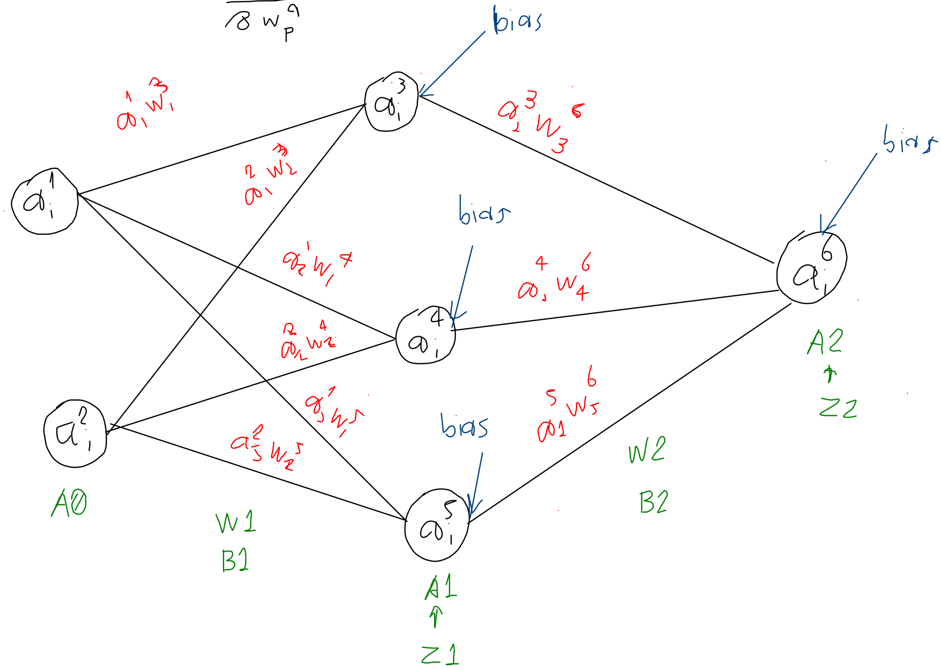
พิจารณาสมหการ Gradient Descent $\frac{\partial Error}{\partial w_p^q}$ และ Backpropagation สำหรับ Regression



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

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

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

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

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

$$= A0W1 + B1$$

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

a_1^1

a_1^2

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

ส่งจากโหนด q → โหนด p

$$Error = \frac{\sum_{i=1}^{N} (t_i - \hat{t}_i)^2}{N} = \frac{\sum_{i=1}^{N} Error_i}{N}$$

→ Mean Square Error

$$Error_i = (t_i - \hat{t}_i)^2$$

→ $t = y$; $\hat{t} = \hat{y}$

หา Error ชั้นสุดท้าย แล้ว Backpropagation กลับ เพื่อปรับ weight แต่ละรอบ

เราจะพิจารณาสมการ Gradient Descent 4 สมการ ดังนี้

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

หลังไปหน้า

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

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

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

α

1

$$w2 = w2 - \alpha \frac{\partial Error}{\partial w2}$$

Gradient Descent

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

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

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

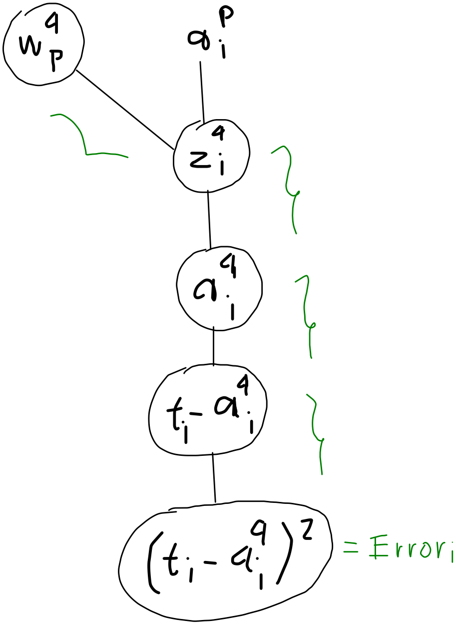
$$= w_p^q - \alpha \frac{\partial \frac{\sum_{i=1}^{N} Error_i}{N}}{\partial w_p^q}$$

