$a_i^{(l)}$ — (l) — layer, i — unit   activation

$$a_1^{(2)} = g(z_1^2) \qquad z_1^2 = \Theta_{10}^{(1)} x_0 + \Theta_{11}^1 x_1 + \Theta_{12}^1 x_2 + \Theta_{13}^1 x_3$$

g — sig $\Theta^T \cdot x$

$$= \frac{1}{1 + e^{-(\Theta_{10}^1 x_0 + \Theta_{11}^1 x_1 + \cdots)}}$$

$$a_2^{(2)} = g(z_2^2) = \frac{1}{1 + e^{-(\Theta_{20}^{(1)} x_0 + \Theta_{21}^{(1)} x_1 + \Theta_{22}^{(1)} x_2 + \Theta_{23}^{(1)} x_3)}}$$

$\downarrow$ $z_2^{(2)}$

$$a_3^{(2)} = g(z_3^2) = \frac{1}{1 + e^{-(\Theta_{30}^1 x_0 + \Theta_{31}^1 x_1 +}}$$

$$\Theta^{(1)} : [3 \times 4]$$

z^2 $\Theta^1 x$ $[3 \times 1] \times [4 \times 1]$

$\underline{z} = \underline{\underline{\Theta}}\ \underline{x} = [3 \times 4] \times [4 \times 1]$

$= [3 \times 1]$

$a^2 = g(z^2) = [3 \times 1]$

$z^3 = \Theta^{(2)}\ \underline{a}^2$

$\downarrow$

$[1 \times 4]\ [4 \times 1] = [1]$

$z^3 = \underbrace{\Theta^{(2)}_{10} a^2_0 + \Theta^{(2)}_{11} a^2_1 + \Theta^{(2)}_{12} a^2_2 + \Theta^{(2)}_{13} a^2_3}$

$a^3 = g(\underline{z}^3)$

$= \frac{1}{}$

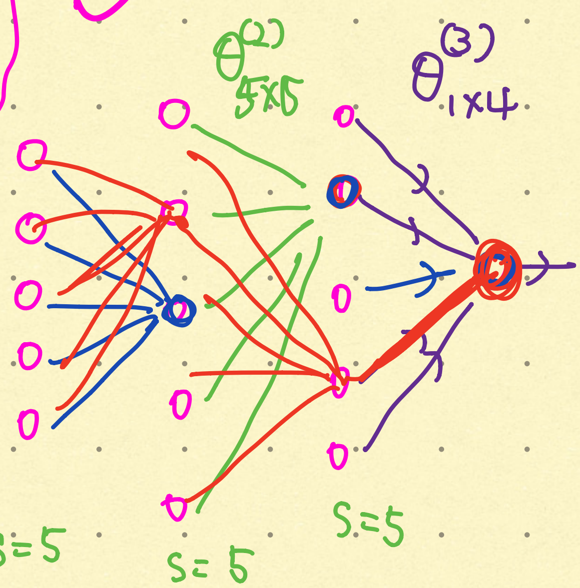
$$a_1^3 \qquad \frac{\quad}{1 + e^{-(\theta_{10}^2 a_0^2 + \theta_{11}^2 a_1^{(2)} + \cdots)}}$$

$$a_1^{(2)} = \frac{1}{1 + e^{-(\theta_{10}^{(1)} x_0 + \theta_{11}^{(1)} x_1 + \theta_{12}^{(1)} x_2)}}$$

$$z^{(l)} = \sum_k \theta_{jk}^{(l-1)} a_k^{(l-1)} + \theta_{j0}^{(l-1)}$$

$$= \underline{\theta}^{(l-1)} \underline{a}^{(l-1)}$$

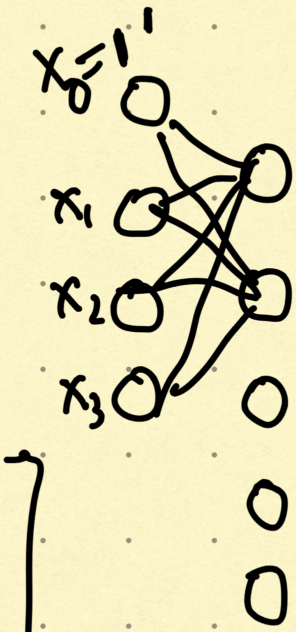
$$a_j^{(l)} = g\left(\sum_k \theta_{jk}^{(l-1)} a_k^{(l-1)} + \theta_{j0}^{(l-1)}\right)$$



In a vector form

$$z^2 = \Theta^{(1)} a^1$$

$5\times1 \quad 5\times4 \quad 4\times1$



$\underline{\underline{5\times4}}$

$$\underline{\underline{\Theta^{(1)}}}_{[5\times4]} \quad \begin{bmatrix} \theta^1_{10} & \theta^1_{11} & \theta^1_{12} & \theta^1_{13} \\ \vdots & & & \\ \vdots & & & \\ \theta^1_{50} & \theta^1_{51} & \theta^1_{52} & \theta^1_{53} \end{bmatrix}$$

$$a^2 = g(z^2) \quad 5\times1 \rightarrow (a^2_0 \; a^2_1 \; a^2_2 \cdots a^2_5)$$

$a^2_0 = 1$, $\quad a^2_1 = \dfrac{1}{1+e^{-z^2_1}}$, $\quad a^2_2 = \dfrac{1}{1+e^{-z^2_2}}$, $\quad a^2_5 = \dfrac{1}{1+e^{-z^2_5}}$

$$z^3 = \Theta^{(2)} \underline{a^2}$$

$5\times1 \quad 5\times6 \quad 6\times1$

$$\begin{bmatrix} \theta^2_{10} & \theta^2_{11} \cdots & \theta^2_{15} \end{bmatrix}$$

$$\underline{\underline{\Theta^{(2)}}}_{[5\times6]} \quad \begin{bmatrix} \theta^2_{20} & \theta^2_{21} & \cdots & \theta^2_{25} \\ \vdots & & & \\ \theta^2_{50} & \theta^2_{51} & \cdots & \theta^2_{55} \end{bmatrix} \quad 5\times6$$

$$\Theta^{(3)}_{[4\times6]}$$

$$a^4 = (a^4_1 \quad a^4_2 \quad a^4_3 \quad a^4_4)$$

$$y^{(1)} = (y_1 \quad y_2 \quad y_3 \quad y_4)$$

$$\underset{\sim}{\delta^4_j} = a^4_j - y_j \qquad j = 1, 2, 3, 4$$

$$\underset{5\times1}{\tilde{\delta}^3} = \underline{\left(\Theta^{(3)}\right)^T \underset{4\times1}{\delta^4}} \qquad 5\times4 \cdot 4\times1 \qquad 5\times1$$

$$\boxed{4\times5} \longrightarrow \underline{5\times4}$$

$$(a^3_1 \quad a^3_2 \quad a^3_3 \quad a^3_4 \quad a^3_5)^T$$

$$\underset{5\times 1}{\delta^3} = \overbrace{\underset{5\times 4}{(\Theta^3)^T} \cdot \underset{4\times 1}{\delta^4}}^{\delta^3} .* \, g'(z^3)$$

$$\boxed{\underset{5\times 1}{\delta^3}}$$

$$y = f(x)$$

m

$$\Delta y = \frac{\partial f}{\partial x} \cdot \Delta x$$

$$\underset{5\times 1}{\delta^3} = \underset{5\times 1}{(\Theta^3)^T \delta^4} .* \underset{5\times 1}{g'(z^3)}$$

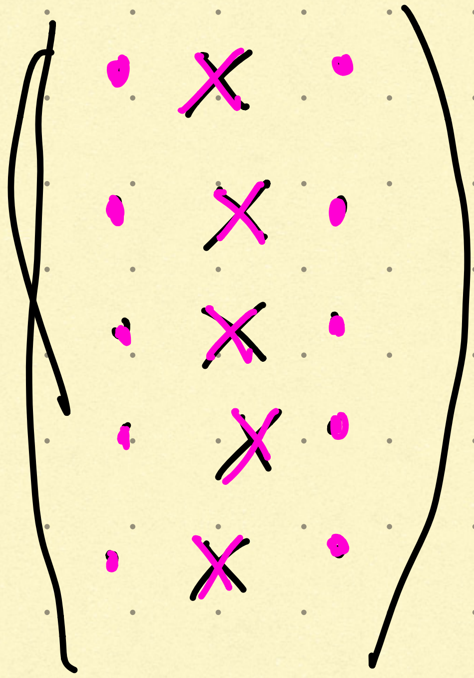
$$\begin{pmatrix} \vdots \end{pmatrix} .* \begin{pmatrix} \vdots \end{pmatrix}$$

