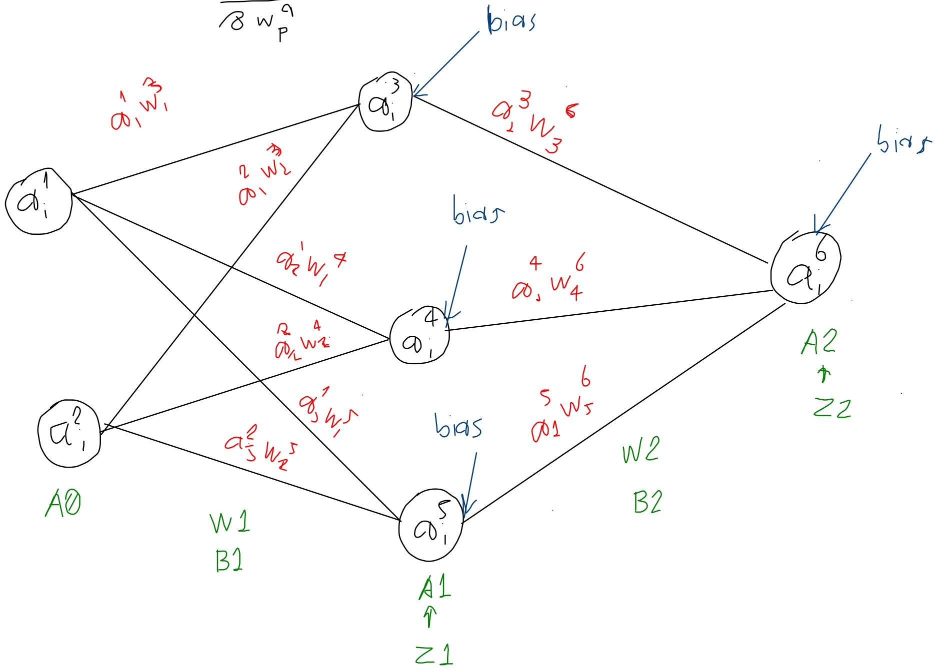
พิจารณาสมหการ Gradient Descent $\frac{\partial Error}{\partial w_p^q}$ และ Backpropagation สำหรับ Classification



$$A2 = \begin{bmatrix} a_1^6 \\ a_2^6 \\ \vdots \\ a_N^6 \end{bmatrix} = \begin{bmatrix} af(z_1^6) \\ af(z_2^6) \\ \vdots \\ af(z_N^6) \end{bmatrix} = af(Z2)$$

$$Z2 = \begin{bmatrix} z_1^6 \\ z_2^6 \\ \vdots \\ z_N^6 \end{bmatrix} = \begin{bmatrix} w_3^6 a_1^3 + w_4^6 a_1^4 + w_5^6 a_1^5 + b^6 \\ w_3^6 a_2^3 + w_4^6 a_2^4 + w_5^6 a_2^5 + b^6 \\ \vdots \\ w_3^6 a_N^3 + w_4^6 a_N^4 + w_5^6 a_N^5 + b^6 \end{bmatrix} = A1\,W2 + B2$$

$$W2 = \begin{bmatrix} w_3^6 \\ w_4^6 \\ w_5^6 \end{bmatrix} ; \quad B2 = [\,b^6\,]$$

$$A1 = \begin{bmatrix} a_1^3 & a_1^4 & a_1^5 \\ a_2^3 & a_2^4 & a_2^5 \\ \vdots & \vdots & \vdots \\ a_N^3 & a_N^4 & a_N^5 \end{bmatrix} = \begin{bmatrix} af(z_1^3) & af(z_1^4) & af(z_1^5) \\ af(z_2^3) & af(z_2^4) & af(z_2^5) \\ \vdots & \vdots & \vdots \\ af(z_3^3) & af(z_N^4) & af(z_N^5) \end{bmatrix} = af(Z1)$$

$$
Z1 = \begin{bmatrix} z_1^3 & z_1^4 & z_1^5 \\ z_2^3 & z_2^4 & z_2^5 \\ & \vdots & \\ z_N^3 & z_N^4 & z_N^5 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{2} w_j^3 a_1^j + b^3 & \sum_{j=1}^{2} w_j^4 a_1^j + b^4 & \sum_{j=1}^{2} w_j^5 a_1^j + b^5 \\ \sum_{j=1}^{2} w_j^3 a_2^j + b^3 & \sum_{j=1}^{2} w_j^4 a_2^j + b^4 & \sum_{j=1}^{2} w_j^5 a_2^j + b^5 \\ \vdots & \vdots & \vdots \\ \sum_{j=1}^{2} w_j^3 a_N^j + b^3 & \sum_{j=1}^{2} w_j^4 a_N^j + b^4 & \sum_{j=1}^{2} w_j^5 a_N^j + b^5 \end{bmatrix}
$$

$$
= A0\,W1 + B1
$$

$$
W1 = \begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} \; ; \quad B1 = \begin{bmatrix} b^3 & b^4 & b^5 \end{bmatrix} \; ; \quad A0 = \begin{bmatrix} a_1^1 & a_1^2 \\ a_2^1 & a_2^2 \\ \vdots & \vdots \\ a_N^1 & a_N^2 \end{bmatrix}
$$

$$w_p^q = w_p^q - \alpha \frac{\partial Error}{\partial w_p^q}$$

ส่งจากโหนด q → โหนด p

Binary Class

↳ ใช้กับ Multiclass ก็ได้

$$Error = \frac{-\sum_{i=1}^{N} t_i \log(\hat{t}_i) + (1-t_i)\log(1-\hat{t}_i)}{N} = \frac{-\sum_{i=1}^{N} Error_i}{N}$$