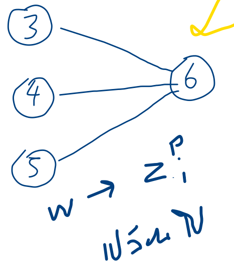
$$= w_p^q - \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial Error_i}{\partial w_p^q}$$

$$w_p^q = w_p^q - \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial (t_i - \hat{t}_i)^2}{\partial w_p^q}$$

chain rule

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





$$w_p^q = w_p^q - \frac{\alpha}{N}\sum_{i=1}^{N} \frac{\partial (t_i - t_i)^2}{\partial w_p^q}$$

$$z_i^q = \sigma(w_3^q a_i^3 + w_4^q a_i^4 + w_5^q a_i^5 + b^q)$$

$$\overline{\partial w_p^q}$$

$$w_p^q = w_p^q - \frac{\alpha}{N}\sum_{i=1}^{N} \frac{\partial (t_i - a_i^q)^2}{\partial (t_i - a_i^q)} \frac{\partial (t_i - a_i^q)}{\partial a_i^q} \frac{\partial a_i^q}{\partial z_i^q} \frac{\partial z_i^q}{\partial w_p^q}$$

$$w_p^q = w_p^q - \frac{\alpha}{N}\sum_{i=1}^{N} 2(t_i - a_i^q)(-1) \frac{\partial a_i^q}{\partial z_i^q} a_i^p$$

ดึง 2 ไว้หน้า
รวมกับ α
เป็นค่าคงที่

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

Activation function

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

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

ดังนั้น สามารถเขียนให้อยู่ในรูป Matrix ได้ดังนี้

$$\begin{bmatrix} w_3^6 \\ w_4^6 \\ w_5^6 \end{bmatrix} = \begin{bmatrix} w_3^6 \\ w_4^6 \\ w_5^6 \end{bmatrix} + \frac{\alpha}{N} \begin{bmatrix} \sum_{i=1}^{N} (t_i - a_i^6)\, diff_i^6\, a_i^3 \\ \sum_{i=1}^{N} (t_i - a_i^6)\, diff_i^6\, a_i^4 \\ \sum_{i=1}^{N} (t_i - a_i^6)\, diff_i^6\, a_i^5 \end{bmatrix}$$

$q \rightarrow$ โหนด 6

$$\begin{bmatrix} w_3^6 \\ w_4^6 \\ w_5^6 \end{bmatrix} = \begin{bmatrix} w_3^6 \\ w_4^6 \\ w_5^6 \end{bmatrix} + \frac{\alpha}{N} \begin{bmatrix} a_1^3 & a_2^3 & \dots & a_N^3 \\ a_1^4 & a_2^4 & \dots & a_N^4 \\ a_1^5 & a_2^5 & \dots & a_N^5 \end{bmatrix} \begin{bmatrix} (t_1 - a_1^6) * diff_1^6 \\ (t_1 - a_2^6) * diff_2^6 \\ \vdots \\ (t_1 - a_3^6) * diff_3^6 \end{bmatrix}$$

$$\begin{bmatrix} w_3^6 w \\ w_4^6 \\ w_5^6 \end{bmatrix} = \begin{bmatrix} w_3^6 \\ w_4^6 \\ w_5^6 \end{bmatrix} + \frac{\alpha}{N} \underbrace{\begin{bmatrix} a_1^3 & a_2^3 & \dots & a_N^3 \\ a_1^4 & a_2^4 & \dots & a_N^4 \\ a_1^5 & a_2^5 & \dots & a_N^5 \end{bmatrix}}_{A1^T} \left(\begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} * \begin{bmatrix} diff_1^6 \\ diff_2^6 \\ \vdots \\ diff_N^6 \end{bmatrix} \right)$$

กำหนดให้

$$\delta 2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} \; ; \quad Diff2 = \begin{bmatrix} diff_1^6 \\ diff_2^6 \\ \vdots \\ diff_N^6 \end{bmatrix}$$

คูณแบบ "*" ตัวต่อตัว

$$Err2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} * \begin{bmatrix} diff_1^6 \\ diff_2^6 \\ \vdots \\ diff_N^6 \end{bmatrix} = \delta 2 * Diff2$$