Hadamard product

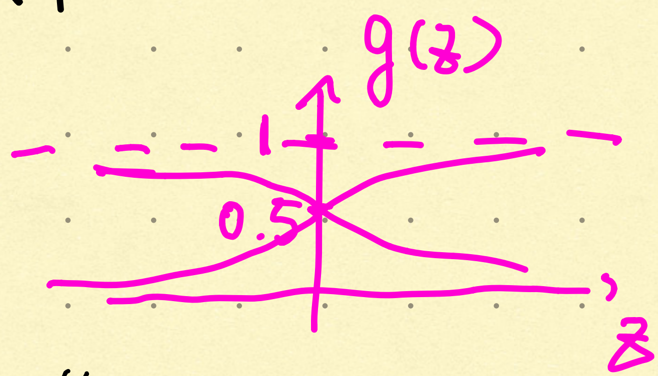
$.*$

$\odot$

$\begin{bmatrix}1\\2\end{bmatrix} \odot \begin{bmatrix}3\\4\end{bmatrix} = \begin{bmatrix}1\times 3\\2\times 4\end{bmatrix} = \begin{bmatrix}3\\8\end{bmatrix}$

$=$ 

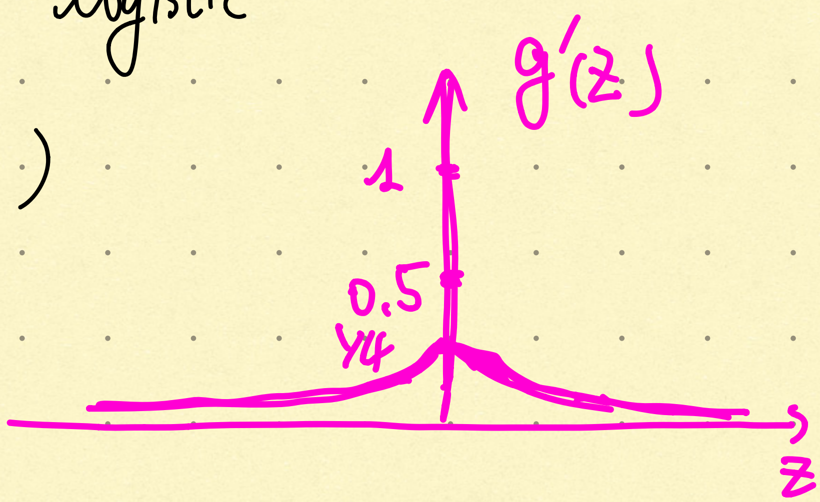
$$\delta^2 = \underbrace{(\Theta^2)^T \cdot \delta^3}_{5\times 1} .* \underline{g'(z^2)}$$



$g(z) = \frac{1}{1+e^{-\theta x}}$

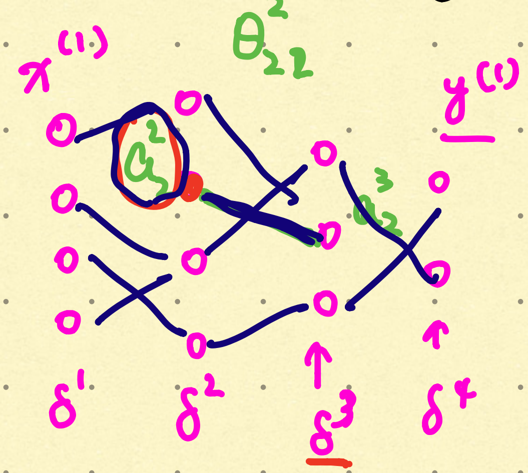
$g'(z) = g(z)g(-z)$ logistic

$= g(z)(1-g(z))$



once saturated,
learning slowly

$$z_1^{(3)} = \theta_{10}^2 \cdot 1 + \theta_{11}^2 a_1^2 + \theta_{12}^2 a_2^2$$



$$\delta^4 = y - a_1^4$$

$$\tilde{\delta}_2^2 = \theta_{12}^2 \delta_1^3 + \theta_{22}^2 \delta_2^3$$

$(\delta_1^3, \delta_2^3, \delta_3^3)$

$$\delta_2^2 = \tilde{\delta}_2^2 \cdot g'(z_2^2)$$

$$\underbrace{\frac{\partial J}{\partial \theta_{10}^2}}_{i\,j} \ell = \underbrace{a_0^2}_{\overset{\shortparallel}{1}} \cdot \underline{\delta_1^3}$$

$$\frac{\partial J(\Theta)}{\partial \Theta_{ij}^{\ell}} = a_j^{\ell} \delta_i^{\ell+1}$$

$$\frac{\partial J}{\partial \Theta^{2 \sim l}_{12}} = a^2_2 \cdot \delta^3_1$$

$$\underset{i}{1}\ \underset{j}{2}$$

$$\Theta' = \Theta - \alpha \frac{\partial J}{\partial \Theta}$$

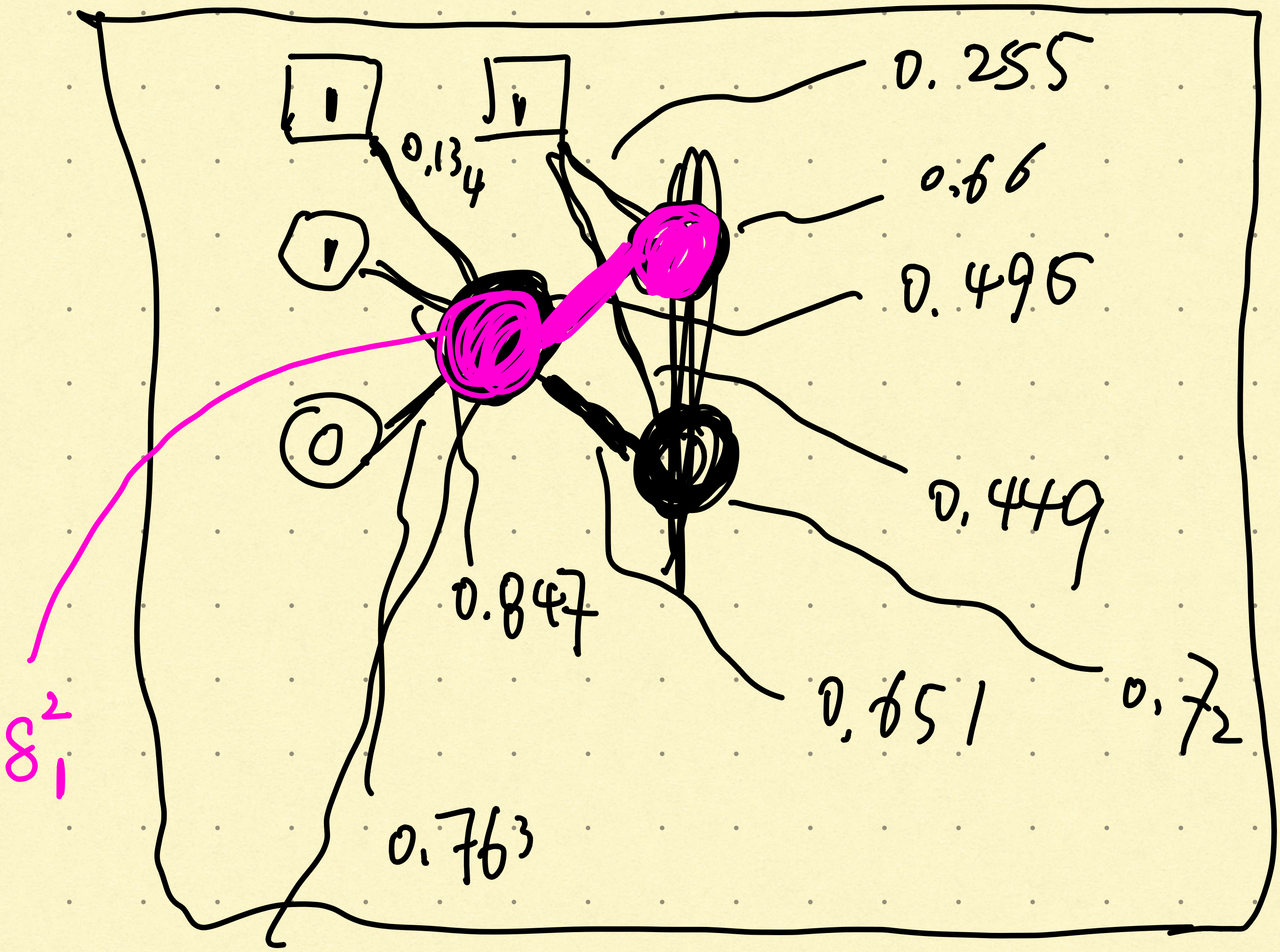
$$\frac{\partial J}{\partial \Theta^2_{..}} = \underbrace{\frac{\partial J}{\partial z^3_{.}}}_{\delta^3_{.}} \underbrace{\frac{\partial z^3_{.}}{\partial \Theta^2_{..}}}_{a^2_{.}}$$

