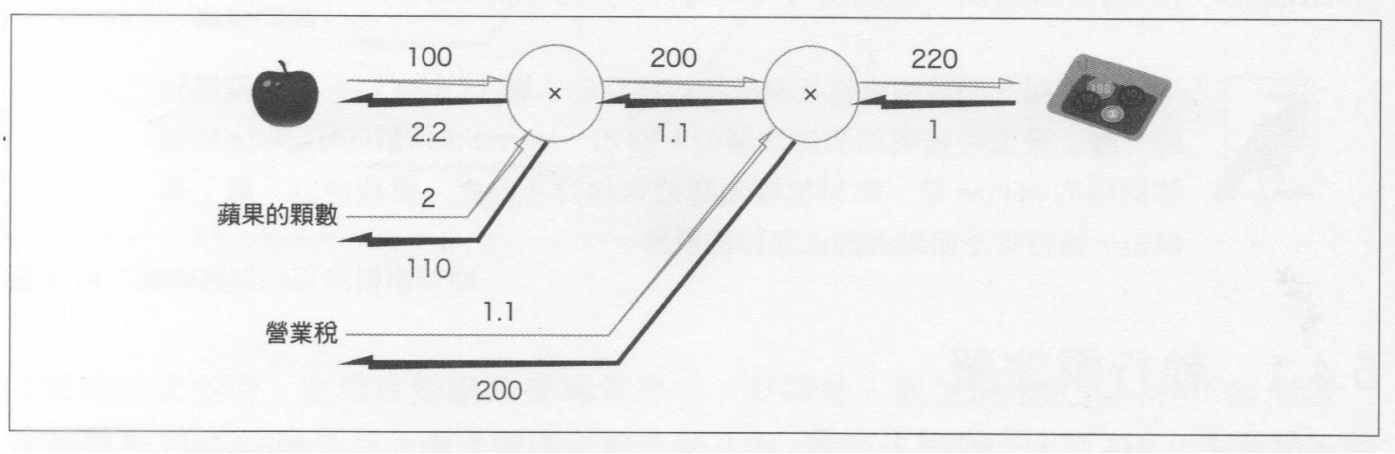
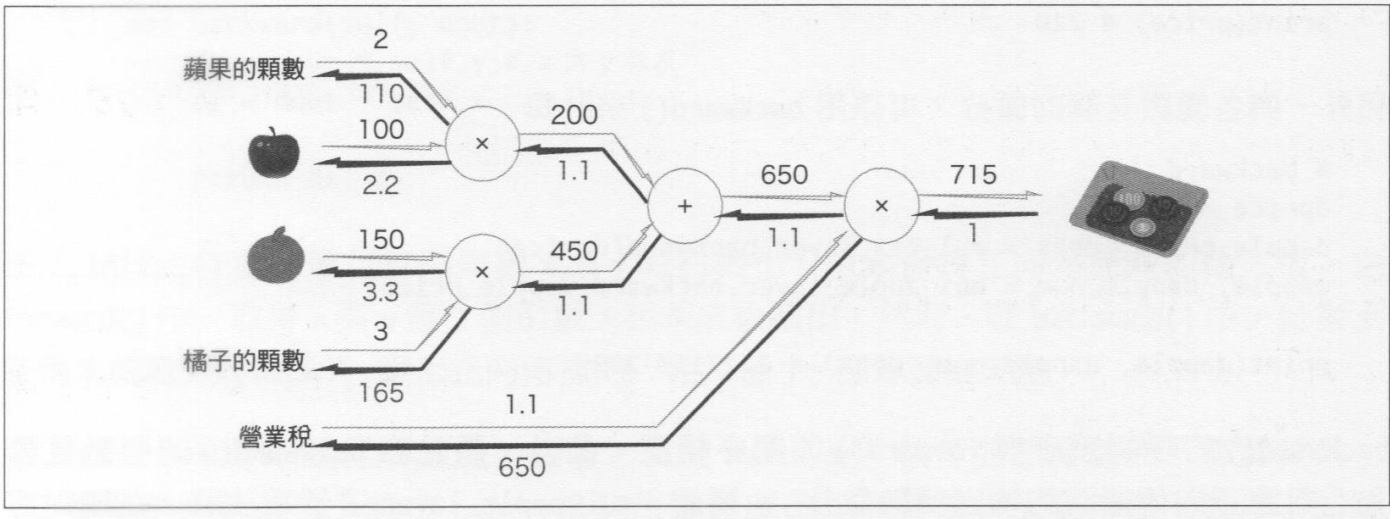
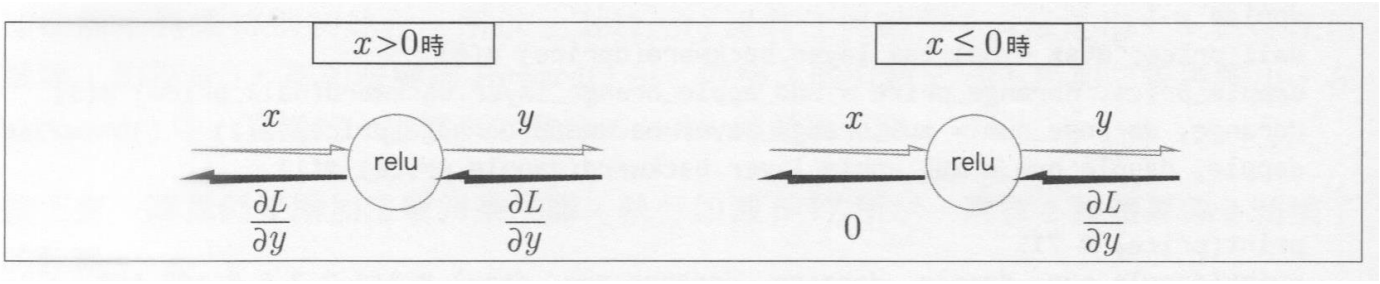


