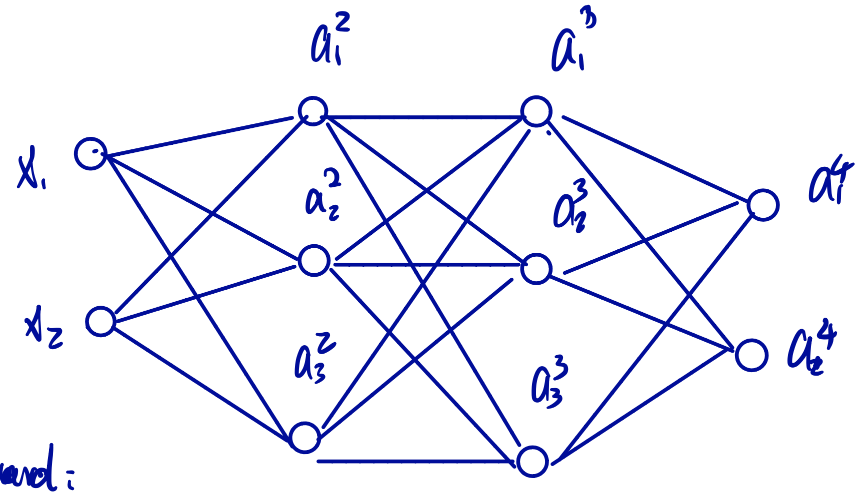
2.



forward:

$$\begin{pmatrix} z_1^2 \\ z_2^2 \\ z_3^2 \end{pmatrix} = \begin{pmatrix} w_{12}^2 & w_{12}^2 & b_1^2 \\ w_{21}^2 & w_{22}^2 & b_2^2 \\ w_{31}^2 & w_{32}^2 & b_3^2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix},$$ that is $z^2 = (w^2 \ b^2)\begin{pmatrix} x \\ 1 \end{pmatrix}$

$$a^2 = RELU(z^2) = \max(z^2, 0).$$

So, $z^2 = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}$, $a^2 = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}$.

Similarity, $z^3 = (w^3 \ b^3)\begin{pmatrix} a^2 \\ 1 \end{pmatrix} = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}$, $a^3 = z^3 = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}$

$z^4 = (w^4 \ b^4)\begin{pmatrix} a^3 \\ 1 \end{pmatrix} = \begin{pmatrix} 31 \\ 31 \end{pmatrix}$, $a^4 = z^4 = \begin{pmatrix} 31 \\ 31 \end{pmatrix}$

$$C = (y_1 - a_1^4)^2 + (y_2 - a_2^4)^2.$$

backpropagation:

$$\begin{pmatrix} \frac{\partial C}{\partial z_1^4} \\ \frac{\partial C}{\partial z_2^4} \end{pmatrix} = \begin{pmatrix} \frac{\partial C}{\partial a_1^4} \cdot \frac{\partial a_1^4}{\partial z_1^4} \\ \frac{\partial C}{\partial a_2^4} \cdot \frac{\partial a_2^4}{\partial z_2^4} \end{pmatrix} = \begin{pmatrix} \frac{\partial C}{\partial a_1^4} \\ \frac{\partial C}{\partial a_2^4} \end{pmatrix} \odot \begin{pmatrix} \frac{\partial a_1^4}{\partial z_1^4} \\ \frac{\partial a_2^4}{\partial z_2^4} \end{pmatrix},$$ that is $\delta^4 = \frac{\partial C}{\partial a^4} \odot \sigma'(z^4)$.

BP1

$$\delta^4 = \begin{pmatrix} 2(a_1^4 - y_1) \\ 2(a_2^4 - y_2) \end{pmatrix} \odot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 60 \\ 60 \end{pmatrix} \ldots$$

$$\begin{pmatrix} \frac{\partial C}{\partial z_1^3} \\ \frac{\partial C}{\partial z_2^3} \\ \frac{\partial C}{\partial z_3^3} \end{pmatrix} = \begin{pmatrix} \frac{\partial C}{\partial a_1^3} \cdot \frac{\partial a_1^3}{\partial z_1^3} \\ \frac{\partial C}{\partial a_2^3} \cdot \frac{\partial a_2^3}{\partial z_2^3} \\ \frac{\partial C}{\partial a_3^3} \cdot \frac{\partial a_3^3}{\partial z_3^3} \end{pmatrix} = \begin{pmatrix} \frac{\partial C}{\partial a_1^3} \\ \frac{\partial C}{\partial a_2^3} \\ \frac{\partial C}{\partial a_3^3} \end{pmatrix} \odot \sigma'(z^3)$$

$$= \begin{pmatrix} \frac{\partial C}{\partial z_1^4} \cdot \frac{\partial z_1^4}{\partial a_1^3} + \frac{\partial C}{\partial z_2^4} \cdot \frac{\partial z_2^4}{\partial a_1^3} \\ \frac{\partial C}{\partial z_1^4} \cdot \frac{\partial z_1^4}{\partial a_2^3} + \frac{\partial C}{\partial z_2^4} \cdot \frac{\partial z_2^4}{\partial a_3^3} \\ \frac{\partial C}{\partial z_1^4} \cdot \frac{\partial z_1^4}{\partial a_3^3} + \frac{\partial C}{\partial z_2^4} \cdot \frac{\partial z_2^4}{\partial a_3^3} \end{pmatrix} \odot \sigma'(z^3)$$

$$= ((W^4)^T \delta^4) \odot \sigma'(z^3) \qquad \text{(BP2)}$$

So, $\delta^3 = \begin{pmatrix} 120 \\ 120 \\ 120 \end{pmatrix}$.