Cross Entropy

$$Error_i = t_i \log(\hat{t}_i) + (1-t_i)\log(1-\hat{t}_i)$$

หา Error ชั้นสุดท้าย แล้ว Backpropagation กลับ เพื่อปรับ weight แต่ละรอบ

เราจะพิจารณาสมการ Gradient Descent 4 สมการ ดังนี้

$$W2 = W2 - \alpha \frac{\partial Error}{\partial W2}$$

หลังไปหน้า

$$B2 = B2 - \alpha \frac{\partial Error}{\partial B2}$$

$$W1 = W1 - \alpha \frac{\partial Error}{\partial W1}$$

$$B1 = B1 - \alpha \frac{\partial Error}{\partial B1}$$

α

(1) $w2 = w2 - \alpha \frac{\partial \text{Error}}{\partial w2}$

Gradient Descent

$$\begin{bmatrix} w_3^6 \\ w_4^6 \\ w_5^6 \end{bmatrix} = \begin{bmatrix} w_3^6 \\ w_4^6 \\ w_5^6 \end{bmatrix} - \alpha \begin{bmatrix} \frac{\partial \text{Error}}{\partial w_3^6} \\ \frac{\partial \text{Error}}{\partial w_4^6} \\ \frac{\partial \text{Error}}{\partial w_5^6} \end{bmatrix}$$

พิจารณา $\frac{\partial \text{Error}}{\partial w_p^q}$ เมื่อ $p \in \{3,4,5\}$ และ $q \in \{6\}$

$$w_p^q = w_p^q - \alpha \frac{\partial \text{Error}}{\partial w_p^q}$$

$$= w_p^q - \alpha \frac{\partial \frac{\sum_{i=1}^{N} \text{Error}_i}{N}}{\partial w_p^q}$$

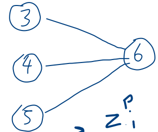
$$= w_p^q + \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial \text{Error}_i}{\partial w_p^q}$$

$$w_p^q = w_p^q + \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial \left(t_i \log(a_i^q) + (1-t_i)\log(1-a_i^q)\right)}{\partial w_p^q}$$

$$w_p^q = w_p^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(\frac{\partial t_i \log(a_i^q)}{\partial w_p^q} + \frac{(1-t_i)\log(1-a_i^q)}{\partial w_p^q} \right)$$

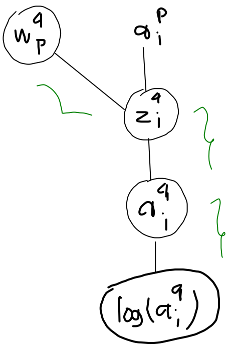
chain rule

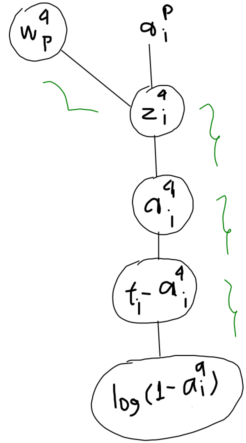
พิจารณา $\frac{\partial Error}{\partial w_p^q}$ เมื่อ $p \in \{3,4,5\}$ และ $q \in \{6\}$



$w \rightarrow z_i^p$

แปวันไป





$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{\partial t_i \log(a_i^q)}{\partial w_p^q} + \frac{\partial (1-t_i)\log(1-a_i^q)}{\partial w_p^q}\right)$$

$$\frac{\partial \log(x)}{\partial x} = \frac{1}{x}$$

$$w_p^q = w_p^q + \frac{\alpha}{N} \sum_{1}^{N} \left(\frac{\partial t_i \log(a_i^q)}{\partial w_p^q} + \frac{(1-t_i)\log(1-a_i^q)}{\partial w_p^q} \right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(t_i \frac{\partial \log(a_i^q)}{\partial w_p^q} \frac{\partial a_i^q}{\partial z_i^q} \frac{\partial z_i^q}{\partial w_p^q} + (1-t_i) \frac{\partial \log(1-a_i^q)}{\partial (1-a_i^q)} \frac{\partial (1-a_i^q)}{\partial a_i^q} \frac{\partial a_i^q}{\partial z_i^q} \frac{\partial z_i^q}{\partial w_p^q} \right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(t_i \left(\frac{1}{a_i^q}\right) \frac{\partial a_i^q}{\partial z_i^q} a_i^p + (1-t_i) \left(\frac{1}{1-a_i^q}\right)(-1) \frac{\partial a_i^q}{\partial z_i^q} (a_i^p) \right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(\frac{t_i}{a_i^q} \frac{\partial a_i^q}{\partial z_i^q} a_i^p - \frac{1-t}{1-a_i^q} \frac{\partial a_i^q}{\partial z_i^q} (a_i^p) \right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(\frac{t_i}{a_i^q} - \frac{1-t}{1-a_i^q} \right) \mathrm{diff}_i^q \, (a_i^p)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i}{a_i^q} - \frac{1-t}{1-a_i^q}\right) \mathrm{diff}_i^q(a_i^p)$$

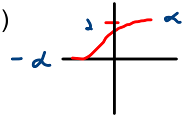
$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i(1-a_i^q) - (1-t_i)a_i^q}{a_i^q(1-a_i^q)}\right) \mathrm{diff}_i^q(a_i^p)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i - \cancel{t_i a_i^q} - a_i^q + \cancel{t_i a_i^q}}{a_i^q(1-a_i^q)}\right) \mathrm{diff}_i^q(a_i^p)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i - a_i^q}{a_i^q(1-a_i^q)}\right) \mathrm{diff}_i^q(a_i^p)$$

ยังไม่ผ่าน Activation function
(ตัวอย่างนี้ให้เป็น simoid ในชั้นสุดท้าย)

ยังไม่ผ่าน Activation function (ตัวอย่างนี้ให้เป็น simoid ในชั้นสุดท้าย)
เพื่อให้เห็นภาพ เรามา Diff simoid กัน



$$\frac{1}{1+e^{-z}}$$

$$\frac{d}{dz}\left[\frac{1}{1+e^{-z}}\right] \Rightarrow \frac{d}{dz}(1+e^{-z})^{-1}$$

$$= (-1)(1+e^{-z})^{-2}\frac{d}{dz}(1+e^{-z})$$

$$= -(1+e^{-z})^{-2}(-e^{-z}) \rightarrow \frac{e^{-z}}{(1+e^{-z})(1+e^{-z})}$$

สรุป

$$\frac{1}{1+e^{-z}}\cdot\frac{e^{-z}}{1+e^{-z}}$$

$e^{-z}+1-1$

$$\rightarrow \frac{1}{1+e^{-z}}\cdot\frac{(1+e^{-z})-1}{1+e^{-z}}$$

$$\rightarrow \frac{1}{1+e^{-z}}\cdot\left(\frac{1+e^{-z}}{1+e^{-z}}-\frac{1}{1+e^{-z}}\right)$$

$$\rightarrow \frac{1}{1+e^{-z}}\cdot\left(1-\frac{1}{1+e^{-z}}\right)$$

$$a_i^a(1-a_i^a)$$

ในที่นี้ให้ Activation functionชั้นสุดท้ายเป็น simoid ดังนั้น..