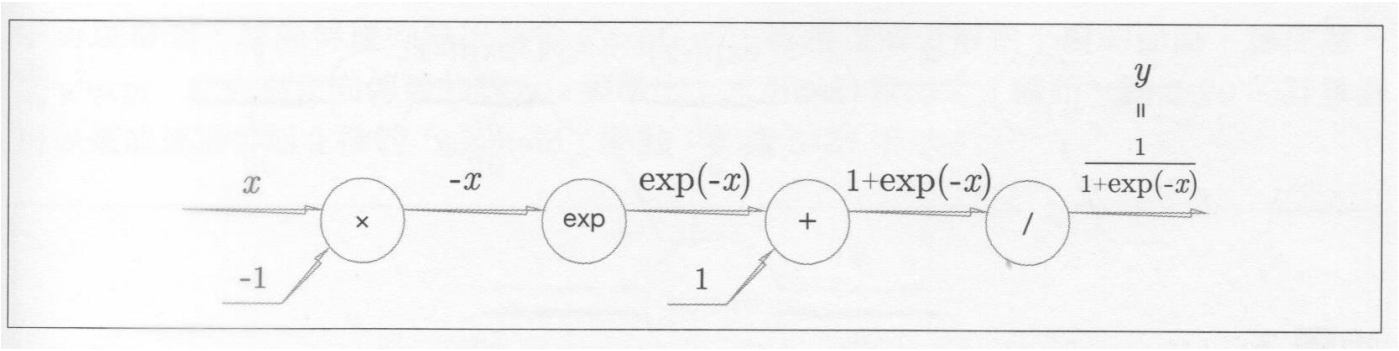


$$y = \begin{cases} x & (x > 0) \\ 0 & (x \leqq 0) \end{cases}$$

$$\frac{\partial y}{\partial x} = \begin{cases} 1 & (x > 0) \\ 0 & (x \leqq 0) \end{cases}$$