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

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

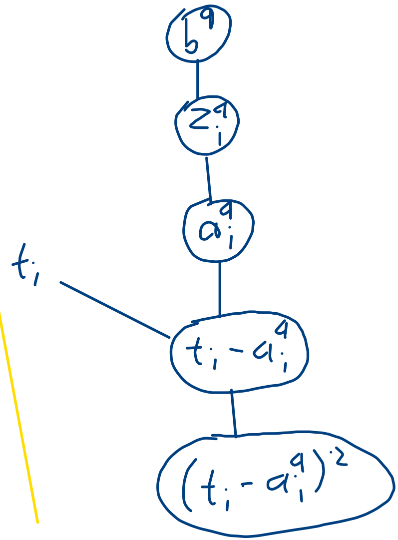
พิจารณา $\frac{\partial Error}{\partial b^q}$ เมื่อ $q \in \{6\}$

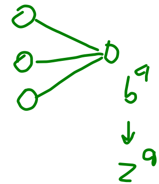
$$b^q = b^q - \alpha \frac{\partial Error}{\partial b^q}$$

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

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

$$b^q = b^q - \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial (t_i - a_i^q)^2}{\partial b^q}$$





$$b^q = b^q - \frac{\alpha}{N}\sum_{i=1}^{N} \frac{\partial(t_i - a_i^q)^2}{\partial(t_i - a_i^q)^2}\frac{\partial(t_i - a_i^q)}{\partial a_i^q}\frac{\partial a_i^q}{\partial z_i^q}\frac{\partial z_i^q}{\partial b^q}$$

$$\frac{\partial z_i^q = \partial(w_3^q a_i^3 + w_4^q a_i^4 + w_5^q a_i^5 + b^q)}{\partial b^q}$$

$$b^q = b^q - \frac{\alpha}{N}\sum_{i=1}^{N} 2(t_i - a_i^q)(-1)\frac{\partial a_i^q}{\partial z_i^q}(1)$$

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

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

สามารถเขียนในรูป Matrix ได้ดังนี้

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

$$[b^6] = [b^6] + \frac{\alpha}{N}[1\ 1\ \cdots\ 1]\begin{bmatrix} (t_1 - a_1^6) * diff_1^6 \\ (t_1 - a_1^6) * diff_2^6 \\ \vdots \\ (t_1 - a_1^6) * diff_N^6 \end{bmatrix}$$

ให้

$$\delta 2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} ; \quad Diff2 = \begin{bmatrix} diff_1^6 \\ diff_2^6 \\ \vdots \\ diff_N^6 \end{bmatrix}$$

$$Err2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} * \begin{bmatrix} diff_1^6 \\ diff_2^6 \\ \vdots \\ diff_N^6 \end{bmatrix} = \delta 2 * Diff2$$

ดังนั้นสมการ Gradient Descent ของ B2 ที่ได้

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

รวมแนวคอลัมน์

$$[0\ 0\ 0]$$

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

สมการ Gradient Descen ของ W1 จะอยู่ในรูปนี้นะ

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

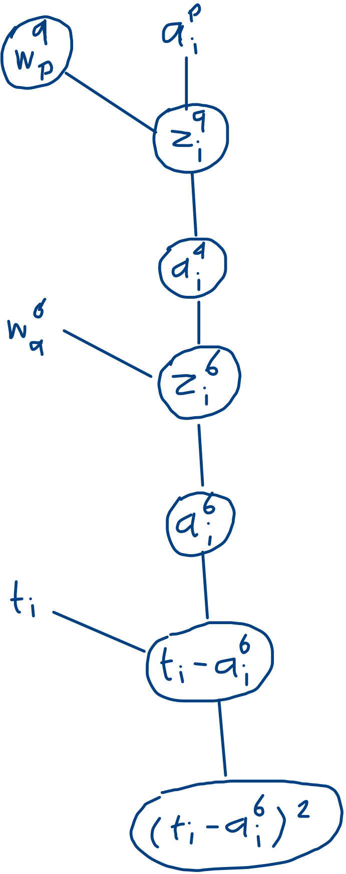
พิจารณา $\frac{\partial Error}{\partial w_p^q}$ เมื่อ $p \in \{1,2\}$ และ $q \in \{3,4,5\}$

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

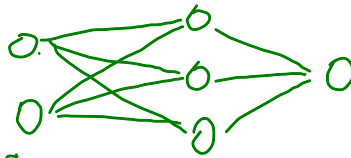
$$w_p^q = w_p^q - \alpha \frac{\partial \frac{\sum_{i=1}^{N} Error_i}{N}}{\partial w_p^q}$$