$$z^3 = \Theta^2 \cdot a^2 \longrightarrow$$

$$= \begin{bmatrix} \Theta^2_{10} & \Theta^2_{11} & \cdots & \Theta^2_{15} \\ & 5 \times 6 & & \\ \Theta^2_{50} & \Theta^2_{51} & \cdots & \Theta^2_{55} \end{bmatrix} \cdot \begin{bmatrix} a^2_0 = 1 \\ a^2_1 \\ a^2 \\ \vdots \\ a^2_5 \end{bmatrix}$$

$$\delta^2 = (\Theta^2)^T \delta^3 .* g'(z^2)$$

Take an explicit example :



$$\boxed{z = \underline{\underline{\theta}} \cdot \underline{x}}$$

$$\boxed{a = g(z) = \frac{1}{1+e^{-z}}}$$

$$a_1^{(2)} = \frac{1}{1 + e^{-(\underbrace{\theta^1_{10}}_{0.1343} + \underbrace{\theta^1_{11}}_{0.847} \underbrace{x_1}_{1} + \underbrace{\theta^1_{12}}_{0.76} \underbrace{x_2}_{0})}}$$

$$= \frac{1}{1 + e^{-(0.1343 + 0.847)}} \quad 0.72736$$

$$a_1^{(3)} = \frac{1}{1 + e^{-(\theta^2_{10} + \theta^2_{11} a^2_1)}}$$

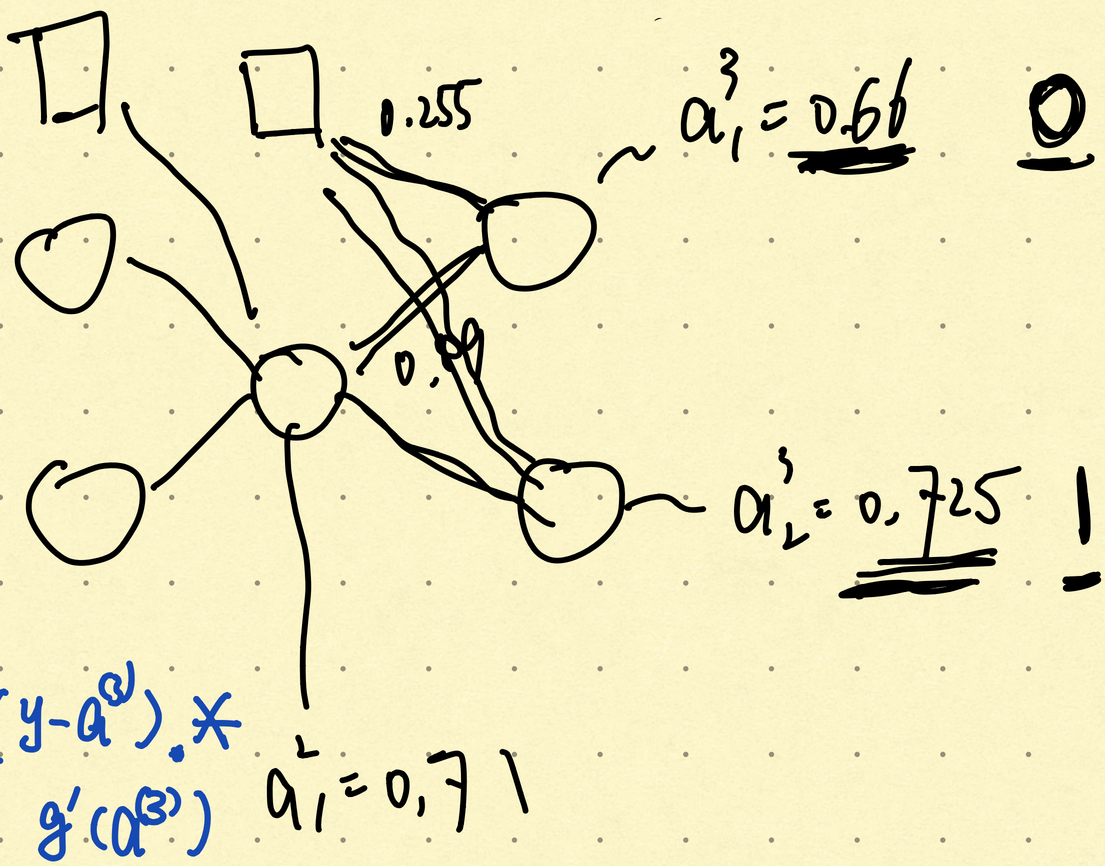
$$= \frac{1}{1 + e^{-(0.255 + 0.496)}}$$

$$= 0.66 \ldots$$

$$a_2^{(3)} = \frac{1}{1 + e^{-(\theta_{20}^2 + \theta_{21}^2 a_1^2)}}$$

$$= \frac{1}{1 + e^{-(0.449 + 0.6516 \cdot 0.72736)}}$$

$$= 0.72\ldots$$



$\delta^{(3)} = (y - a^{(3)}) .* g'(a^{(3)})$

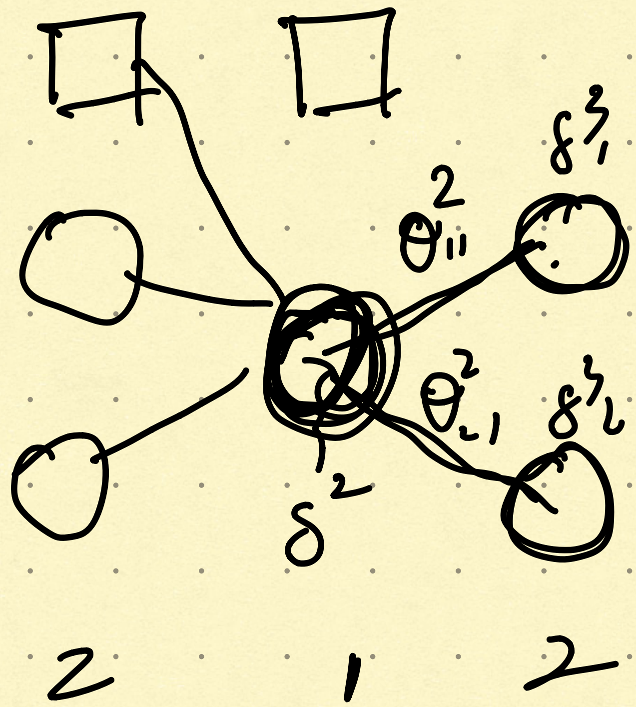
$$\tilde{\delta}^{(3)} \begin{pmatrix} \delta_1^{(3)} \\ \\ \delta_2^{(3)} \end{pmatrix} \begin{matrix} = -0.66 \cdot 0.66 \cdot (1 - 0.66) = -0.148 \\ \\ = 0.235 \cdot 0.725 \cdot 0.235 = 0.0547 \end{matrix}$$

$$\delta^2 = (\Theta^2)^T \cdot \delta^3$$
$$\quad\quad\quad\quad 2\times 2 \quad\quad 2 1$$

$$= \begin{pmatrix} \Theta^2_{10} & \Theta^2_{20} \\ \Theta^2_{11} & \Theta^2_{21} \end{pmatrix} \begin{pmatrix} \delta^3_1 \\ \delta^3_2 \end{pmatrix} = \begin{pmatrix} \Theta^2_{10}\delta^3_1 + \Theta^2_{20}\delta^3_2 \\ \Theta^2_{11}\delta^3_1 + \Theta^2_{21}\delta^3_2 \end{pmatrix}$$

