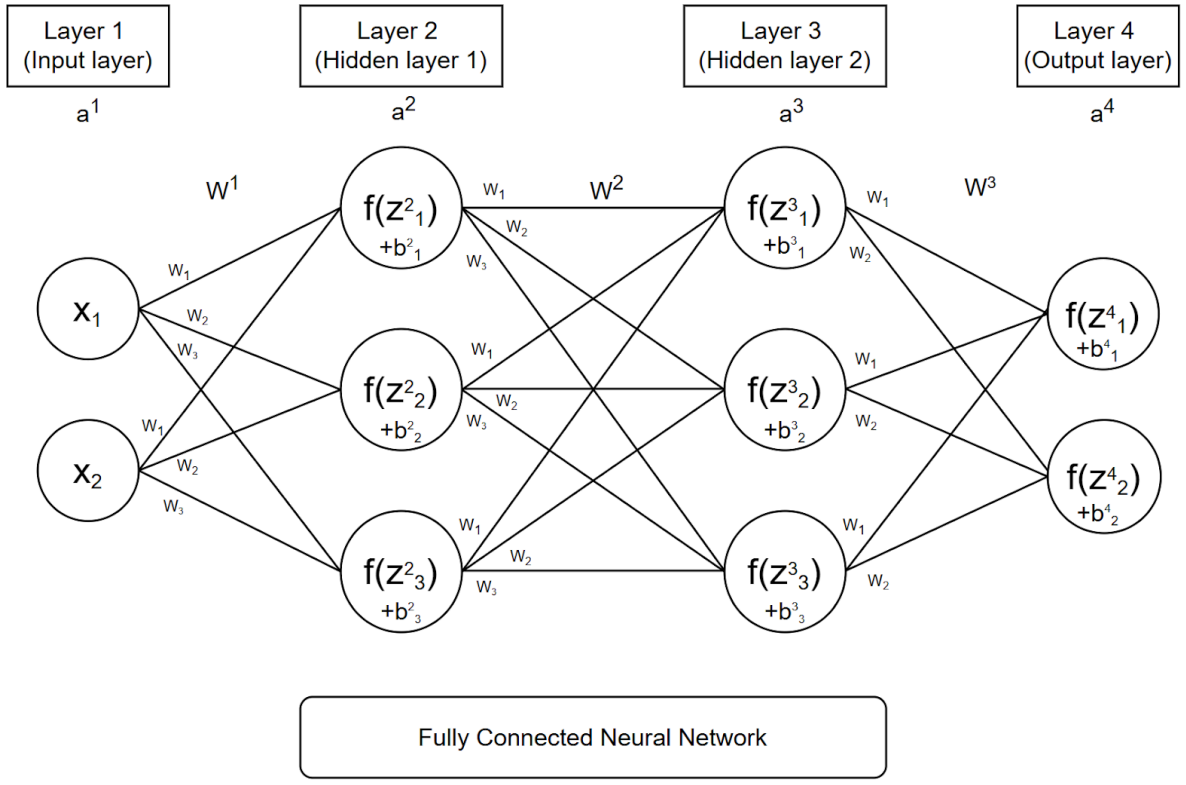
นายพรชัย คำแดง
รหัส 6305000942

ทฤษฎีของ Fully Connected Layers ใน Neural Networks



รูปตัวอย่างโครงสร้างของ Neural Network ที่ใช้ศึกษาการทำงาน

สมการของนิวรอน (Neuron Equation)

$$z = \sum_{i=1}^{n} w_i x_i + b$$

$$y = f(z)$$

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

สมการของนิวรอน (Neuron Equation) แบบเมทริกซ์
เมื่อ W, x และ b อยู่ในรูปของ เมทริกซ์จะได้สมการดังนี้

$$y = f(Wx + b)$$

Wx คือผลคูณเมทริกซ์ และบวกด้วยเมทริกซ์ b แล้วค่อยหาฟังก์ชันกระตุ้นของเมทริกซ์ลัพธ์ (z) ก็จะได้คำตอบของสมการนิวรอน

นายพรชัย คำแดง
รหัส 6305000942

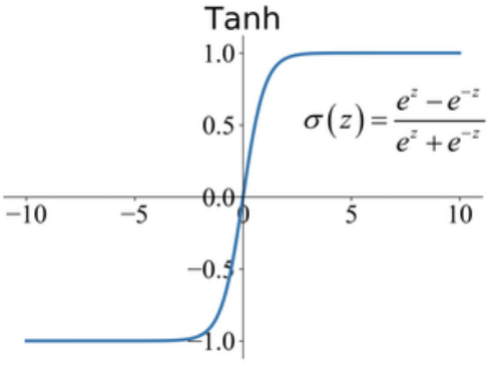
ขั้นตอนการคำนวณของ Fully Connected Layers ใน Neural Networks