$$diff_i^q = a_i^q(1-a_i^q)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t-a_i^q}{a_i^q(1-a_i^q)}\right)\underline{a_i^q(1-a_i^q)}\,a_i^p$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i-a_i^q}{a_i^q(1-a_i^q)}\right)\underline{diff_i^q}\,(a_i^p)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}(t_i-a_i^q)\,a_i^p \rightarrow$$

เหมือน Logistic Regression

$$w_d + \frac{\alpha}{N}\sum_{i=1}^{N}(Y_i-\hat{Y}_i)\,x_i^d$$

ดังนั้น สามารถเขียนให้อยู่ในรูป Matrix ได้ดังนี้

$$\begin{bmatrix} w^6_3 \\ w^6_4 \\ w^6_5 \end{bmatrix} = \begin{bmatrix} w^6_3 \\ w^6_4 \\ w^6_5 \end{bmatrix} + \frac{\alpha}{N} \begin{bmatrix} \sum_{i=1}^{N} (t_i - a^6_i)\, diff^6_i\, a^3_i \\ \sum_{i=1}^{N} (t_i - a^6_i)\, diff^6_i\, a^4_i \\ \sum_{i=1}^{N} (t_i - a^6_i)\, diff^6_i\, a^5_i \end{bmatrix}$$

$a \to$ โหนดที่ 6

เนื่องจากผ่าน simoid ทำให้ Diff2 = 1

$$\begin{bmatrix} w^6_3 \\ w^6_4 \\ w^6_5 \end{bmatrix} = \begin{bmatrix} w^6_3 \\ w^6_4 \\ w^6_5 \end{bmatrix} + \frac{\alpha}{N} \begin{bmatrix} \sum_{i=1}^{N} (t_i - a^6_i)\, a^3_i \\ \sum_{i=1}^{N} (t_i - a^6_i)\, a^4_i \\ \sum_{i=1}^{N} (t_i - a^6_i)\, a^5_i \end{bmatrix}$$

$$\begin{bmatrix} w^6_3 \\ w^6_4 \\ w^6_5 \end{bmatrix} = \begin{bmatrix} w^6_3 \\ w^6_4 \\ w^6_5 \end{bmatrix} + \frac{\alpha}{N} \underbrace{\begin{bmatrix} a^3_1 & a^3_2 & \cdots & a^3_N \\ a^4_1 & a^4_2 & \cdots & a^4_N \\ a^5_1 & a^5_2 & \cdots & a^5_N \end{bmatrix}}_{A1^T} \underbrace{\begin{bmatrix} (t_1 - a^6_1) \\ (t_1 - a^6_2) \\ \vdots \\ (t_1 - a^6_3) \end{bmatrix}}_{\delta 2}$$

กำหนดให้

$$\delta 2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} \; ; \quad Diff2 = 1$$

เนื่องจากผ่าน simoid ทำให้ Diff2 = 1

คูณแบบ " * " ตัวต่อตัว

$$Err2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} * 1 = \delta 2 * Diff2$$

เพราะ

$$A1^T = \begin{bmatrix} a_1^3 & a_2^3 & \dots & a_N^3 \\ a_1^4 & a_2^4 & \dots & a_N^4 \\ a_1^5 & a_2^5 & \dots & a_N^5 \end{bmatrix}$$

ดังนั้นสมการ Gradient Descent ของ W2 ที่ได้

$$\boxed{W2 = W2 + \frac{\alpha}{N} A1^T Err2}$$

② $B2 = B2 - \alpha \frac{\partial Error}{\partial B2}$

สมการ Gradient Descen ของ B2 จะอยู่ในรูปนี้นะ

$$[b^6] = [b^6] - \alpha \left[\frac{\partial Error}{\partial b^6} \right]$$

พิจารณา $\frac{\partial Error}{\partial b^q}$ เมื่อ $q \in \{6\}$

$$b^q = b^q - \alpha \frac{\partial Error}{\partial b^q}$$

$$= b^q - \alpha \frac{\partial \frac{\sum_{i=1}^{N} Error_i}{N}}{\partial b^q}$$

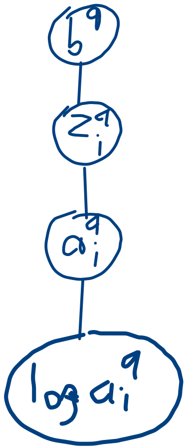
$$= b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial Error_i}{\partial b^q}$$

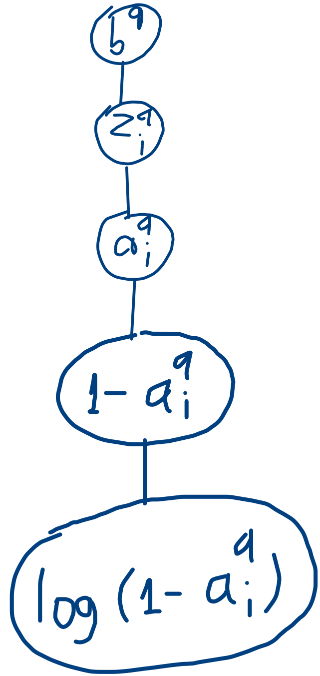
$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial \left(t_i \log(a_i^q) + (1-t_i) \log(1-a_i^q) \right)}{\partial b^q}$$

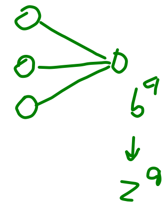
$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(\frac{\partial t_i \log(a_i^q)}{\partial b^q} + \frac{(1-t_i)\log(1-a_i^q)}{\partial b^q} \right)$$

chain rule

$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(\frac{t_i \log(a_i^q)}{b^q} + \frac{(1-t_i)\log(1-a_i^q)}{b^q} \right)$$







$$\frac{\partial \log(x)}{\partial x} = \frac{1}{x}$$

$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(\frac{\partial t_i \log(a_i^q)}{b^q} + \frac{(1-t_i)\log(1-a_i^q)}{b^q} \right)$$