0,44 0,65

$$\Theta^2 = \begin{pmatrix} \Theta^2_{10} & \Theta^2_{11} \\ \Theta^2_{20} & \Theta^2_{21} \end{pmatrix}$$



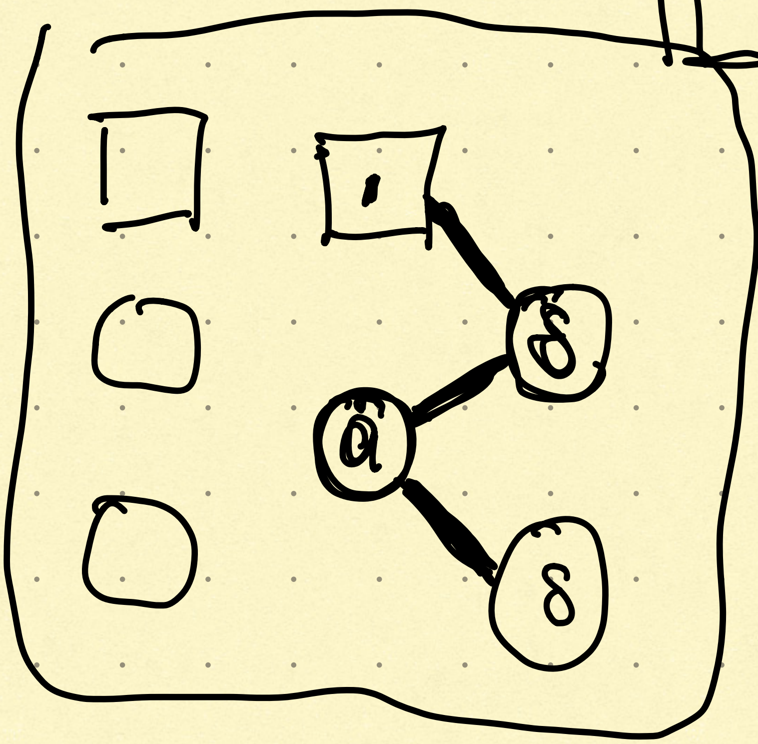
$$\delta^2 = (\Theta^2)^T \cdot \delta^3 .* \underline{g'}(z^2)$$

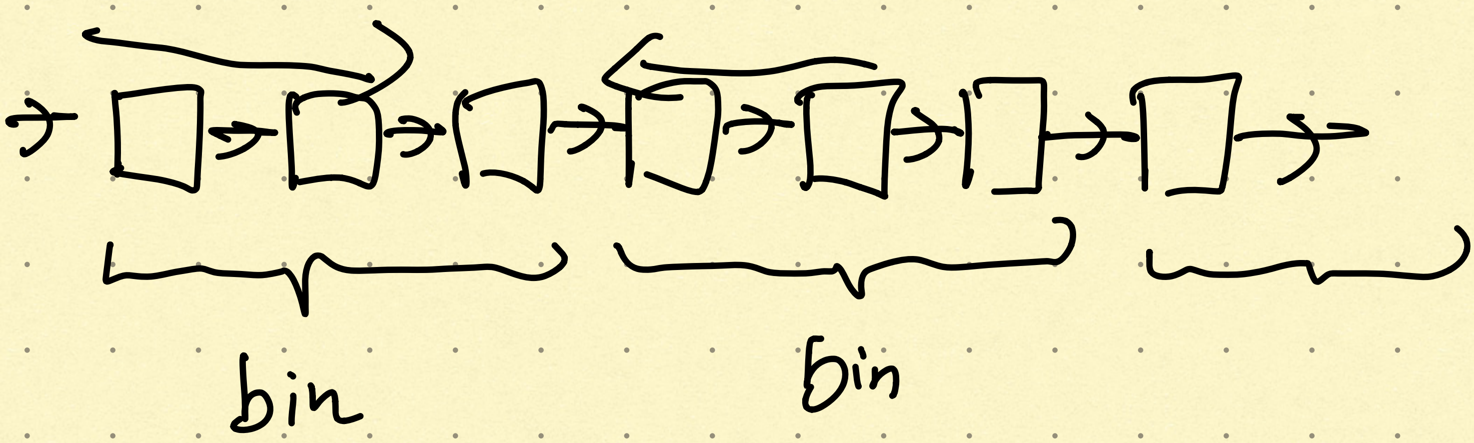
$$\theta^{new} = \theta^{old} - \alpha \cdot \frac{\partial J}{\partial \theta}$$

$$\boxed{a_j^{(l)} \cdot \delta_i^{(l+1)}}$$

$y - a$

$a - y$





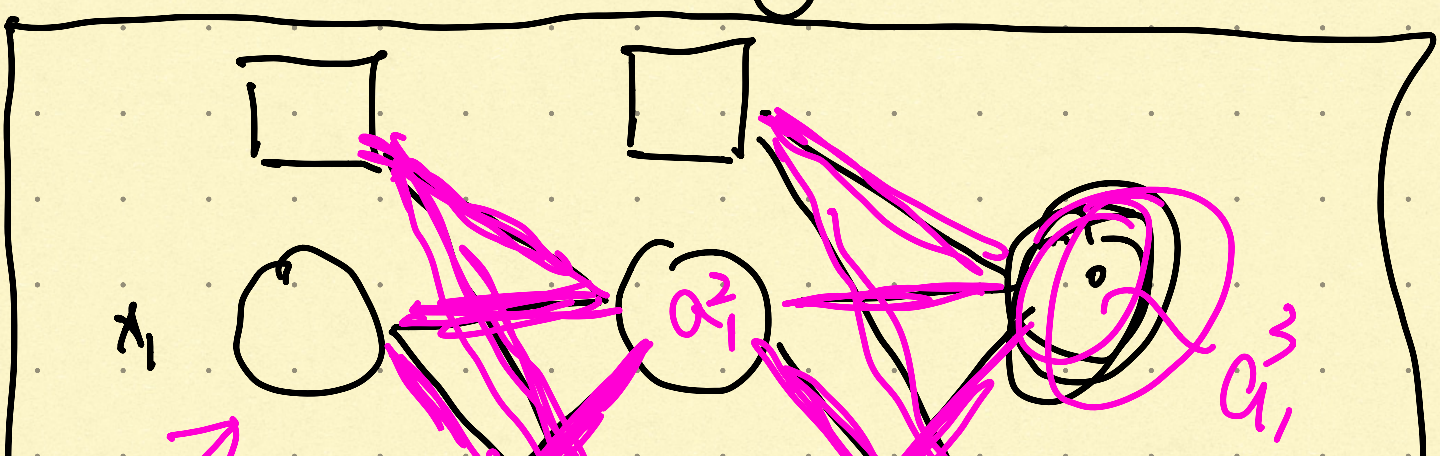
epoch epoch epoch

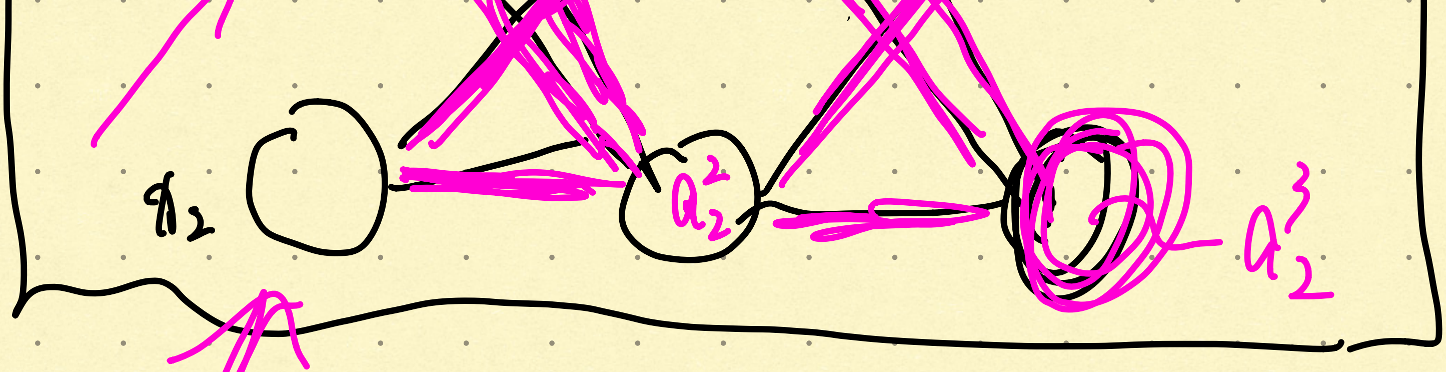
$$\{(\underline{x^1}, y^1), (x^2, y^2), \ldots, (x^m, y^m)\}$$