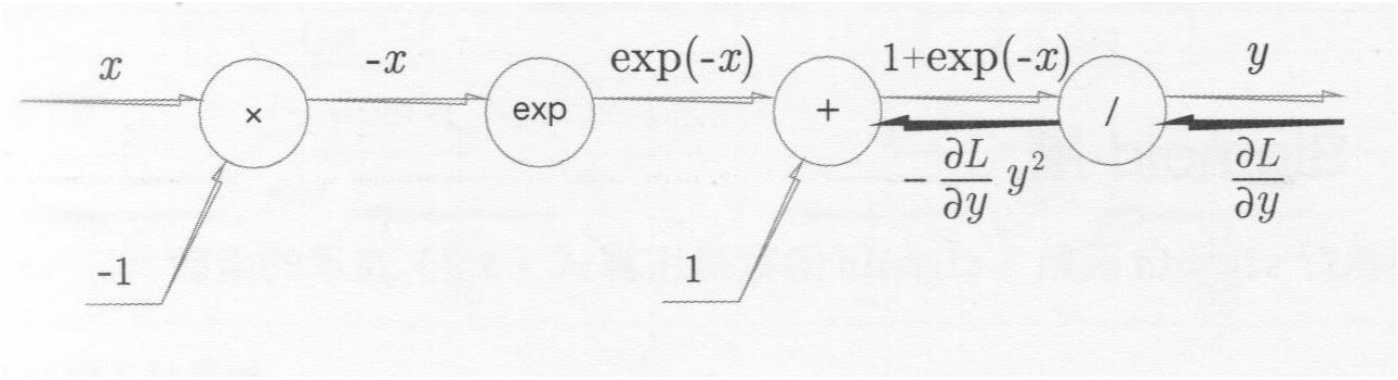


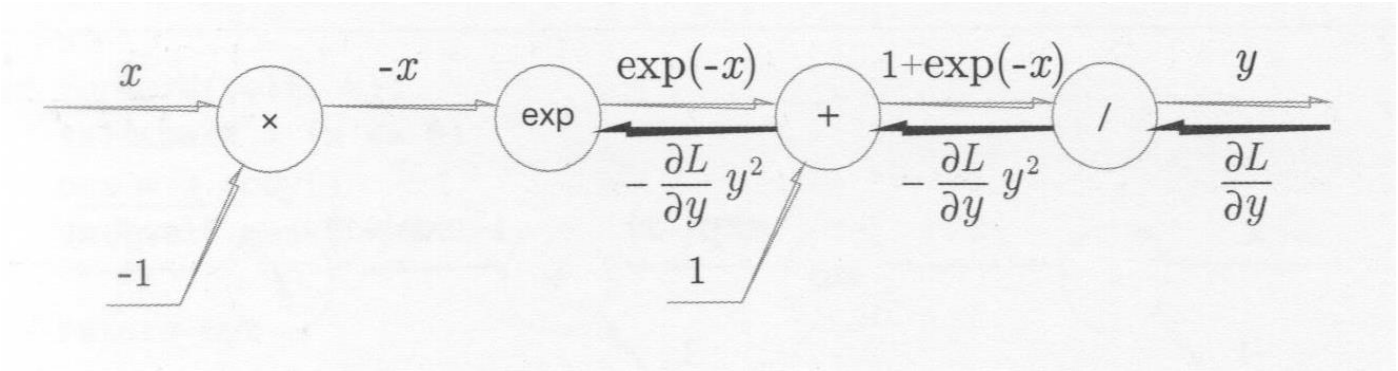
$$y = \frac{1}{1 + \exp(-x)}$$



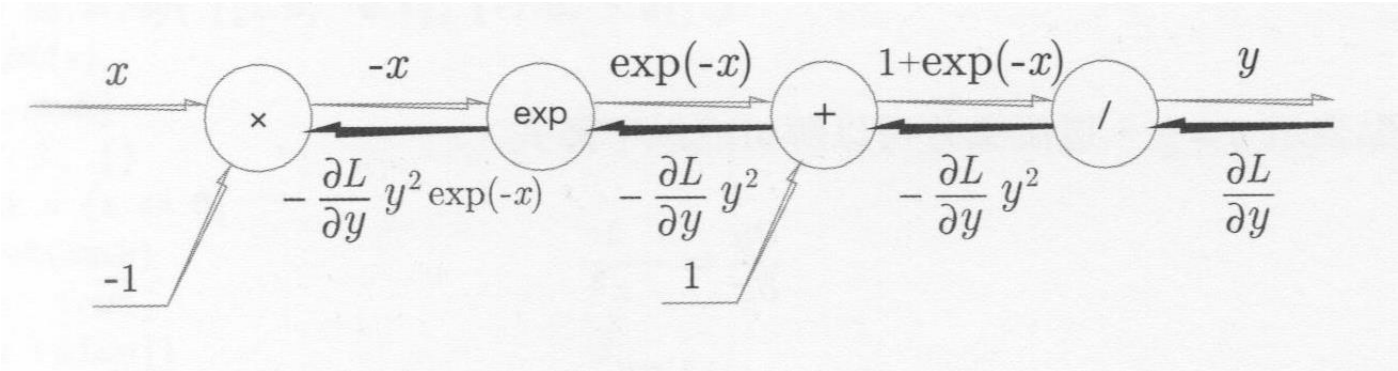
$$\frac{\partial y}{\partial x} = -\frac{1}{x^2}$$