$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(t_i \frac{\partial \log(a_i^q)}{\partial b^q} \frac{\partial a_i^q}{\partial z_i^q} \frac{\partial z_i^q}{\partial b^q} + (1-t_i) \frac{\partial \log(1-a_i^q)}{\partial (1-a_i^q)} \frac{\partial (1-a_i^q)}{\partial a_i^q} \frac{\partial a_i^q}{\partial z_i^q} \frac{\partial z_i^q}{\partial b^q} \right)$$

$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(t_i \left(\frac{1}{a_i^q}\right) \frac{\partial a_i^q}{\partial z_i^q} (1) + (1-t_i) \left(\frac{1}{1-a_i^q}\right) (-1) \frac{\partial a_i^q}{\partial z_i^q} (1) \right)$$

$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(\frac{t_i}{a_i^q} \frac{\partial a_i^q}{\partial z_i^q} - \frac{1-t}{1-a_i^q} \frac{\partial a_i^q}{\partial z_i^q} \right)$$

$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(\frac{t_i}{a_i^q} - \frac{1-t}{1-a_i^q} \right) diff_i^q$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i}{a_i^q} - \frac{1-t}{1-a_i^q}\right) diff_i^q$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i(1-a_i^q) - (1-t_i)a_i^q}{a_i^q(1-a_i^q)}\right) diff_i^q$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i - t_i a_i^q - a_i^q + t_i a_i^q}{a_i^q(1-a_i^q)}\right) diff_i^q$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i - a_i^q}{a_i^q(1-a_i^q)}\right) diff_i^q$$

ยังไม่ผ่าน Activation function
(ตัวอย่างนี้ให้เป็น simoid ในชั้นสุดท้าย)

ในที่นี้ให้ Activation functionชั้นสุดท้ายเป็น simoid ดังนั้น..

$$diff_i^q = a_i^q(1-a_i^q)$$

$$b^q = b^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t-a_i^q}{a_i^q(1-a_i^q)}\right)a_i^q(1-a_i^q)$$

$$b^q = b^q + \frac{\alpha}{N}\sum_{i=1}^{N}(t_i-a_i^q)$$

ดังนั้น สามารถเขียนให้อยู่ในรูป Matrix ได้ดังนี้ $[b^6] = [b^6] + \frac{\alpha}{N}\left[\sum_{i=1}^{N}(t_i - a_i^6)\right]$

$$[b^6] = [b^6] + \frac{\alpha}{N}[1\ 1\ \dots\ 1]\begin{bmatrix}(t_i - a_i^6)\\(t_i - a_i^6)\\(t_i - a_i^6)\end{bmatrix}$$

q → โหนดที่ 6

เนื่องจากผ่าน simoid ทำให้ Diff2 = 1

กำหนดให้

$$\delta 2 = \begin{bmatrix}t_1 - a_1^6\\t_2 - a_2^6\\\vdots\\t_N - a_N^6\end{bmatrix} \quad ; \quad \text{Diff}2 = 1$$

$$\text{Err}2 = \begin{bmatrix}t_1 - a_1^6\\t_2 - a_2^6\\\vdots\\t_N - a_N^6\end{bmatrix} * 1 = \delta 2 * \text{Diff}2$$

ดังนั้นสมการ Gradient Descent ของ W2 ที่ได้

$$\boxed{B2 = B2 + \frac{\alpha}{N}\text{Err}2.\text{Sum}(\text{axis}=0)}$$

③ $$W1 = W1 - \alpha \cdot \frac{\partial Error}{\partial W1}$$

สมการ Gradient Descen ของ W1 จะอยู่ในรูปนี้นะ

$$\begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} = \begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} - \alpha \begin{bmatrix} \frac{\partial Error}{\partial w_1^3} & \frac{\partial Error}{\partial w_1^4} & \frac{\partial Error}{\partial w_1^5} \\ \frac{\partial Error}{\partial w_2^3} & \frac{\partial Error}{\partial w_2^4} & \frac{\partial Error}{\partial w_2^5} \end{bmatrix}$$

พิจารณา $\frac{\partial Error}{\partial w_p^q}$ เมื่อ $p \in \{1,2\}$ และ $q \in \{3,4,5\}$

$$w_p^q = w_p^q - \alpha \frac{\partial Error}{\partial w_p^q}$$

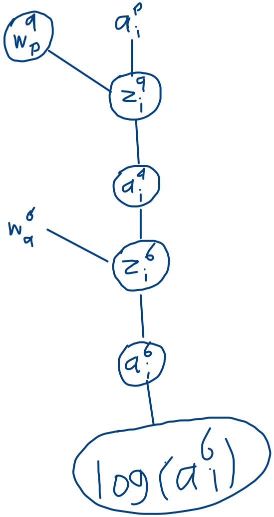
$$w_p^q = w_p^q - \alpha \frac{\partial \frac{\sum_{i=1}^{N} Error_i}{N}}{\partial w_p^q}$$

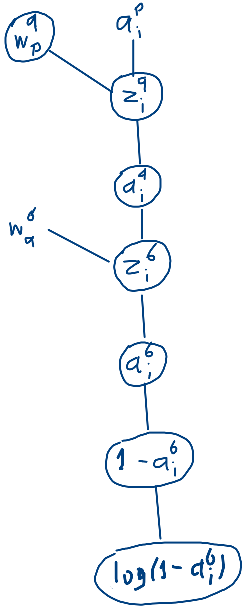
$$w_p^q = w_p^q + \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial Error_i}{\partial w_p^q}$$

$$w_p^q = w_p^q + \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial \left(t_i \log(a_i^q) + (1-t_i)\log(1-a_i^q)\right)}{\partial w_p^q}$$

$$w_p^q = w_p^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(\frac{\partial t_i \log(a_i^q)}{\partial w_p^q} + \frac{(1-t_i)\log(1-a_i^q)}{\partial w_p^q} \right)$$