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

$$w_p^q = w_p^q - \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial (t_i - a_i^6)^2}{\partial w_p^q}$$



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



$q \in \{6\}$

$z_i^q \quad (z_i^6)$

$w_p^q \longleftrightarrow z_i^q \longleftrightarrow$

ทำให้
z เปลี่ยน 2 ชั้น

$$w_p^q = w_p^q - \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial (t_i - a_i^6)^2}{\partial w_p^q}$$

$$w_p^q = w_p^q - \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial (t_i - a_i^6)^2}{\partial (t_i - a_i^6)} \frac{\partial (t_i - a_i^6)}{\partial a_i^6} \frac{\partial a_i^6}{\partial z_i^6} \frac{\partial z_i^6}{\partial a_i^6} \frac{\partial a_i^q}{\partial z_i^q} \frac{\partial z_i^q}{\partial w_p^q}$$

$$\frac{z_i^q = \partial (w_3^q a_i^3 + w_4^q a_i^4 + w_5^q a_i^5 + b^q)}{\partial w_p^q}$$

$$w_p^q = w_p^q - \frac{\alpha}{N} \sum_{i=1}^{N} 2 (t_i - a_i^6) (-1) \frac{\partial a_i^6}{\partial z_i^6} w_q^6 \frac{\partial a_i^q}{\partial z_i^q} a_i^p$$

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

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

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

<u>เนื่องจาก</u>

$$\delta 2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} \; ; \quad Diff2 = \begin{bmatrix} diff_1^6 \\ diff_2^6 \\ \vdots \\ diff_N^6 \end{bmatrix}$$

$$Err2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} * \begin{bmatrix} diff_1^6 \\ diff_2^6 \\ \vdots \\ diff_N^6 \end{bmatrix} = \delta 2 * Diff2 \Rightarrow$$

ใน

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

จะได้

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

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

รวมกัน $a = \{3, 4, 5\}$

$n = a_i^1$

$u = a_i^2$

สามารถเขียนในรูป Matrix ได้ดังนี้

$$\begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} = \begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} + \frac{\alpha}{N} \begin{bmatrix} \sum_{i=1}^{N} Err2_i\, w_3^6\, diff_i^3\, a_i^1 & \sum_{i=1}^{N} Err2_i\, w_3^6\, diff_i^4\, a_i^1 & \sum_{i=1}^{N} Err2_i\, w_3^6\, diff_i^5\, a_i^1 \\ \sum_{i=1}^{N} Err2_i\, w_3^6\, diff_i^3\, a_i^2 & \sum_{i=1}^{N} Err2_i\, w_3^6\, diff_i^4\, a_i^2 & \sum_{i=1}^{N} Err2_i\, w_3^6\, diff_i^5\, a_i^2 \end{bmatrix} \begin{matrix} a^1 \\ a^2 \end{matrix}$$

$$\begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} = \begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} + \frac{\alpha}{N} \begin{bmatrix} a_1^1 & a_2^1 & \dots & a_N^1 \\ a_1^2 & a_2^2 & \dots & a_N^2 \end{bmatrix} \begin{bmatrix} Err2_1\, w_3^6\, diff_1^3 & Err2_1\, w_4^6\, diff_1^4 & Err2_1\, w_5^6\, diff_1^5 \\ Err2_1\, w_3^6\, diff_2^3 & Err2_1\, w_4^6\, diff_2^4 & Err2_1\, w_5^6\, diff_2^5 \\ \vdots & \vdots & \vdots \\ Err2_1\, w_3^6\, diff_N^3 & Err2_1\, w_4^6\, diff_N^4 & Err2_1\, w_5^6\, diff_N^5 \end{bmatrix}$$

$$\begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} = \begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} + \frac{\alpha}{N} \begin{bmatrix} a_1^1 & a_2^1 & \dots & a_N^1 \\ a_1^2 & a_2^2 & \dots & a_N^2 \end{bmatrix} \left(\begin{bmatrix} Err2_1\, w_3^6 & Err2_1\, w_4^6 & Err2\, w_5^6 d \\ Err2_1\, w_3^6 & Err2_1\, w_4^6 & Err2_1\, w_5^6 d \\ Err2_1\, w_3^6 & Err2_1\, w_4^6 & Err2_1\, w_5^6 d \end{bmatrix} * \begin{bmatrix} diff_1^3 & diff_1^1 & dff_1 \\ diff_2^3 & diff_2^4 & dff_2^5 \\ diff_N^3 & diff_N^4 & dff_N^5 \end{bmatrix} \right)$$