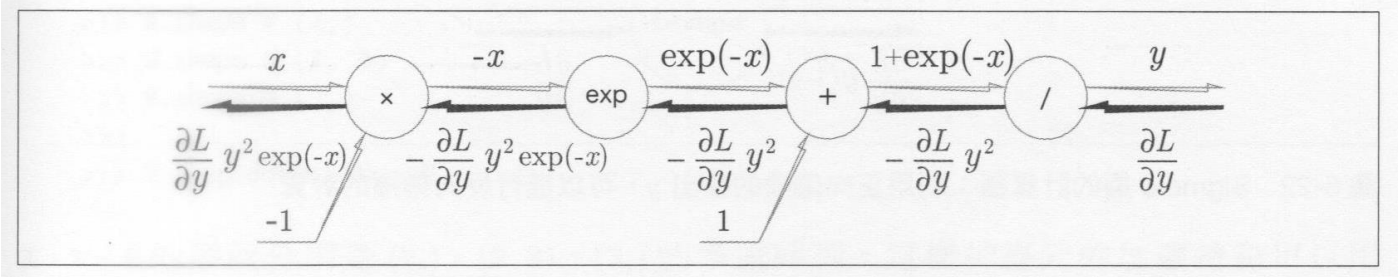
$$= -y^2$$

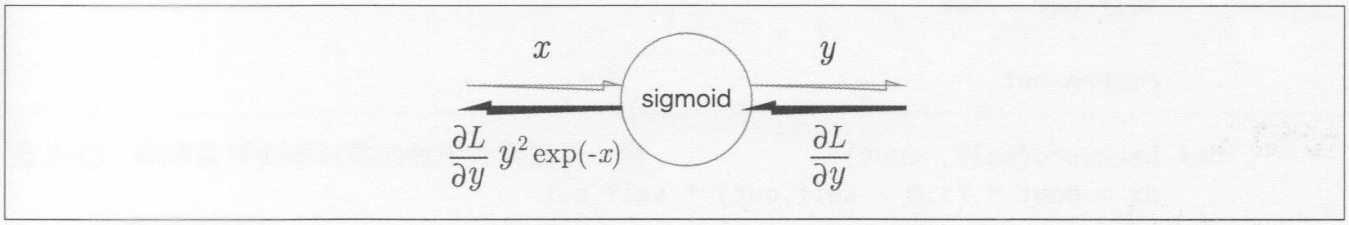




$$\frac{\partial y}{\partial x} = \exp(x)$$

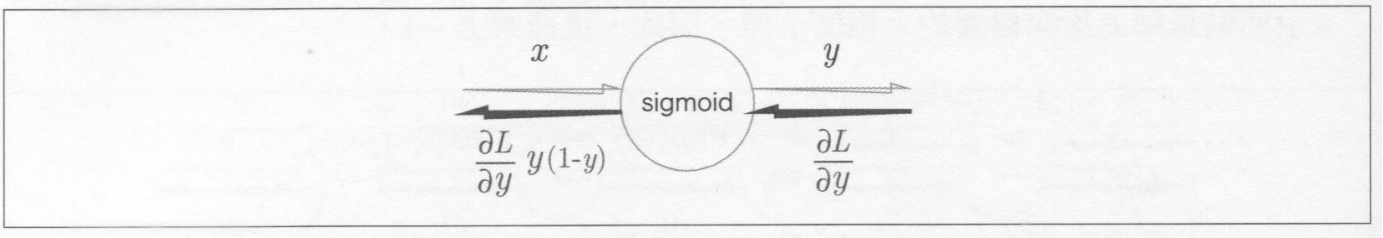


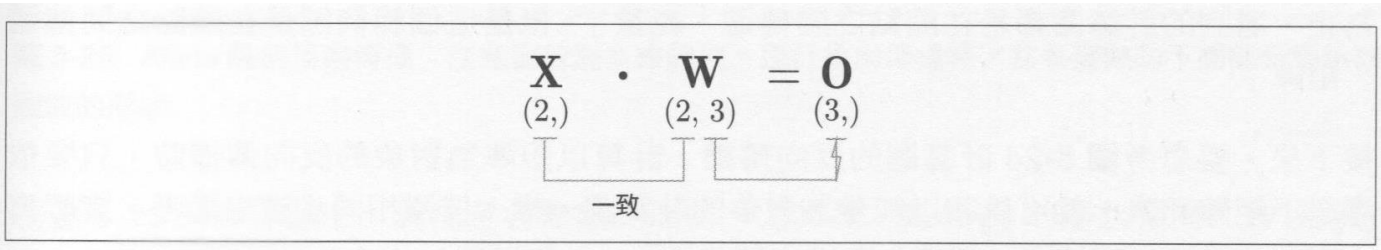


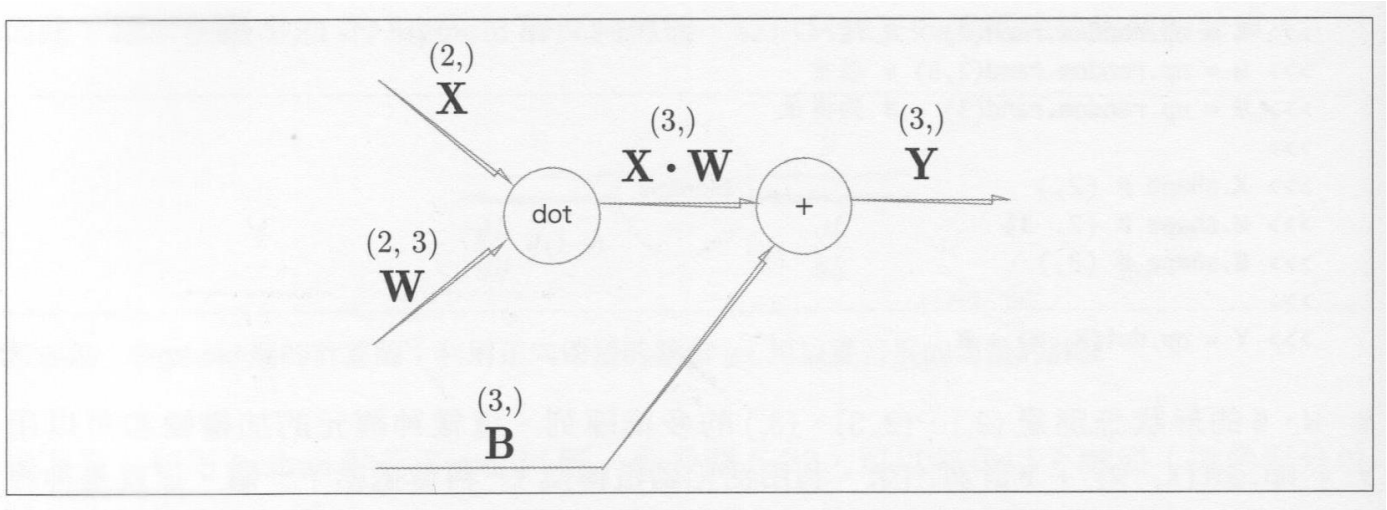


$\frac{\partial L}{\partial y} y^2 \exp(-x)$ 可以整理如下。

$$\begin{aligned}\frac{\partial L}{\partial y} y^2 \exp(-x) &= \frac{\partial L}{\partial y} \frac{1}{(1+\exp(-x))^2} \exp(-x) \\ &= \frac{\partial