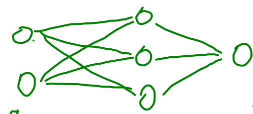
chain rule





$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{\partial t_i \log(a_i^q)}{\partial w_p^q} + \frac{(1-t_i)\log(1-a_i^q)}{\partial w_p^q}\right)$$

$p \in \{1,2\}$ $q \in \{3,4,5\}$



$q \in \{6\}$

$w_p^q \longleftrightarrow z_i^q \longleftrightarrow z_i^q \; (z_i^6)$

ทำให้
z เปลี่ยน 2 อัน

$$\frac{\partial \log(x)}{\partial x} = \frac{1}{x}$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{\partial t_i \log(a_i^q)}{\partial w_p^q} + \frac{(1-t_i)\log(1-a_i^q)}{\partial w_p^q}\right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\Bigg(t_i \frac{\partial \log(a_i^q)}{\partial a_i^6}\frac{\partial a_i^6}{\partial z_i^6}\frac{\partial z_i^6}{\partial a_i^q}\frac{\partial a_i^q}{\partial z_i^q}\frac{\partial z_i^q}{\partial w_p^q}$$

$$+ (1-t_i)\frac{\partial \log(1-a_i^6)}{\partial(1-a_i^6)}\frac{\partial(1-a_i^6)}{\partial a_i^6}\frac{\partial a_i^6}{\partial z_i^6}\frac{\partial z_i^6}{\partial a_i^q}\frac{\partial a_i^q}{\partial z_i^q}\frac{\partial z_i^q}{\partial w_p^q}\Bigg)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(t_i\left(\frac{1}{a_i^6}\right)\frac{\partial a_i^6}{\partial z_i^6} w_q^6 \frac{\partial a_i^q}{\partial z_i^q} a_i^p + (1-t_i)\left(\frac{1}{1-a_i^6}\right)(-1)\frac{\partial a_i^6}{\partial z_i^6} w_q^6 \frac{\partial a_i^q}{\partial z_i^q} a_i^p\right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i}{a_i^4}\underline{\frac{\partial a_i^6}{\partial z_i^6}} w_q^6 \underline{\frac{\partial a_i^4}{\partial z_i^q}} a_i^p - \frac{1-t}{1-a_i^6}\underline{\frac{\partial a_i^6}{\partial z_i^6}} w_q^6 \underline{\frac{\partial a_i^4}{\partial z_i^q}}(a_i^p)\right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i}{a_i^4}\, diff_i^6\, w_q^6\, diff_i^4\, a_i^p - \frac{1-t}{1-a_i^6}\, diff_i^6\, w_q^6\, diff_i^4\, (a_i^p)\right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i}{a_i^q}\,diff_i^6\,w_q^6\,diff_i^q\,a_i^p - \frac{1-t}{1-a_i^q}\,diff_i^6\,w_q^6\,diff_i^q\,(a_i^p)\right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i}{a_i^q} - \frac{1-t}{1-a_i^q}\right)diff_i^6\,w_q^6\,diff_i^q\,a_i^p$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i(1-a_i^q) - (1-t_i)a_i^q}{a_i^q(1-a_i^q)}\right)diff_i^6\,w_q^6\,diff_i^q\,a_i^p$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i - t_i a_i^q - a_i^q + t_i a_i^q}{a_i^q(1-a_i^q)}\right)diff_i^6\,w_q^6\,diff_i^q\,a_i^p$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i - a_i^q}{a_i^q(1-a_i^q)}\right)diff_i^6\,w_q^6\,diff_i^q\,a_i^p$$

ในที่นี้ให้ Activation functionชั้นสุดท้ายเป็น simoid ดังนั้น..

$$diff_i^6 = a_i^6(1-a_i^6)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t-a_i^q}{a_i^q(1-a_i^q)}\right)a_i^q(1-a_i^q)w_q^6 diff_i^q a_i^p$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}(t_i-a_i^q)w_q^6 diff_i^q a_i^p$$

เนื่องจากผ่าน simoid ทำให้ Diff2 = 1

$$\delta 2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} \; ; \; Diff2 = 1$$

$$Err2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ 6 \end{bmatrix} * 1 = \delta 2 * Diff2$$

ใช้

$$Err2_i = (t_i - a_i^6) * diff_i^6$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N} Err2_i \; w_q^6 diff_i^q a_i^p$$

$$\begin{bmatrix} \boxed{1}n & \boxed{2}n & \boxed{3}n \\ \boxed{1}v & \boxed{2}v & \boxed{3}v \end{bmatrix} = \begin{bmatrix} n \\ v \end{bmatrix} \begin{bmatrix} \boxed{1} & \boxed{2} & \boxed{3} \end{bmatrix}$$

รษนก $a = \{3, 4, 5\}$

$n = a^1_i$

$v = a^2_i$

สามารถเขียนในรูป Matrix ได้ดังนี้

$$\begin{bmatrix} w^3_1 & w^4_1 & w^5_1 \\ w^3_2 & w^4_2 & w^5_2 \end{bmatrix} = \begin{bmatrix} w^3_1 & w^4_1 & w^5_1 \\ w^3_2 & w^4_2 & w^5_2 \end{bmatrix} + \frac{\alpha}{N} \begin{array}{c} \begin{bmatrix} \sum_{i=1}^{N} Err2_i\, w^6_3\, diff^3_i\, a^1_i & \sum_{i=1}^{N} Err2_i\, w^6_3\, diff^4_i\, a^1_i & \sum_{i=1}^{N} Err2_i\, w^6_3\, diff^5_i\, a^1_i \\ \sum_{i=1}^{N} Err2_i\, w^6_3\, diff^3_i\, a^2_i & \sum_{i=1}^{N} Err2_i\, w^6_3\, diff^4_i\, a^2_i & \sum_{i=1}^{N} Err2_i\, w^6_3\, diff^5_i\, a^2_i \end{bmatrix} \end{array} \begin{matrix} a^1 \\ a^2 \end{matrix}$$

