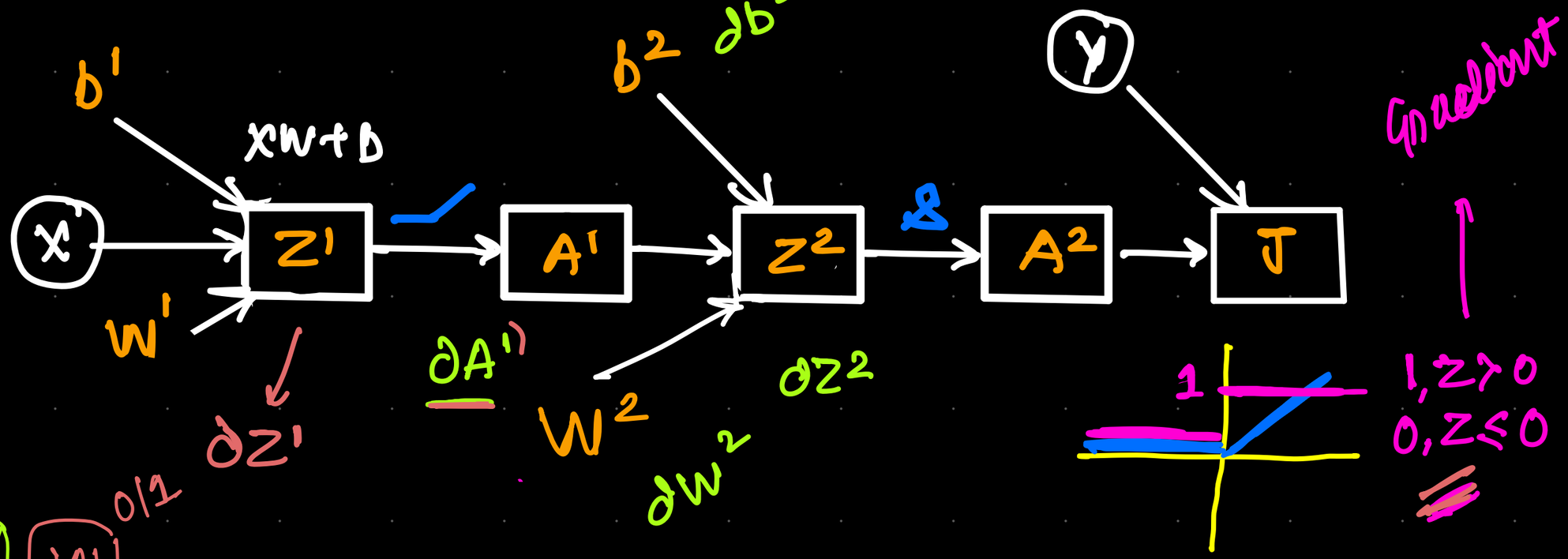
① $\frac{\partial Z^2}{\partial A^1} = \frac{\partial (A^1 W^2 + b^2)}{\partial A^1} = W^2$

$\partial A_1 = \frac{\partial J}{\partial Z^2} \cdot \frac{\partial Z^2}{\partial A^1}$

$$\boxed{\partial A^1 = \partial Z^2 \cdot W^{2T}}$$

Backward Prop (∂Z^1)

$\frac{\partial Z^1}{\partial W^1}$ 3 $\quad$ $\frac{\partial A^1}{\partial Z^1}$ 2 $\quad$ $\frac{\partial Z^2}{\partial A^1}$ 1 $\quad$ dZ^2



$$\partial Z^1 = \boxed{\partial A_1} \cdot \boxed{\frac{\partial A^1}{\partial Z_1}} \quad 0/1$$