$$\downarrow$$

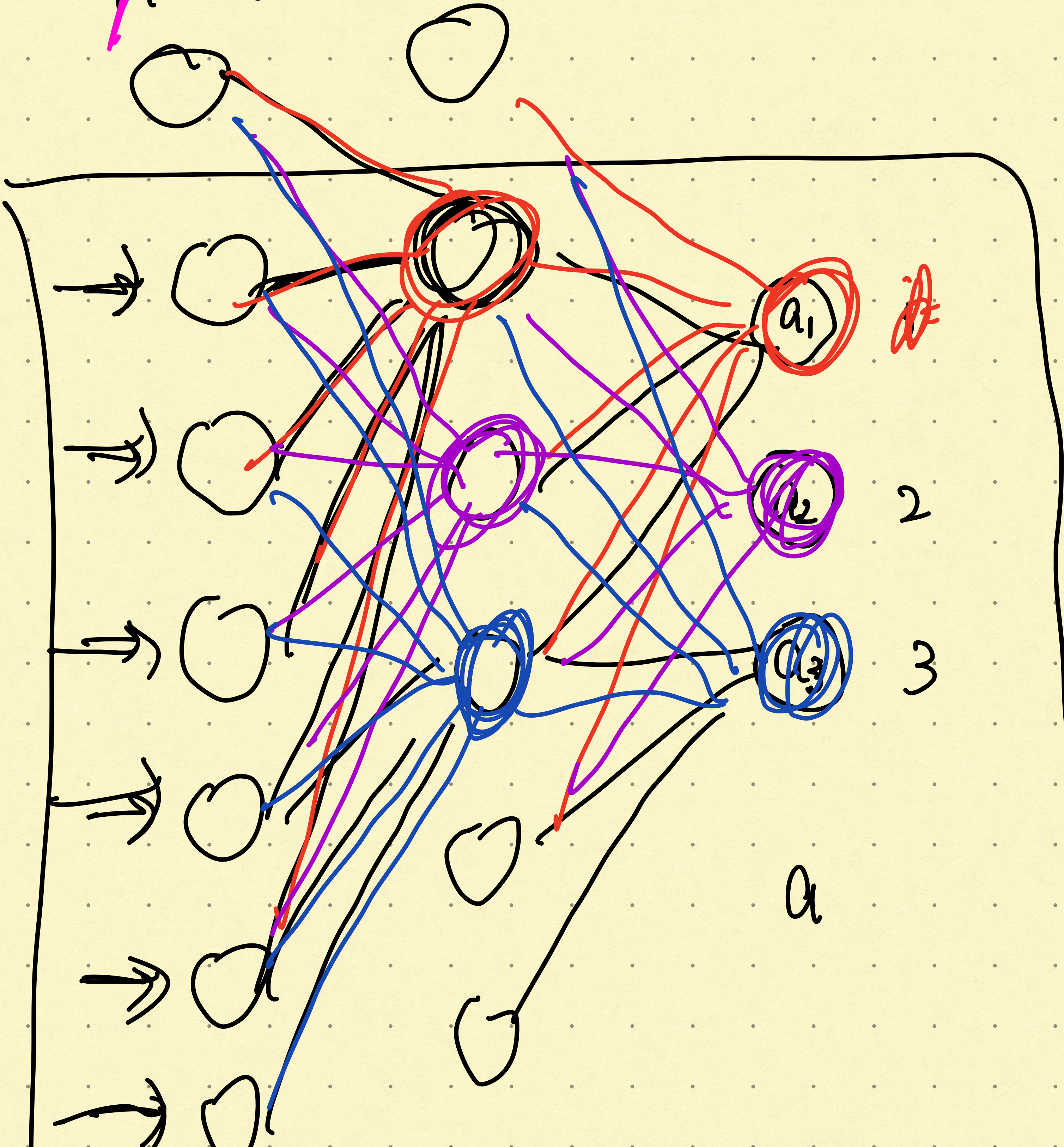
$$\begin{array}{r|l}
 & 1, 2, 3 \\
\hline
1 = & [1, 0, 0] \\
2 = & [0, 1, 0] \\
3 = & [0, 0, 1]
\end{array}$$

$$a = \frac{1}{1 + e^{-z}} \in (0, 1)$$





m = 10



0

$\rightarrow$ 0

3

7 5

MNIST

$\underline{28 \times 28 = \boxed{784}}$ 60,000

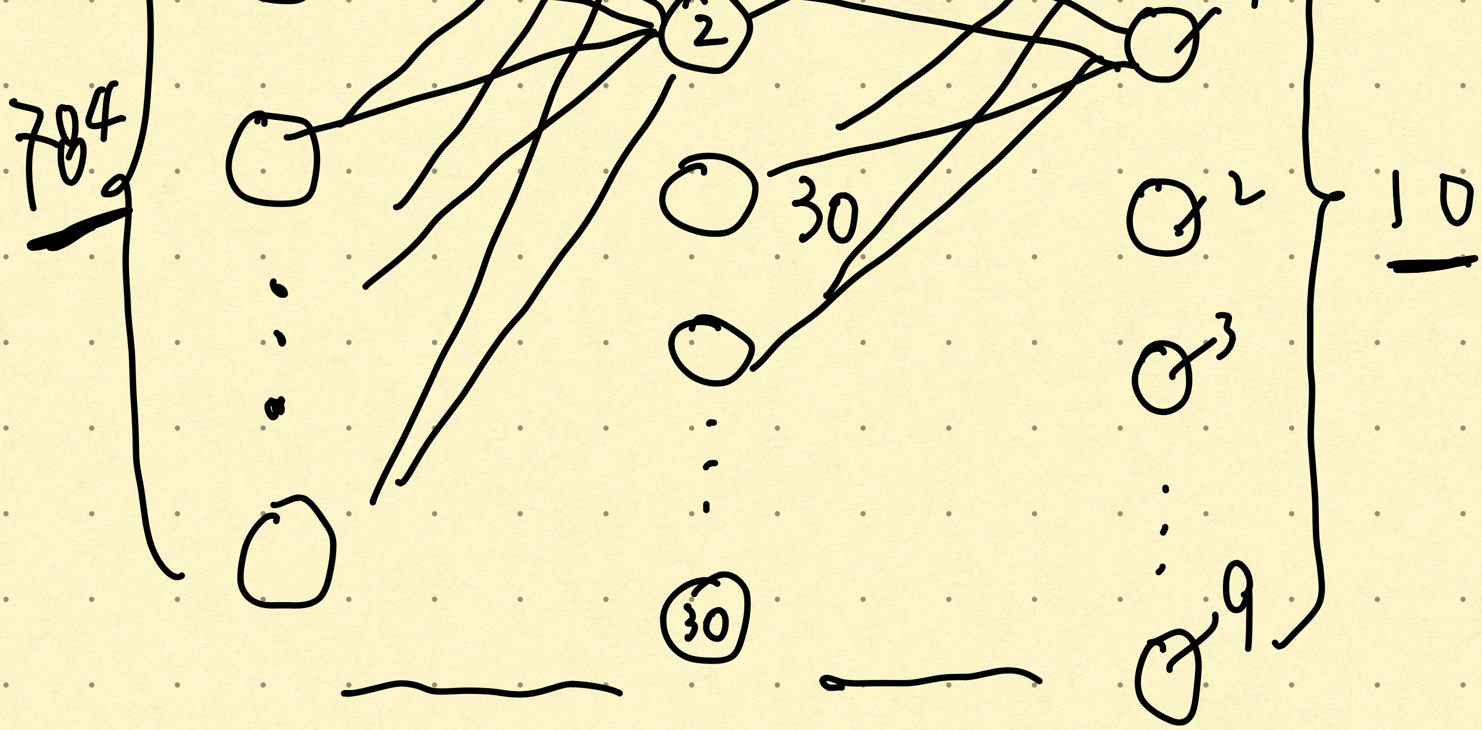
$[\underline{0}, 1, 2, \cdots, \underline{255}]$