$$\begin{bmatrix} w^3_1 & w^4_1 & w^5_1 \\ w^3_2 & w^4_2 & w^5_2 \end{bmatrix} = \begin{bmatrix} w^3_1 & w^4_1 & w^5_1 \\ w^3_2 & w^4_2 & w^5_2 \end{bmatrix} + \frac{\alpha}{N} \begin{bmatrix} a^1_1 & a^1_2 & \dots & a^1_N \\ a^2_1 & a^2_2 & \dots & a^2_N \end{bmatrix} \begin{bmatrix} Err2_1 w^6_3 diff^3_1 & Err2_1 w^6_4 diff^4_1 & Err2_1 w^6_5 diff^5_1 \\ Err2_1 w^6_3 diff^3_2 & Err2_1 w^6_4 diff^4_2 & Err2_1 w^6_5 diff^5_2 \\ \vdots & \vdots & \vdots \\ Err2_1 w^6_3 diff^3_N & Err2_1 w^6_4 diff^4_N & Err2_1 w^6_5 diff^5_N \end{bmatrix}$$

$$\begin{bmatrix} w^3_1 & w^4_1 & w^5_1 \\ w^3_2 & w^4_2 & w^5_2 \end{bmatrix} = \begin{bmatrix} w^3_1 & w^4_1 & w^5_1 \\ w^3_2 & w^4_2 & w^5_2 \end{bmatrix} + \frac{\alpha}{N} \begin{bmatrix} a^1_1 & a^1_2 & \dots & a^1_N \\ a^2_1 & a^2_2 & \dots & a^2_N \end{bmatrix} \left(\begin{bmatrix} Err2_1\, w^6_3 & Err2_1\, w^6_4 & Err2\, w^6_5 d \\ Err2_1\, w^6_3 & Err2_1\, w^6_4 & Err2_1\, w^6_5 d \\ \\ Err2_1\, w^6_3 & Err2_1\, w^6_4 & Err2_1\, w^6_5 d \end{bmatrix} * \begin{bmatrix} diff^3_1 & diff^1_1 & dff_1 \\ diff^3_2 & diff^4_2 & dff^5_2 \\ \\ diff^3_N & diff^4_N & dff^5_N \end{bmatrix} \right)$$

$$\underbrace{\begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix}}_{W1} = \begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} + \frac{\alpha}{N} \underbrace{\begin{bmatrix} a_1^1 & a_2^1 & \dots & a_N^1 \\ a_1^2 & a_2^2 & \dots & a_N^2 \end{bmatrix}}_{A\Theta^T} \left(\left(\begin{bmatrix} Err2_1 \\ Err2_2 \\ \vdots \\ Err2_N \end{bmatrix} \underbrace{\begin{bmatrix} w_3^6 & w_4^6 & w_5^6 \end{bmatrix}}_{W_2^T} \right) * \begin{bmatrix} diff_1^3 & diff_1 & diff_1 \\ diff_2^3 & diff_2^4 & diff_2^5 \\ \\ diff_N^3 & diff_N^4 & diff_N^5 \end{bmatrix} \right)$$

Ju

$$S1 = \begin{bmatrix} Err2_1 \\ Err2_2 \\ \vdots \\ Err2_N \end{bmatrix} \begin{bmatrix} w_3^6 & w_4^6 & w_5^6 \end{bmatrix}, \quad Diff1 = \begin{bmatrix} diff_1^3 & diff_1 & diff_1 \\ diff_2^3 & diff_2^4 & diff_2^5 \\ \\ diff_N^3 & diff_N^4 & diff_N^5 \end{bmatrix}$$

$$Err1 = \left(\begin{bmatrix} Err2_1 \\ Err2_2 \\ \vdots \\ Err2_N \end{bmatrix} \begin{bmatrix} w_3^6 & w_4^6 & w_5^6 \end{bmatrix} \right) * \begin{bmatrix} diff_1^3 & diff_1 & diff_1 \\ diff_2^3 & diff_2^4 & diff_2^5 \\ \\ diff_N^3 & diff_N^4 & diff_N^5 \end{bmatrix} = \delta 1 * Diff1$$

สมการ Gradient Descen ของ W1 จะอยู่ในรูป Matrix นี้นะ

$$\boxed{W1 = W1 + \frac{\alpha}{N} A0^T Err1}$$

④ $B_1 = B_1 - \alpha \frac{\partial Error}{\partial B_1}$

สมการ Gradient Descen ของ B1 จะอยู่ในรูปนี้นะ

$$[b^3 \ b^4 \ b^5] = [b^3 \ b^4 \ b^5] - \alpha \left[\frac{\partial Error}{\partial b^3} \ \frac{\partial Error}{\partial b^4} \ \frac{\partial Error}{\partial b^5} \right]$$

พิจารณา $\frac{\partial Error}{\partial b^q}$ เมื่อ $q \in \{3,4,5\}$

$$b^q = b^q - \alpha \frac{\partial Error}{\partial b^q}$$

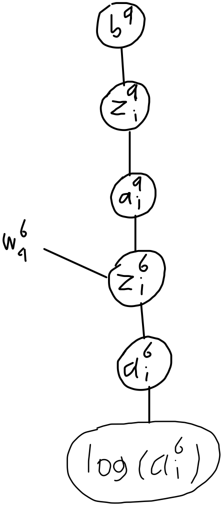
$$= b^q - \alpha \frac{\partial \frac{\sum_{i=1}^{N} Error_i}{N}}{\partial b^q}$$

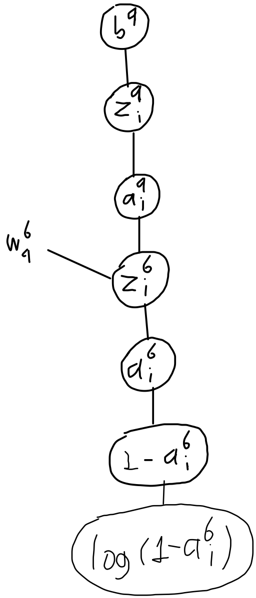
$$= b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial Error_i}{\partial b^q}$$

$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \left(\frac{\partial t_i \log(a_i^6)}{\partial b^q} + \frac{(1-t_i)\log(1-a_i^6)}{\partial b^q} \right)$$

chain rule

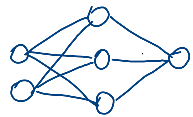
$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial\left(t_i \log(a_i^6) + (1-t_i)\log(1-a_i^6)\right)}{\partial b^q}$$





$$b^9 = b^9 + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{\partial t_i \log(a_i^6)}{\partial b^9} + \frac{(1-t_i)\log(1-a_i^6)}{\partial b^9}\right)$$