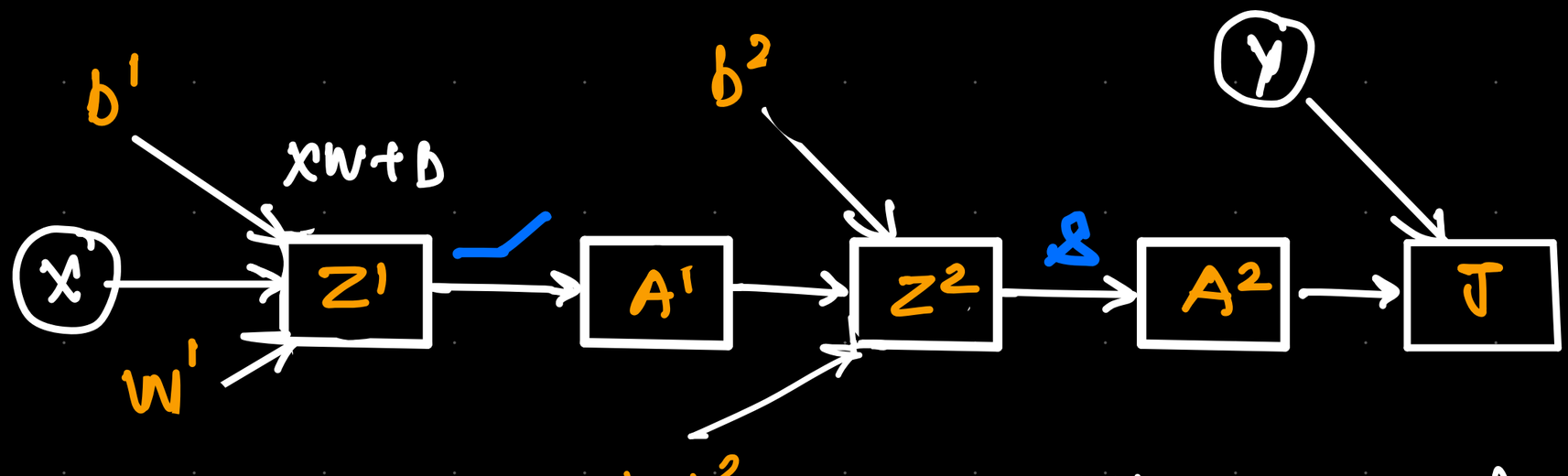
$$\frac{\partial A_1}{\partial Z_1} = \frac{\partial(\text{ReLU}(Z_1))}{\partial Z_1}$$

$$\partial Z^1 = \boxed{\partial A^1}$$

$$\partial Z^1[Z^1 \leq 0] = 0$$

Backward Prop(∂W^1)

$\frac{\partial Z^1}{\partial W^1}$ 3 $\quad$ $\frac{\partial A^1}{\partial Z^1}$ 2 $\quad$ $\frac{\partial Z^2}{\partial A^1}$ 1 $\quad$ dZ^2



$$\frac{\partial Z^1}{\partial W^1} = \frac{\partial (XW^1 + b^1)}{\partial W^1} = X$$

$\partial Z^1 \longrightarrow (300, 4)$

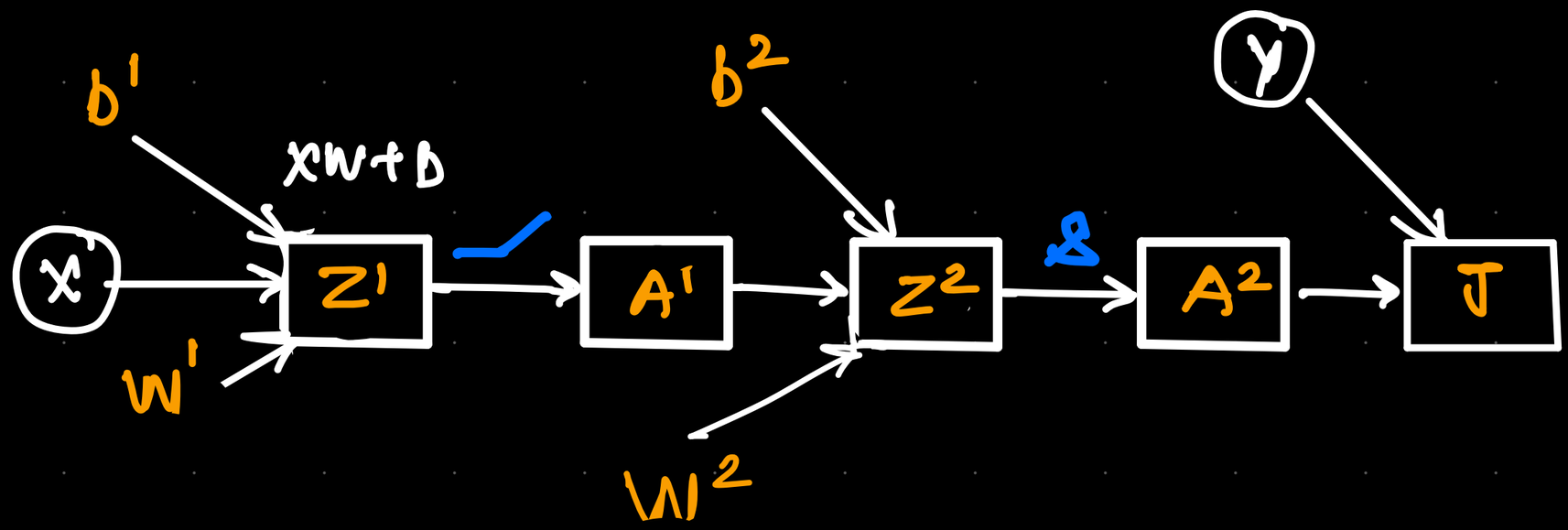
$X \longrightarrow (300, 2)$

$\partial W^1 \longrightarrow (2, 4)$

$$\boxed{\partial W^1 = X^T \cdot \partial Z^1}$$

Backward Prop(∂b^1)