L}{\partial y} \frac{1}{1+\exp(-x)} \frac{\exp(-x)}{1+\exp(-x)} \\ &= \frac{\partial L}{\partial y} y(1-y)\end{aligned}$$







$$\frac{\partial L}{\partial \mathbf{X}} = \frac{\partial L}{\partial \mathbf{Y}} \cdot \mathbf{W}^{\mathrm{T}}$$

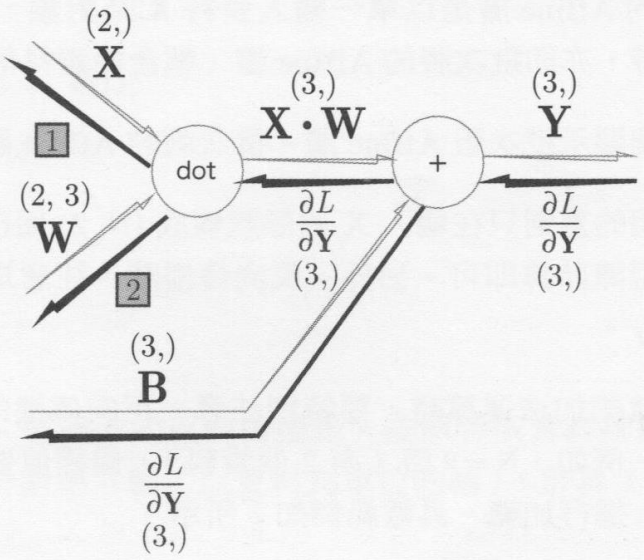
$$\frac{\partial L}{\partial \mathbf{W}} = \mathbf{X}^{\mathrm{T}} \cdot \frac{\partial L}{\partial \mathbf{Y}}$$

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix}$$

$$\mathbf{W}^{\mathrm{T}} = \begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{pmatrix}$$

1 $\underset{(2,)}{\frac{\partial L}{\partial \mathbf{X}}} = \underset{(3,)}{\frac{\partial L}{\partial \mathbf{Y}}} \cdot \underset{(3,\ 2)}{\mathbf{W}^{\mathrm{T}}}$

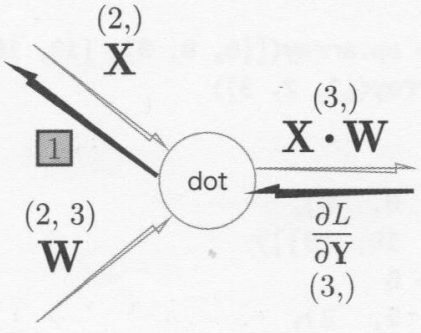
2 $\underset{(2,\ 3)}{\frac{\partial L}{\partial \mathbf{W}}} = \underset{(2,\ 1)}{\mathbf{X}^{\mathrm{T}}} \cdot \underset{(1,\ 3)}{\frac{\partial L}{\partial \mathbf{Y}}}$