1) กำหนดโครงสร้างของ Neural Network

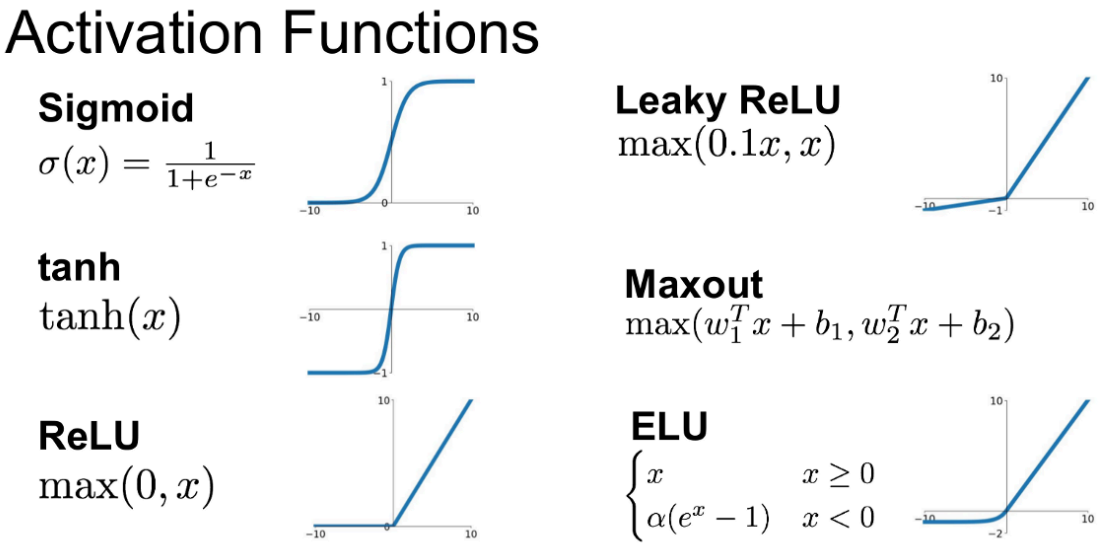
- **Layer 1 (Input layer):** 2 นิวรอน
- **Layer 2 (Hidden layer 1):** 3 นิวรอน
- **Layer 3 (Hidden layer 2):** 3 นิวรอน
- **Layer 4 (Output layer):** 2 นิวรอน
- ฟังก์ชันกระตุ้น f(x)=tanh(x) = tanh(x)

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

รูปแบบของ ฟังก์ชันกระตุ้น Tanh พบว่าคำตอบจะอยู่ในช่วง [-1,1]

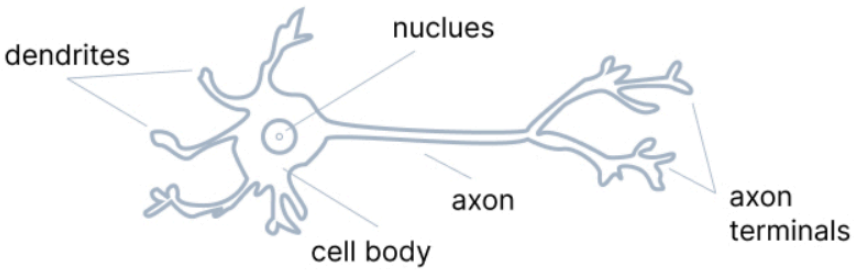


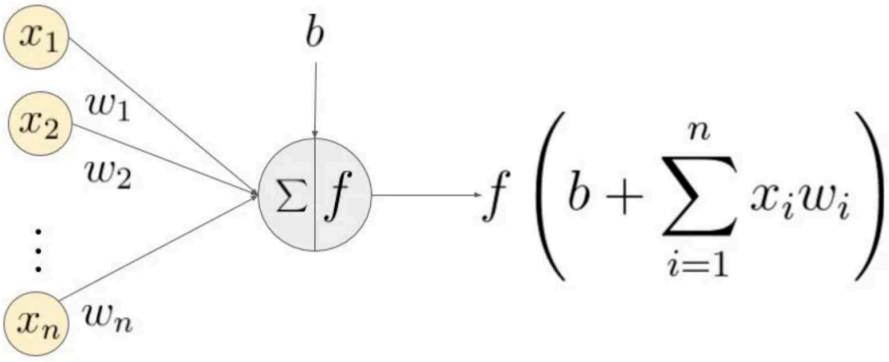
รูปแบบของ ฟังก์ชันกระตุ้นประเภทอื่นๆ



นายพรชัย คำแดง
รหัส 6305000942

เปรียบเทียบเซลล์ประสาททางชีววิทยา และ เซลล์ประสาทเทียม





2) Forward Propagation (แบบเมทริกซ์)