$a \in \{3, 4, 5\}$



$b^6 \to z^6$

$b^9 \to z_i^9$

$$\frac{\partial \log(x)}{\partial x} = \frac{1}{x}$$

$$b^9 = b^9 + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{\partial t_i \log(a_i^9)}{b^9} + \frac{(1-t_i)\log(1-a_i^9)}{b^9}\right)$$

$$b^9 = b^9 + \frac{\alpha}{N}\sum_{i=1}^{N}\Bigg(t_i \frac{\partial \log(a_i^6)}{\partial b^9}\frac{\partial a_i^6}{\partial z_i^6}\frac{\partial z_i^6}{\partial a^6}\frac{\partial a_i^9}{\partial z_i^9}\frac{\partial z_i^9}{\partial b^9}$$

$$+ (1-t_i)\frac{\partial \log(1-a_i^6)}{\partial(1-a_i^6)}\frac{\partial(1-a_i^6)}{\partial a_i^6}\frac{\partial a_i^6}{\partial z_i^6}\frac{\partial z_i^6}{\partial a_i^6}\frac{\partial a_i^9}{\partial z_i^9}\frac{\partial z_i^9}{\partial b^9}$$

$$b^9 = b^9 + \frac{\alpha}{N}\sum_{i=1}^{N}\left(t_i\left(\frac{1}{a_i^6}\right)\frac{\partial a_i^6}{\partial z_i^6} w_9^6 \frac{\partial a_i^4}{\partial z_i^9}(1) + (1-t_i)\left(\frac{1}{1-a_i^6}\right)(-1)\frac{\partial a_i^6}{\partial z_i^6} w_9^6 \frac{\partial a_i^4}{\partial z_i^9}(1)\right)$$

$$b^9 = b^9 + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i}{a_i^6}\underline{\frac{\partial a_i^6}{\partial z_i^6}} w_9^6 \underline{\frac{\partial a_i^4}{\partial z_i^9}} - \left(\frac{1-t_i}{1-a_i^6}\right)\underline{\frac{\partial a_i^6}{\partial z_i^6}} w_9^6 \underline{\frac{\partial a_i^4}{\partial z_i^9}}\right)$$

$$b^9 = b^9 + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i}{a_i^6}\, diff_i^6\, w_9^6\, diff_i^9 - \left(\frac{1-t_i}{1-a_i^6}\right) diff_i^6\, w_9^6\, diff_i^9\right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i}{a_i^6}\, diff_i^6\, w_q^6\, diff_i^q - \frac{1-t}{1-a_i^6}\, diff_i^6\, w_q^6\, diff_i^q\right)$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i}{a_i^6} - \frac{1-t}{1-a_i^6}\right) diff_i^6\, w_q^6\, diff_i^q$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i(1-a_i^q) - (1-t_i)a_i^q}{a_i^q(1-a_i^q)}\right) diff_i^6\, w_q^6\, diff_i^q$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i - t_i a_i^q - a_i^q + t_i a_i^q}{a_i^q(1-a_i^q)}\right) diff_i^6\, w_q^6\, diff_i^q$$

$$w_p^q = w_p^q + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t_i - a_i^q}{a_i^q(1-a_i^q)}\right) diff_i^6\, w_q^6\, diff_i^q$$

ในที่นี้ให้ Activation functionชั้นสุดท้ายเป็น simoid ดังนั้น..

$$diff_i^6 = a_i^6(1-a_i^6)$$

$$b^9 = b^9 + \frac{\alpha}{N}\sum_{i=1}^{N}\left(\frac{t-a_i^9}{a_i^9(1-a_i^9)}\right)\underline{a_i^9(1-a_i^9)}w_9^6 diff_i^9$$

$$b^9 = b^9 + \frac{\alpha}{N}\sum_{i=1}^{N}(t_i-a_i^9)w_9^6 diff_i^9$$

เนื่องจากผ่าน simoid ทำให้ Diff2 = 1

$$\delta 2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} \quad ; \quad Diff2 = 1$$

$$Err2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} * 1 = \delta 2 * Diff2$$

ได้

$$Err2_i = (t_i - a_i^6) * diff_i^6$$

$$b^9 = b^9 + \frac{\alpha}{N}\sum_{i=1}^{N} Err2_i \, w_9^6 diff_i^9$$

$$b^q = b^q + \frac{\alpha}{N} \sum_{i=1}^{N} Err2_i \; w_q^6 \, diff_i^q$$

สามารถเขียนในรูป Matrix ได้ดังนี้

$$[b^3 \; b^4 \; b^5] = [b^3 \; b^4 \; b^5] + \frac{\alpha}{N}\left[\sum_{i=1}^{N} Err2_i w_3^6 \, diff_i^3 \quad \sum_{i=1}^{N} Err2_i w_4^6 \, diff_i^4 \quad \sum_{i=1}^{N} Err2_i w_5^6 \, diff_i^5 \right]$$

$$[b^3 \; b^4 \; b^5] = [b^3 \; b^4 \; b^5] + \frac{\alpha}{N}[1 \; 1 \ldots 1] \begin{bmatrix} \sum_{i=1}^{N} Err2_1 w_3^6 \, diff_1^3 & \sum_{i=1}^{N} Err2_1 w_4^6 \, diff_1^4 & \sum_{i=1}^{N} Err2_1 w_5^6 \, diff_1^5 \\ \sum_{i=1}^{N} Err2_2 w_3^6 \, diff_2^3 & \sum_{i=1}^{N} Err2_2 w_4^6 \, diff_2^4 & \sum_{i=1}^{N} Err2_2 w_5^6 \, diff_2^5 \\ & \vdots & \\ \sum_{i=1}^{N} Err2_N w_3^6 \, diff_N^3 & \sum_{i=1}^{N} Err2_N w_4^6 \, diff_N^4 & \sum_{i=1}^{N} Err2_N w_5^6 \, diff_N^5 \end{bmatrix}$$