$$\mathbf{X} = (x_0, x_1, \cdots, x_n)$$

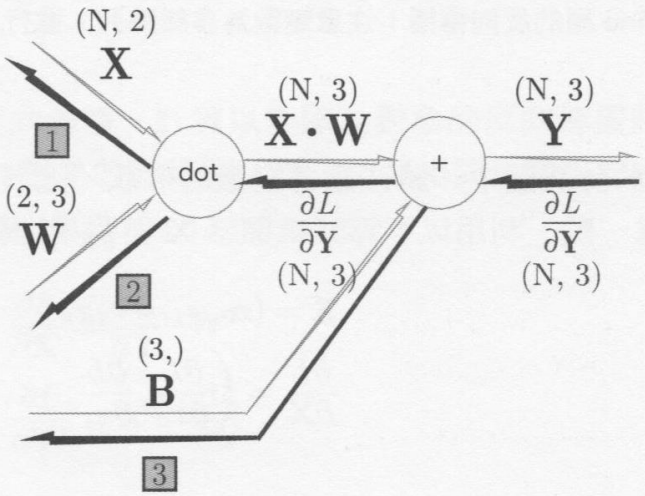
$$\frac{\partial L}{\partial \mathbf{X}} = \left(\frac{\partial L}{\partial x_0}, \frac{\partial L}{\partial x_1}, \cdots, \frac{\partial L}{\partial x_n} \right)$$

1 $\underset{(3,)}{\frac{\partial L}{\partial \mathbf{Y}}} \cdot \underset{(3,\ 2)}{\mathbf{W}^{\mathrm{T}}} = \underset{(2,)}{\frac{\partial L}{\partial \mathbf{X}}}$

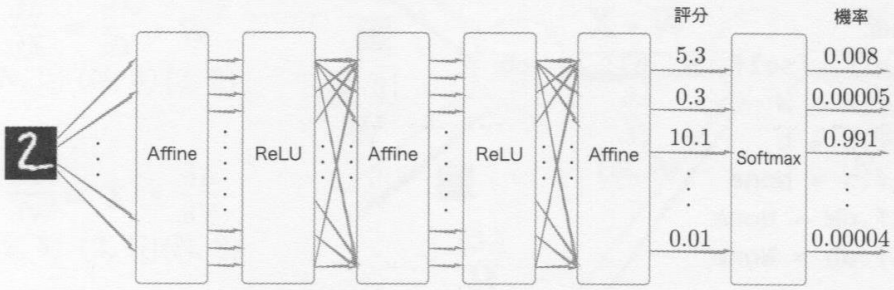


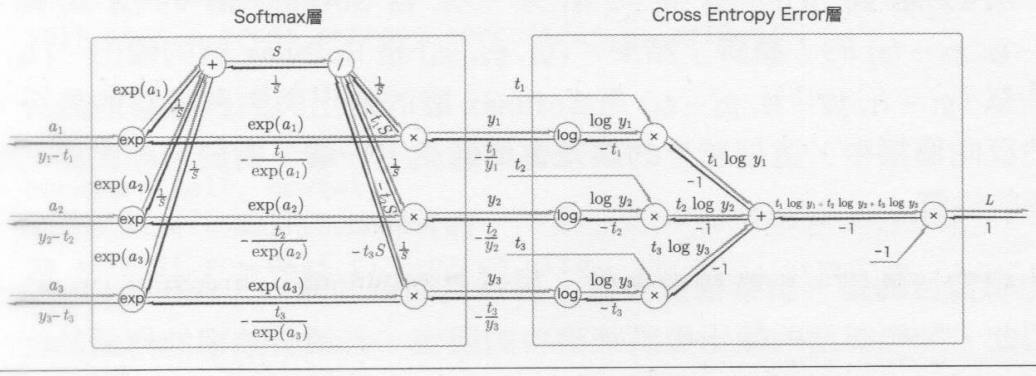
1 $\frac{\partial L}{\partial \mathbf{X}} = \frac{\partial L}{\partial \mathbf{Y}} \cdot \mathbf{W}^{\mathrm{T}}$
(N, 2) (N, 3) (3, 2)