Layer 1 → Layer 2

คำนวณค่า pre-activation:

$$\mathbf{Z}^{(2)} = \mathbf{X}\mathbf{W}^{(1)} + \mathbf{b}^{(1)}$$

อินพุต X เป็นเมทริกซ์ของ 3 ตัวอย่าง (samples) และ 2 ฟีเจอร์ (features):

$$\mathbf{X} = \begin{bmatrix} t_1 & x_1 \\ t_2 & x_2 \\ t_3 & x_3 \end{bmatrix}$$

จากเดิมเป็นเมทริกซ์ขนาด [2 x 3] ทำการ Transpose ให้มีขนาดเป็น [3 x 2] ก่อนนำไปคำนวณสู่นิวรอนใน Layer ที่ 2

นายพรชัย คำแดง
รหัส 6305000942

คำนวณ Weighted Sum $Z^{(2)}$

Layer 2 มี 3 **นิวรอน** และต้องมีน้ำหนักเชื่อมโยงจาก 2 **ฟีเจอร์ของ** Layer 1 ดังนั้น Weight Matrix $W^{(1)}$ **มีขนาด** $(2,3)$ (2 input → 3 neurons):

$$\mathbf{W}^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \end{bmatrix}$$

Bias Vector $\mathbf{b}^{(1)}$ **มีขนาด** $(1,3)$ (หนึ่ง bias ต่อหนึ่งนิวรอนใน Layer 2):

$$\mathbf{b}^{(1)} = \begin{bmatrix} b_1^{(1)} & b_2^{(1)} & b_3^{(1)} \end{bmatrix}$$

คำนวณ Weighted Sum:

$$\mathbf{Z}^{(2)} = \mathbf{X}\mathbf{W}^{(1)} + \mathbf{b}^{(1)}$$

เมื่อแทนค่าเมทริกซ์:

$$\begin{bmatrix} t_1 & x_1 \\ t_2 & x_2 \\ t_3 & x_3 \end{bmatrix} \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(1)} & b_2^{(1)} & b_3^{(1)} \end{bmatrix}$$

ทำเมทริกซ์ดอทโปรดักต์:

$$\mathbf{Z}^{(2)} = \begin{bmatrix} t_1 w_{11}^{(1)} + x_1 w_{21}^{(1)} & t_1 w_{12}^{(1)} + x_1 w_{22}^{(1)} & t_1 w_{13}^{(1)} + x_1 w_{23}^{(1)} \\ t_2 w_{11}^{(1)} + x_2 w_{21}^{(1)} & t_2 w_{12}^{(1)} + x_2 w_{22}^{(1)} & t_2 w_{13}^{(1)} + x_2 w_{23}^{(1)} \\ t_3 w_{11}^{(1)} + x_3 w_{21}^{(1)} & t_3 w_{12}^{(1)} + x_3 w_{22}^{(1)} & t_3 w_{13}^{(1)} + x_3 w_{23}^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(1)} & b_2^{(1)} & b_3^{(1)} \end{bmatrix}$$

ผลลัพธ์:

$$\mathbf{Z}^{(2)} = \begin{bmatrix} t_1 w_{11}^{(1)} + x_1 w_{21}^{(1)} + b_1^{(1)} & t_1 w_{12}^{(1)} + x_1 w_{22}^{(1)} + b_2^{(1)} & t_1 w_{13}^{(1)} + x_1 w_{23}^{(1)} + b_3^{(1)} \\ t_2 w_{11}^{(1)} + x_2 w_{21}^{(1)} + b_1^{(1)} & t_2 w_{12}^{(1)} + x_2 w_{22}^{(1)} + b_2^{(1)} & t_2 w_{13}^{(1)} + x_2 w_{23}^{(1)} + b_3^{(1)} \\ t_3 w_{11}^{(1)} + x_3 w_{21}^{(1)} + b_1^{(1)} & t_3 w_{12}^{(1)} + x_3 w_{22}^{(1)} + b_2^{(1)} & t_3 w_{13}^{(1)} + x_3 w_{23}^{(1)} + b_3^{(1)} \end{bmatrix}$$