$\frac{\partial Z^1}{\partial W^1}$ 3 $\quad$ $\frac{\partial A^1}{\partial Z^1}$ 2 $\quad$ $\frac{\partial Z^2}{\partial A^1}$ 1 $\quad$ dZ^2

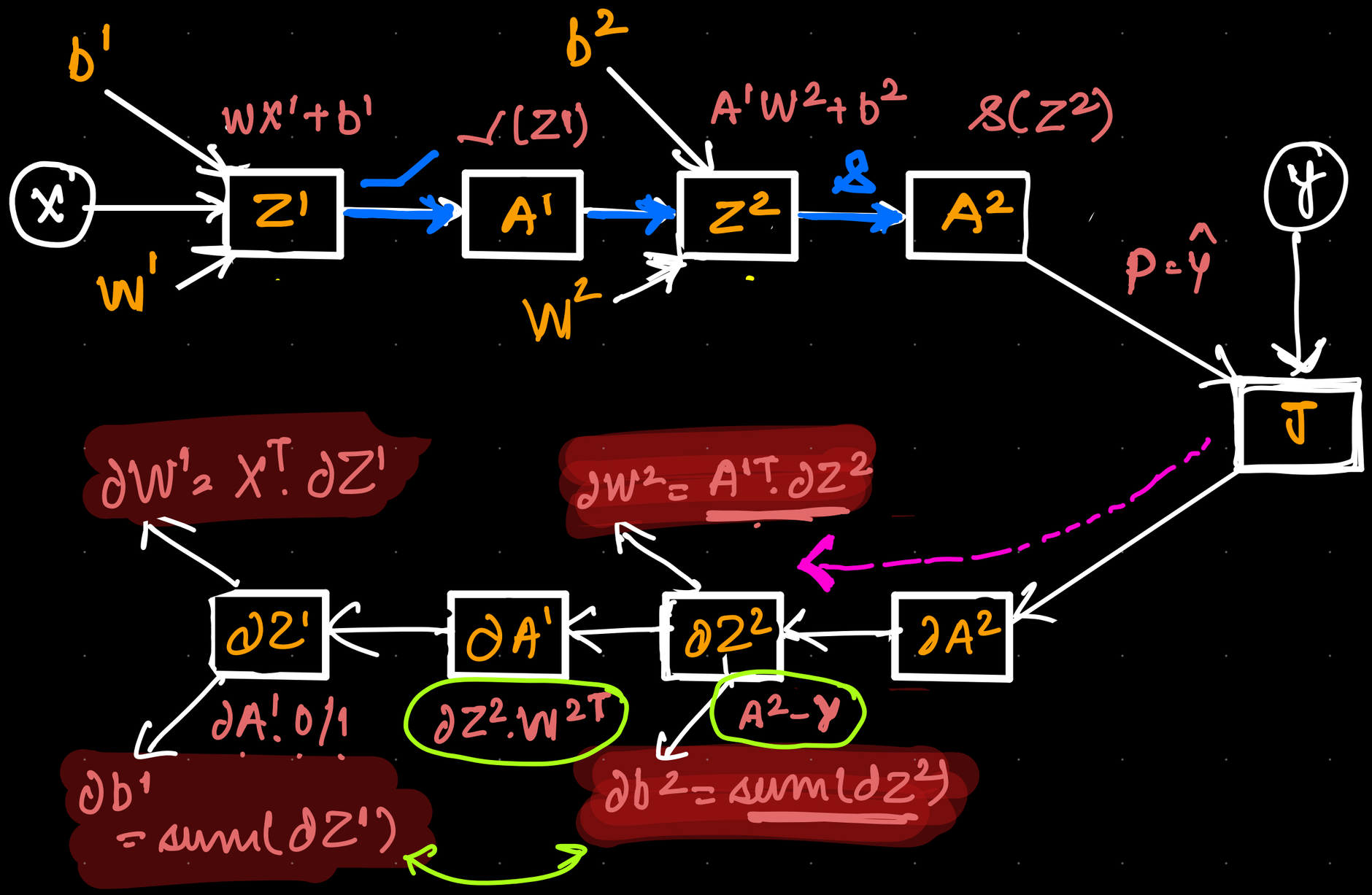


$$\frac{\partial Z^1}{\partial b^1} = \frac{\partial (XW^1 + b^1)}{\partial b^1} = 1$$

$$\partial W^1 = \text{inp.sum}(\partial Z^1)/m$$

Backward Prop (∂b^1)

→ Forward Propagation



Tomorrow
Tuesday → Keras + Tensorflow

Thursday →

Regularization
Hyper tuning
Optimization.
Error Analysis.