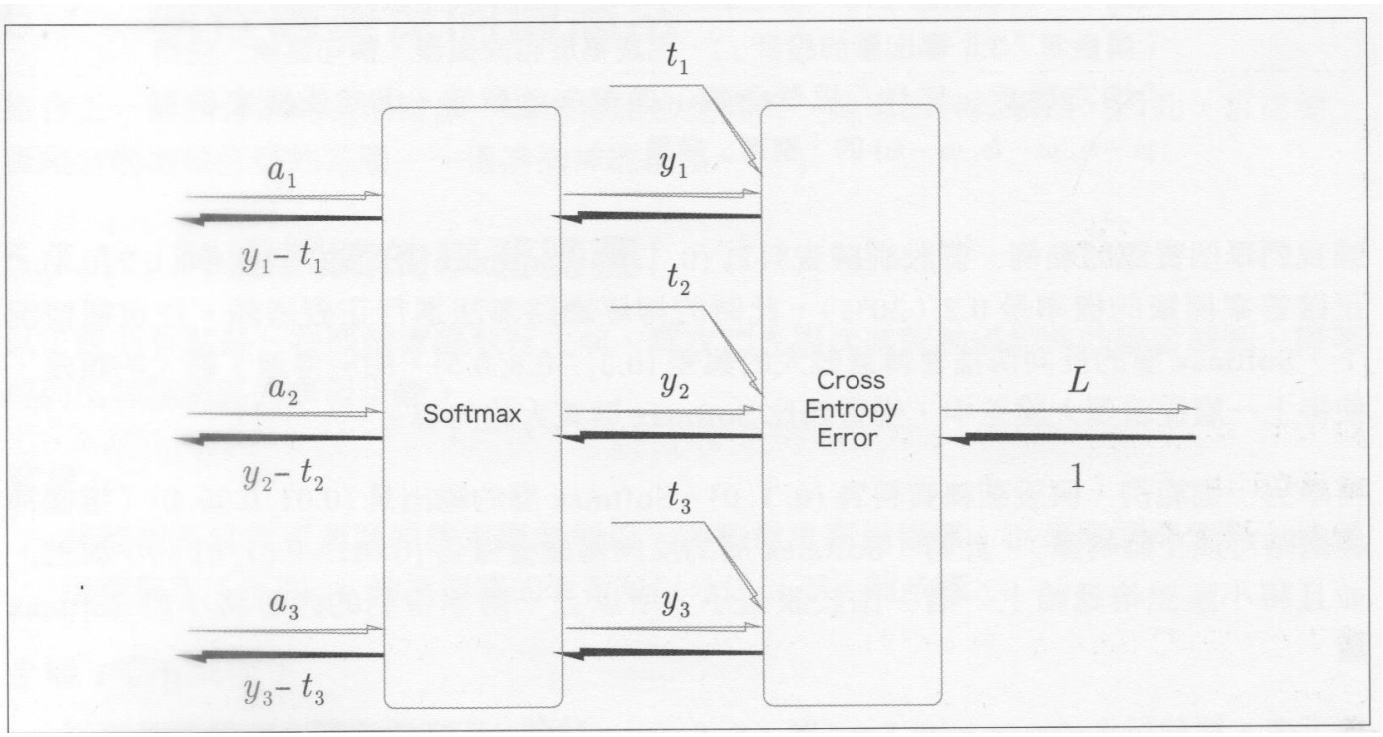
2 $\frac{\partial L}{\partial \mathbf{W}} = \mathbf{X}^{\mathrm{T}} \cdot \frac{\partial L}{\partial \mathbf{Y}}$
(2, 3) (2, N) (N, 3)



3 與 $\frac{\partial L}{\partial \mathbf{B}} = \frac{\partial L}{\partial \mathbf{Y}}$ 最初軸（第0軸）有關的和
(3) (N, 3)






t_1
a_1
y_1
$y_1 - t_1$
t_2
a_2
Softmax
y_2
Cross
Entropy
Error
L
$y_2 - t_2$
1
t_3
a_3
y_3
$y_3 - t_3$