$\underline{[0, 1, 2, \cdots, 9]}$

$\underline{784}$ 10 60,000

$$\underline{a = \frac{1}{1+e^{-z}}}$$

$x_1 \in [0, \cdots, 255]$

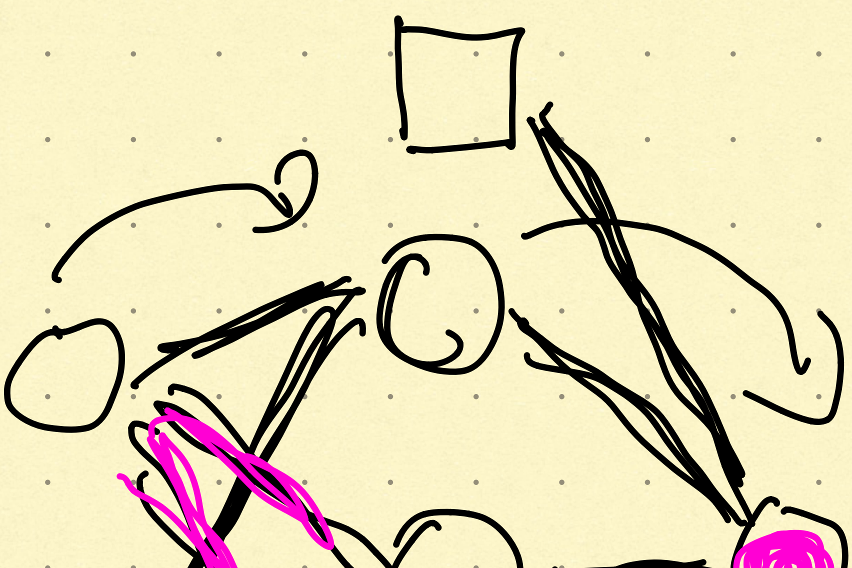


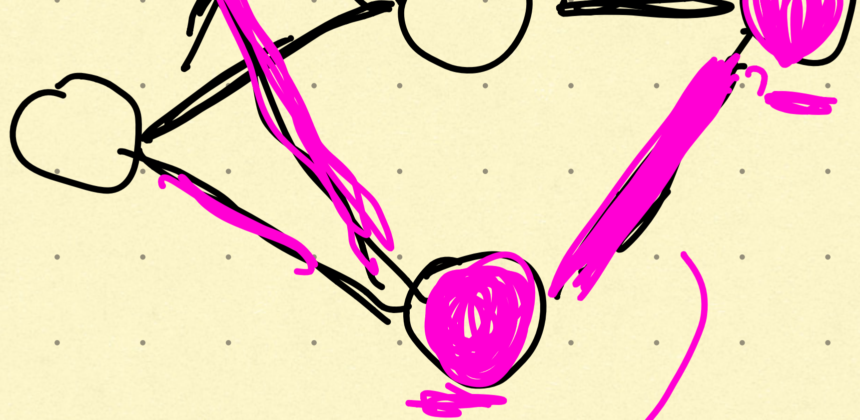
input

output

$$\boxed{\Theta^{(1)}_{30\times 784+1}} \qquad \boxed{\Theta^{(2)}_{10\times 31}}$$

$$23520 + 310 = \boxed{23830}$$



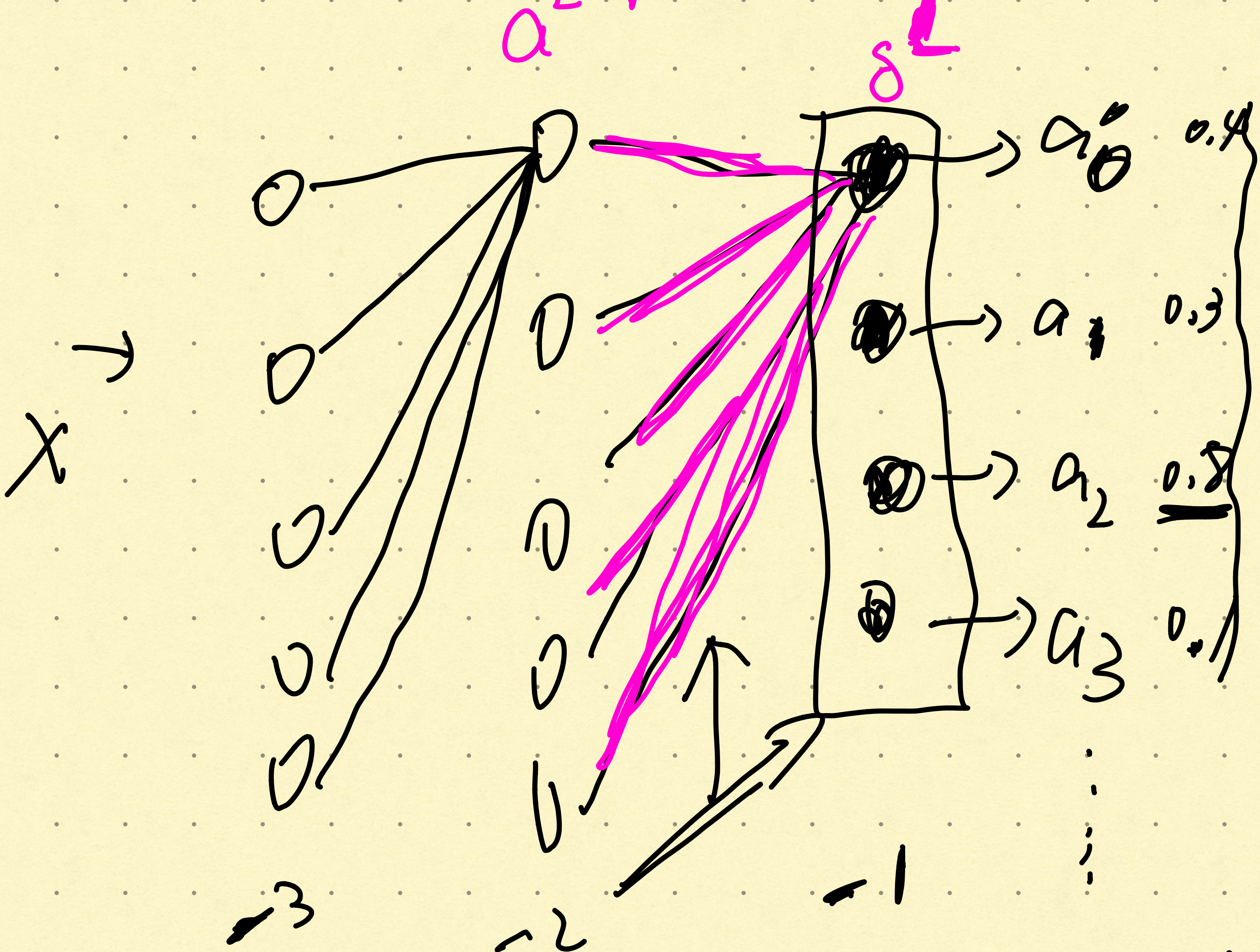


$\Theta^{(1)}$ $\Theta^{(2)}_{1\times 4} = \begin{bmatrix} \boxed{\Theta^{(2)}_{1,0}} \\ \Theta^{(2)}_{1,1} \\ \Theta^{(2)}_{1,2} \\ \Theta^{(2)}_{1,3} \end{bmatrix}$

$$\frac{\partial J}{\partial \Theta_{..}} = \underline{a} \cdot \underline{\delta}$$