$$\underbrace{\begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix}}_{W1} = \begin{bmatrix} w_1^3 & w_1^4 & w_1^5 \\ w_2^3 & w_2^4 & w_2^5 \end{bmatrix} + \frac{\alpha}{N} \underbrace{\begin{bmatrix} a_1^1 & a_2^1 & \dots & a_N^1 \\ a_1^2 & a_2^2 & \dots & a_N^2 \end{bmatrix}}_{A0^T} \left(\left(\begin{bmatrix} Err2_1 \\ Err2_2 \\ \vdots \\ Err2_N \end{bmatrix} \underbrace{\begin{bmatrix} w_3^6 & w_4^6 & w_5^6 \end{bmatrix}}_{W_2^T} \right) * \begin{bmatrix} diff_1^3 & diff_1 & dff_1 \\ diff_2^3 & diff_2^4 & dff_2^5 \\ \\ diff_N^3 & diff_N^4 & diff_N^5 \end{bmatrix} \right)$$

Đu

$$S1 = \begin{bmatrix} Err2_1 \\ Err2_2 \\ \vdots \\ Err2_N \end{bmatrix} \begin{bmatrix} w_3^6 & w_4^6 & w_5^6 \end{bmatrix}, \quad Diff1 = \begin{bmatrix} diff_1^3 & diff_1 & dff_1 \\ diff_2^3 & diff_2^4 & dff_2^5 \\ \\ diff_N^3 & diff_N^4 & diff_N^5 \end{bmatrix}$$

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

สมการ Gradient Descen ของ W1 จะอยู่ในรูป Matrix นี้นะ

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

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

สมการ Gradient Descen ของ B1 จะอยู่ในรูปนี้นะ

$$\begin{bmatrix} b^3 & b^4 & b^5 \end{bmatrix} = \begin{bmatrix} b^3 & b^4 & b^5 \end{bmatrix} - \alpha \begin{bmatrix} \dfrac{\partial Error}{\partial b^3} & \dfrac{\partial Error}{\partial b^4} & \dfrac{\partial Error}{\partial b^5} \end{bmatrix}$$

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

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

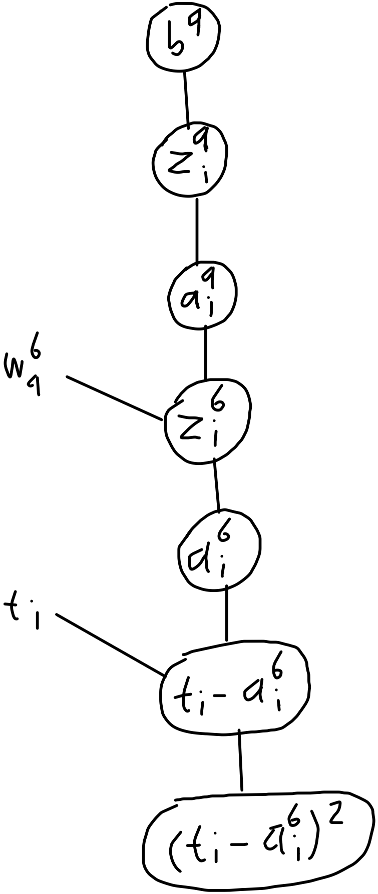
$$= b^q - \alpha \frac{\partial \dfrac{\sum_{i=1}^{N} Error_i}{N}}{\partial b^q}$$