$$[b^3\ b^4\ b^5] = [b^3\ b^4\ b^5] + \frac{\alpha}{N}[1\,1\ldots 1]\left(\begin{bmatrix} \bar{E}rr2_1 w_3^6 & \bar{E}rr2_1 w_4^6 & \bar{E}rr2_1 w_5^6 \\ \bar{E}rr2_2 w_3^6 & \bar{E}rr2_2 w_4^6 & \bar{E}rr2_2 w_5^6 \\ & \vdots & \\ \bar{E}rr2_N w_3^6 & \bar{E}rr2_N w_4^6 & \bar{E}rr2_N w_5^6 \end{bmatrix}\right.$$

$$\left. * \begin{bmatrix} diff_1^3 & diff_1^4 & diff_1^5 \\ diff_2^3 & diff_2^4 & diff_2^5 \\ & & \\ diff_N^3 & diff_N^4 & diff_N^5 \end{bmatrix}\right)$$

$$
\begin{bmatrix} b^3 & b^4 & b^5 \end{bmatrix} = \begin{bmatrix} b^3 & b^4 & b^5 \end{bmatrix} + \frac{\alpha}{N} \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \left(\left(\begin{bmatrix} Err2_1 \\ Err2_2 \\ \vdots \\ Err2_N \end{bmatrix} \begin{bmatrix} w^6_3 & w^6_4 & w^6_5 \end{bmatrix} \right) * \begin{bmatrix} diff_1^3 & diff_1^4 & diff_1^5 \\ diff_2^3 & diff_2^4 & diff_2^5 \\ \vdots & \vdots & \vdots \\ diff_N^3 & diff_N^4 & diff_N^5 \end{bmatrix} \right)
$$

Tú

$$\delta 1 = \begin{bmatrix} Err2_1 \\ Err2_2 \\ \vdots \\ Err2_N \end{bmatrix} \begin{bmatrix} w_3^6 & w_4^6 & w_5^6 \end{bmatrix} ; \quad Diff1 = \begin{bmatrix} diff_1^3 & diff_1^4 & diff_1^5 \\ diff_2^3 & diff_2^4 & diff_2^5 \\ & \vdots & \\ diff_N^3 & diff_N^4 & diff_N^5 \end{bmatrix}$$

$$Err1 = \left(\begin{bmatrix} Err2_1 \\ Err2_2 \\ \vdots \\ Err2_N \end{bmatrix} \begin{bmatrix} w_3^6 & w_4^6 & w_5^6 \end{bmatrix} \right) * \begin{bmatrix} diff_1^3 & diff_1^4 & diff_1^5 \\ diff_2^3 & diff_2^4 & diff_2^5 \\ & \vdots & \\ diff_N^3 & diff_N^4 & diff_N^5 \end{bmatrix}$$

$$= \delta 1 * Diff1$$

สมการ Gradient Descen ของ B1 จะอยู่ในรูป Matrix นี้นะ

$$\boxed{B1 = B1 + \frac{\alpha}{N} Err1.sum(axis=0)}$$

จากการพิจารณาสมการ Gradient Descen ทั้ง 4 สมการ ที่ได้ดังนี้

$$W2 = W2 + \frac{\alpha}{N} A1^{T} Err2$$

$$B2 = B2 + \frac{\alpha}{N} Err2.sum(axis=0)$$

$$W1 = W1 + \frac{\alpha}{N} A0^{T} Err1$$

$$B1 = B1 + \frac{\alpha}{N} Err1.sum(axis=0)$$

โดยที่

$$\delta 2 = T - A2$$

$$Err2 = \delta 2 * Diff2$$

$$\delta 1 = Err2\, W2^{T}$$

$$Err1 = \delta 1 * Diff1$$

$$\delta 2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} \rightarrow T - A2$$

T = t = y = taget (คำตอบ)

A = a = โหนด

ซึ่งจาก Pattern ที่ได้นี้ สามารถสรุปได้ว่า

$$S(m) = T - A(m)$$

เมื่อ m คือ Layer สุดท้าย(output)

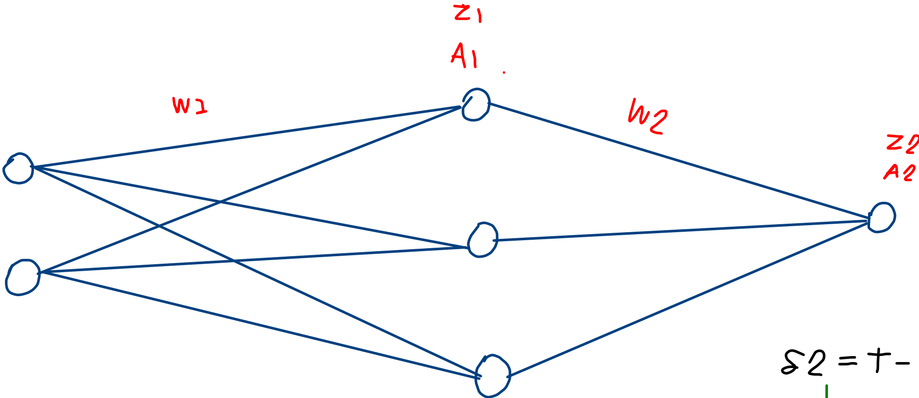
$T = t = y =$ taget (ค่าตอบ)

$$S(m) = Err(m+1)\,W(m+1)^T$$

เมื่อ m ไม่ใช่ Layer สุดท้าย

Backpropagation -> เอาตัวก่อนหน้า มาคำนวณ
ปรับ Weight หา Error

$$Err(m) = S(m) * Diff(m)$$



$$\delta 1 = Err2\, W2^T$$
$$\downarrow$$
$$Diff1$$
$$\downarrow$$
$$Err1 = \delta 1 * Diff1$$

$$W1 = W1 + \frac{\alpha}{N} A0^T \, Err1$$

$$B1 = B1 + \frac{\alpha}{N} Err1.sum(axis=0)$$

$$\delta 2 = T - A2$$
$$\downarrow$$
$$Diff2$$
$$\downarrow$$
$$Err2 = \delta 2 * Diff$$

$$W2 = W2 + \frac{\alpha}{N} A1^T \, Err2$$

$$B2 = B2 + \frac{\alpha}{N} Err2.sum(axis0)$$