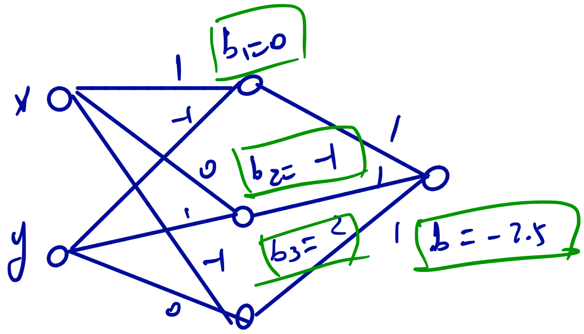
similarly, $\delta^2 = ((W^3)^T \delta^3) \odot \sigma'(z^2) = \begin{pmatrix} 360 \\ 360 \\ 360 \end{pmatrix}$

As $\underline{\frac{\partial C}{\partial b_j^l} = \delta_j^l}$ (BP3), $\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l$ (BP4)

$\frac{\partial C}{\partial b^4} = \begin{pmatrix} 60 \\ 60 \end{pmatrix}$ $\frac{\partial C}{\partial b^3} = \begin{pmatrix} 120 \\ 120 \\ 120 \end{pmatrix}$ $\frac{\partial C}{\partial b^2} = \begin{pmatrix} 360 \\ 360 \\ 360 \end{pmatrix}$.

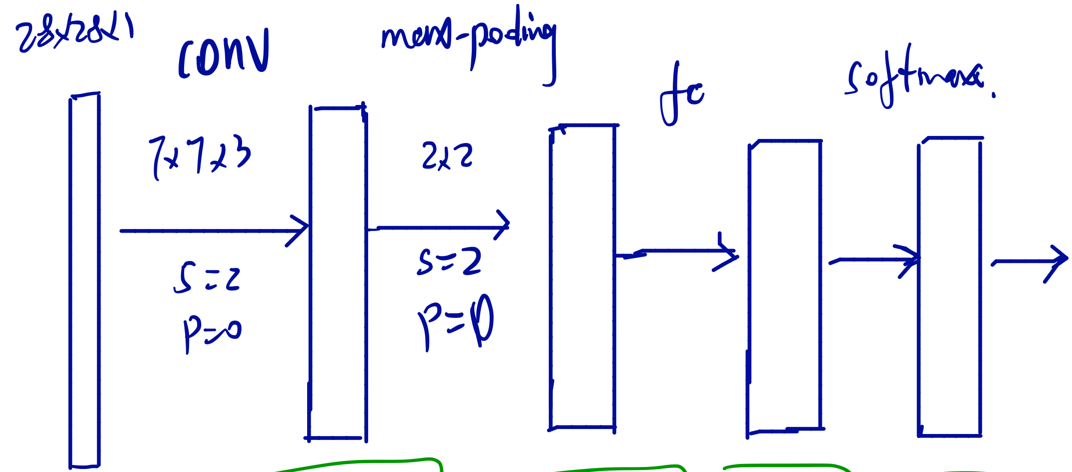
$\frac{\partial C}{\partial W^4} = \begin{pmatrix} 600 & 600 & 600 \\ 600 & 600 & 600 \end{pmatrix}$, $\frac{\partial C}{\partial W^3} = \begin{pmatrix} 360 & 360 & 360 \\ 360 & 360 & 360 \\ 360 & 360 & 360 \end{pmatrix}$, $\frac{\partial C}{\partial W^2} = \begin{pmatrix} 360 & 360 \\ 360 & 360 \\ 360 & 360 \end{pmatrix}$

3.



$$\begin{cases} x - y \geq 0 \\ y \geq 1 \\ x \leq 2 \end{cases} \Rightarrow \begin{cases} x - y \geq 0 \\ y - 1 \geq 0 \\ -x + 2 \geq 0 \end{cases}$$

4.



(a) dimensions: $\boxed{11 \times 11 \times 3}$ $\boxed{5 \times 5 \times 3}$ $\boxed{26 \times 1}$ $\boxed{26 \times 1}$

$N @ ((D - k + 2p)/s + 1)$ $D = 28$, $k = 7$, $p = 0$, $s = 2$

(b) $(7 \times 7 + 1) \times 3 + (5 \times 5 \times 3 + 1) \times 26 = 2126$

(c).

x_1	x_2	x_3	x_4	x_5	
0	0	0	0	0	a
0	0	0	0	1	b
		⋮			⋮
1	1	0	0	1	z

No need to use soft-max.

The ground truth has multiple 1s or no 1. But Soft-Max will force the network to give out a single 1.