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

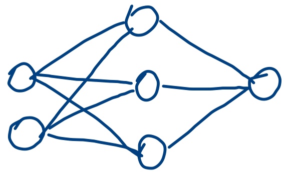
$$b^q = b^q - \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial (t_i - a^q)^2}{\partial b^q}$$

chain rule

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



$= Error_i$



$b^6 \to z^6$

$b^9 \to z_i^9$

$$b^9 = b^9 - \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial (t_i - a^9)^2}{\partial b^9}$$

$$b^9 = b^9 - \frac{\alpha}{N} \sum_{i=1}^{N} \frac{\partial (t_i - a_i^6)^2}{\partial (t_i - a_i^6)} \frac{\partial (t_i - a_i^6)}{\partial a_i^6} \frac{\partial a_i^6}{\partial z_i^6} \frac{\partial z^6}{\partial a_i^9} \frac{\partial a_i^9}{\partial z_i^9} \frac{\partial z_i^9}{\partial b^9}$$

$$b^9 = b^9 - \frac{\alpha}{N} \sum_{i=1}^{N} 2(t_i - a_i^6)(-1) \frac{\partial a_i^6}{\partial z_i^6} w_9^6 \frac{\partial a_i^9}{\partial z_i^9} (1)$$

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

เนื่องจาก

$$\delta 2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} ; \quad Diff2 = \begin{bmatrix} diff_1^6 \\ diff_2^6 \\ \vdots \\ diff_N^6 \end{bmatrix}$$

$$Err2 = \begin{bmatrix} t_1 - a_1^6 \\ t_2 - a_2^6 \\ \vdots \\ t_N - a_N^6 \end{bmatrix} * \begin{bmatrix} diff_1^6 \\ diff_2^6 \\ \vdots \\ diff_N^6 \end{bmatrix} = \delta 2 * Diff2$$

ใช้

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

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

สามารถเขียนในรูป Matrix ได้ดังนี้

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

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

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

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

$$[b^3\ b^4\ b^5] = [b^3\ b^4\ b^5] + \frac{\alpha}{N}[1\,1\ldots1]\left(\left(\begin{bmatrix} Err2_1 \\ Err2_2 \\ \\ Err2_N \end{bmatrix}[w_3^6\ \ w_4^6\ \ w_5^6]\right) * \begin{bmatrix} diff_1^3 & diff_1^4 & diff_1^5 \\ diff_2^3 & diff_2^4 & diff_2^5 \\ \\ diff_N^3 & diff_N^4 & diff_N^5 \end{bmatrix}\right)$$

Ví

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

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

$$= \delta 1 * Diff1$$

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

$$\delta 2 = T - A2$$

$$Err2 = \delta 2 * Diff2$$

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

$$Err1 = \delta 1 * Diff1$$

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

$T = t = y = target$ (คำตอบ)

$A = a =$ โมเดล

ซึ่งจาก Pattern ที่ได้นี้ สามารถสรุปได้ว่า

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

เมื่อ m คือ Layer สุดท้าย(output)

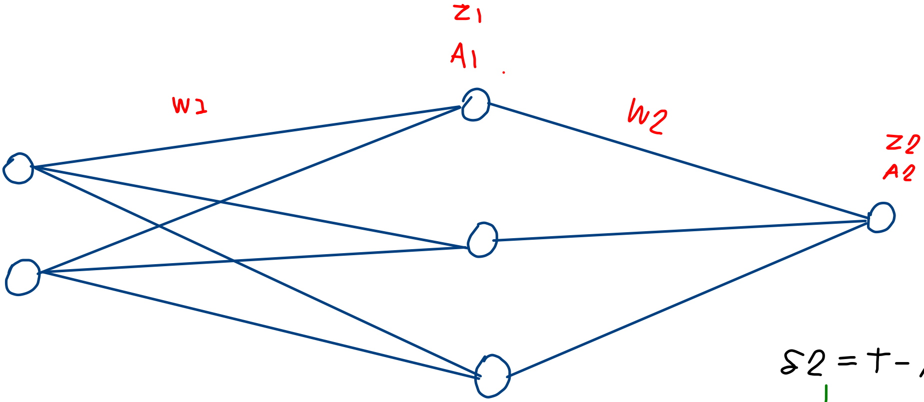
$T = t = y = \text{taget}$ (คำตอบ)

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

เมื่อ m ไม่ใช่ Layer สุดท้าย

Backpropagation -> เอาตัวก่อนหน้า มาคำนวณ
ปรับ Weight หา Error

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



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

$$W2 = W2 + \frac{\alpha}{N} A1^T\, Err\,2$$

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

$$\delta 1 = Err\,2\, W2^T$$
$$\downarrow$$
$$Diff\,1$$
$$\downarrow$$
$$Err1 = \delta 1 * Diff\,1$$

$$W1 = W1 + \frac{\alpha}{N} A0^T\, Err\,1$$

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