$$\delta^L = a^L - y$$

$$\delta = (a^L - y) \cdot g'(z)$$



$$y \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

$\sigma'(z)$

$g'(z)$

$$\tilde{\delta}^{(l)} = \Theta^T \cdot \delta^{(l+1)}$$

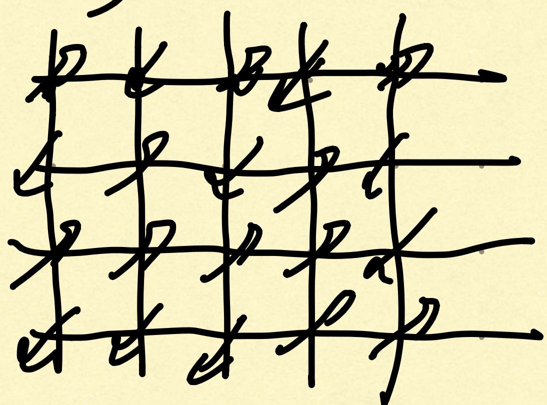
$$g(z) = \frac{1}{1+e^{-z}}$$

2D Ising

8×8, 10×10, 16×16, 32×32

128×128

L

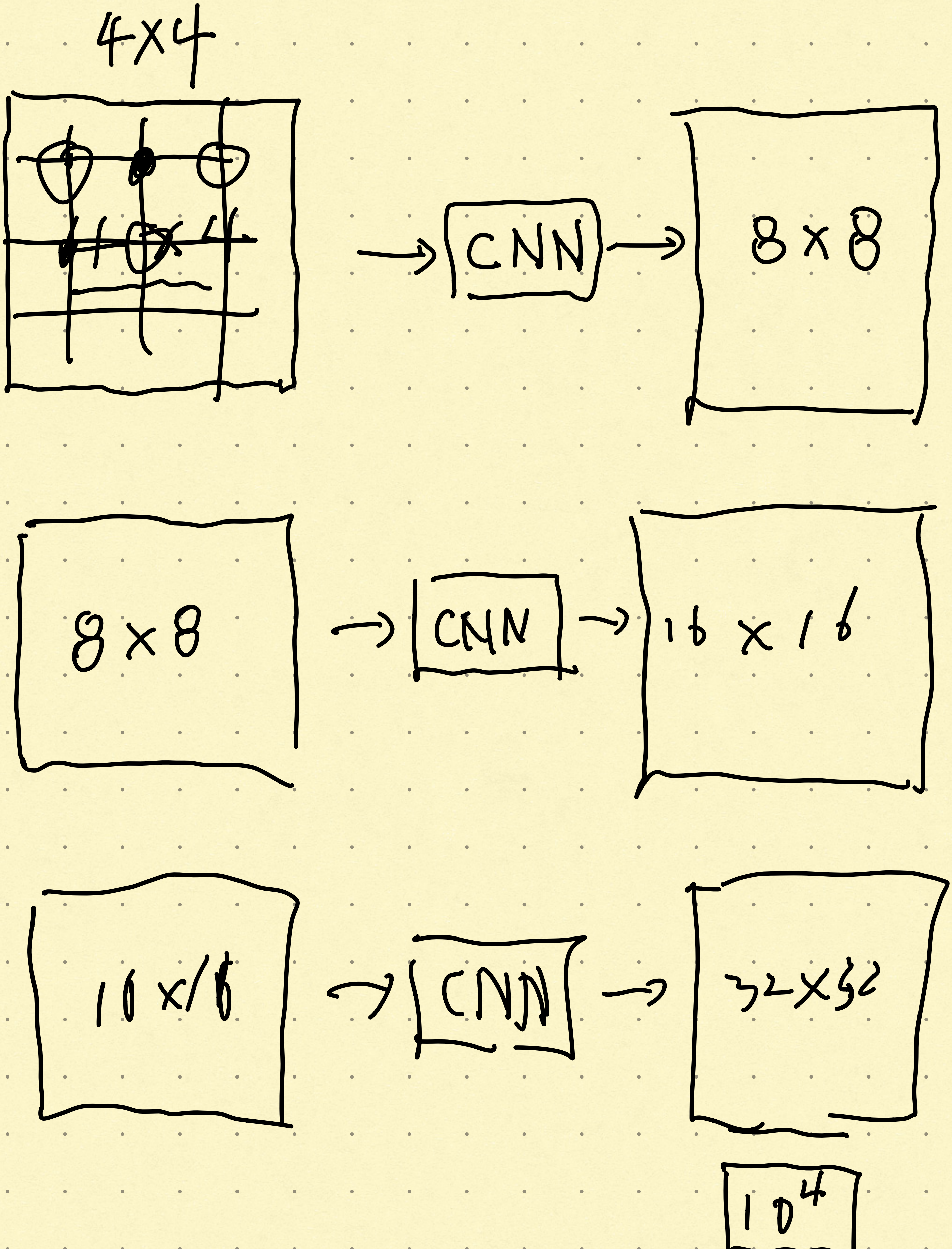


$$Z = \sum_{\{C\}} \underbrace{e^{-\frac{1}{T}E(C)}}\quad C$$

$E(C)$

$$e^{-\frac{1}{T}\cdot E(C)}$$

$L \times L \propto L^2$

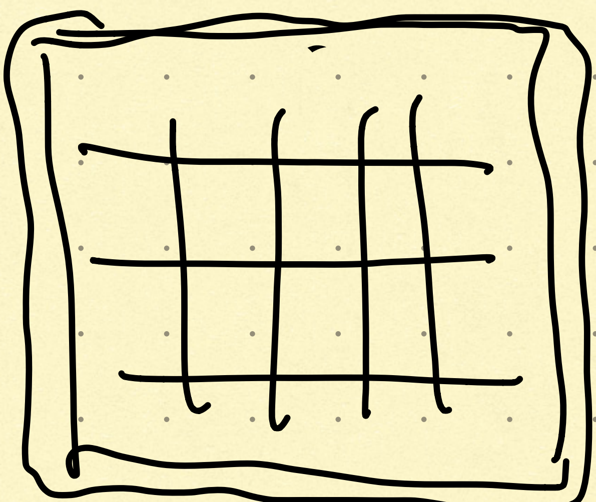


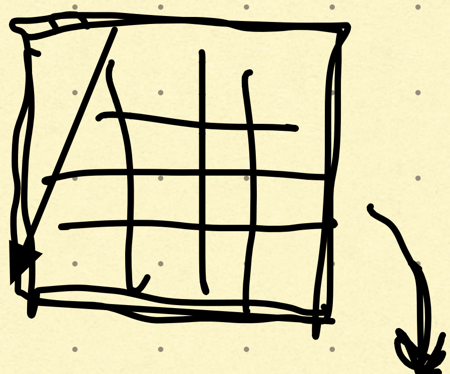
m, E

c, χ

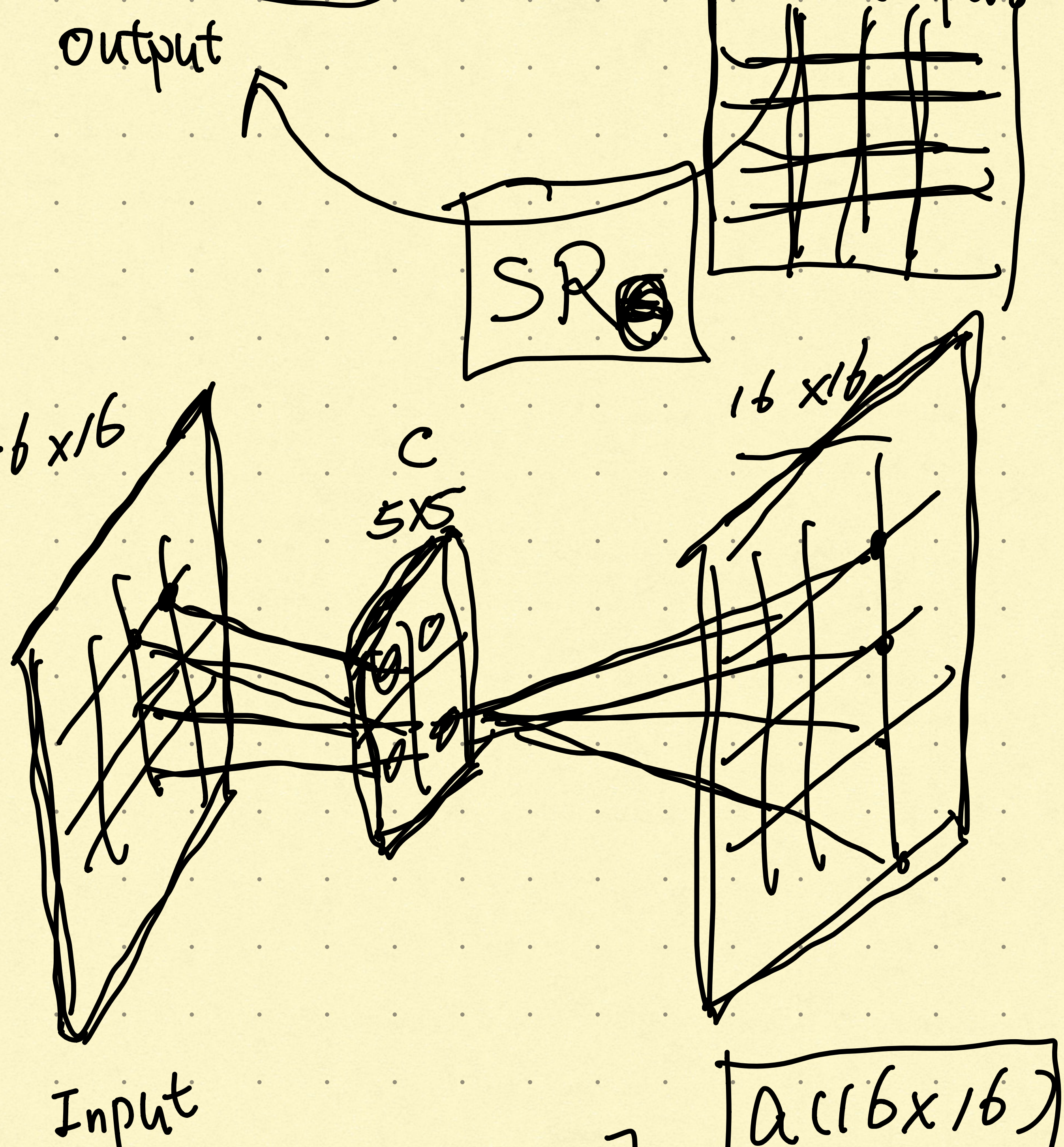
T

MC $16 \times 16 \xrightarrow{MR} 8 \times 8$





Input



$\boxed{a(16\times 16)}$

$a \in (0,1]$

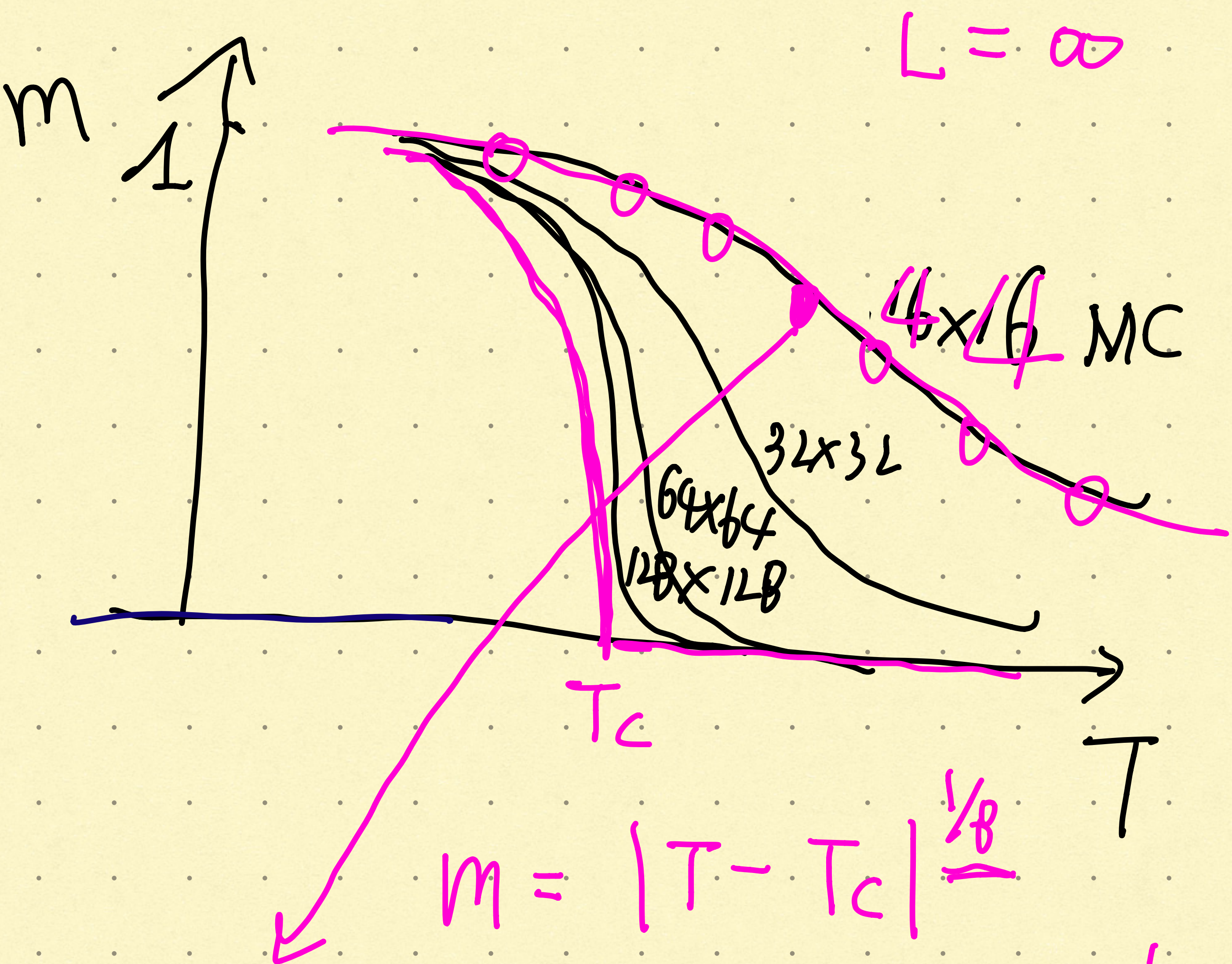
$\downarrow$

$a \in [-1,1]$

$$\frac{1}{1+e^{-\theta\cdot x}}$$

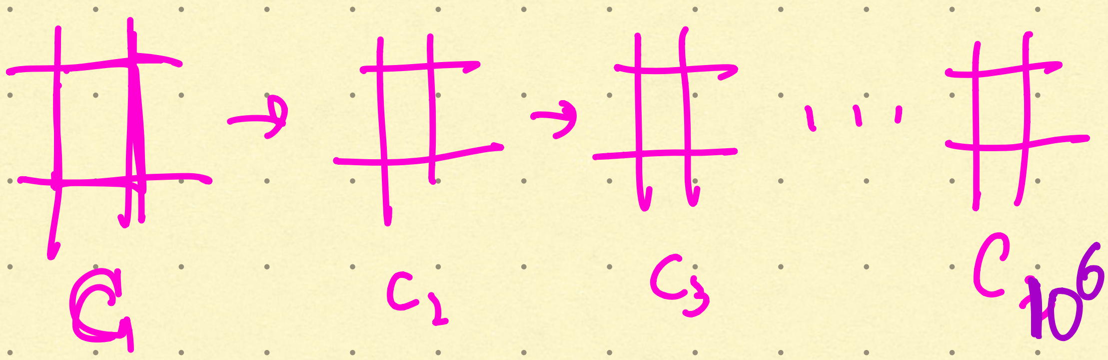
$$2\cdot\left(a-\frac{1}{2}\right)$$

$$(a_i - y_i)^2$$



$$m = |T - T_c|^{1/8}$$

$$\beta = \frac{1}{8} = 0{,}125$$



$$E = -\sum_{\langle ij \rangle} s_i^z s_j^z$$

$$C = \langle E^2 \rangle - \langle E \rangle^2$$

$$\chi = \langle M^2 \rangle - \langle |M| \rangle^2$$

$$E(C_1) \quad E(C_2) \quad E(C_3)$$

$$\underbrace{\langle E \rangle}_{\#}^{2} = \frac{1}{3} \sum_{i=1}^{3} E(C_i)$$

$$\underbrace{\langle E^2 \rangle} = \frac{1}{3} \sum_{i=1}^{3} E^2(C_i)$$