คำนวณค่า activation โดยใช้ฟังก์ชันกระตุ้น tanh:

$$A^{(2)} = \tanh(Z^{(2)})$$

$$\mathbf{A}^{(2)} = \begin{bmatrix} \tanh(z_{11}^{(2)}) & \tanh(z_{12}^{(2)}) & \tanh(z_{13}^{(2)}) \\ \tanh(z_{21}^{(2)}) & \tanh(z_{22}^{(2)}) & \tanh(z_{23}^{(2)}) \\ \tanh(z_{31}^{(2)}) & \tanh(z_{32}^{(2)}) & \tanh(z_{33}^{(2)}) \end{bmatrix}$$

นายพรชัย คำแดง
รหัส 6305000942

Forward Propagation: Layer 2 → Layer 3

คำนวณค่า pre-activation:

$$\mathbf{Z}^{(3)} = \mathbf{A}^{(2)}\mathbf{W}^{(2)} + \mathbf{b}^{(2)}$$

อินพุต ของ Layer 3 คือ $A^{(2)}$ ซึ่งเป็น เมทริกซ์ขนาด [3 x 3] เนื่องจากมี 3 ตัวอย่าง (samples) และ 3 นิวรอนจาก Layer 2

คำนวณ **W**eighted Sum $Z^{(3)}$

Layer 3 มี 3 **นิวรอน**
ดังนั้น Weight Matrix $W^{(2)}$ **มีขนาด** $(3, 3)$

$$\mathbf{W}^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} \end{bmatrix}$$

Bias Vector $\mathbf{b}^{(2)}$ **มีขนาด** $(1, 3)$:

$$\mathbf{b}^{(2)} = \begin{bmatrix} b_1^{(2)} & b_2^{(2)} & b_3^{(2)} \end{bmatrix}$$

คำนวณ Weighted Sum:

$$\mathbf{Z}^{(3)} = \mathbf{A}^{(2)}\mathbf{W}^{(2)} + \mathbf{b}^{(2)}$$

แทนค่าเมทริกซ์:

$$\begin{bmatrix} a_{11}^{(2)} & a_{12}^{(2)} & a_{13}^{(2)} \\ a_{21}^{(2)} & a_{22}^{(2)} & a_{23}^{(2)} \\ a_{31}^{(2)} & a_{32}^{(2)} & a_{33}^{(2)} \end{bmatrix} \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} & b_2^{(2)} & b_3^{(2)} \end{bmatrix}$$

คำนวณเดอทโปรดักต์:

$$\mathbf{Z}^{(3)} = \begin{bmatrix} a_{11}^{(2)}w_{11}^{(2)} + a_{12}^{(2)}w_{21}^{(2)} + a_{13}^{(2)}w_{31}^{(2)} & a_{11}^{(2)}w_{12}^{(2)} + a_{12}^{(2)}w_{22}^{(2)} + a_{13}^{(2)}w_{32}^{(2)} & a_{11}^{(2)}w_{13}^{(2)} + a_{12}^{(2)}w_{23}^{(2)} + a_{13}^{(2)}w_{33}^{(2)} \\ a_{21}^{(2)}w_{11}^{(2)} + a_{22}^{(2)}w_{21}^{(2)} + a_{23}^{(2)}w_{31}^{(2)} & a_{21}^{(2)}w_{12}^{(2)} + a_{22}^{(2)}w_{22}^{(2)} + a_{23}^{(2)}w_{32}^{(2)} & a_{21}^{(2)}w_{13}^{(2)} + a_{22}^{(2)}w_{23}^{(2)} + a_{23}^{(2)}w_{33}^{(2)} \\ a_{31}^{(2)}w_{11}^{(2)} + a_{32}^{(2)}w_{21}^{(2)} + a_{33}^{(2)}w_{31}^{(2)} & a_{31}^{(2)}w_{12}^{(2)} + a_{32}^{(2)}w_{22}^{(2)} + a_{33}^{(2)}w_{32}^{(2)} & a_{31}^{(2)}w_{13}^{(2)} + a_{32}^{(2)}w_{23}^{(2)} + a_{33}^{(2)}w_{33}^{(2)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} & b_2^{(2)} & b_3^{(2)} \end{bmatrix}$$

ผลลัพธ์:

$$\mathbf{Z}^{(3)} = \begin{bmatrix} a_{11}^{(2)}w_{11}^{(2)} + a_{12}^{(2)}w_{21}^{(2)} + a_{13}^{(2)}w_{31}^{(2)} + b_1^{(2)} & a_{11}^{(2)}w_{12}^{(2)} + a_{12}^{(2)}w_{22}^{(2)} + a_{13}^{(2)}w_{32}^{(2)} + b_2^{(2)} & a_{11}^{(2)}w_{13}^{(2)} + a_{12}^{(2)}w_{23}^{(2)} + a_{13}^{(2)}w_{33}^{(2)} + b_3^{(2)} \\ a_{21}^{(2)}w_{11}^{(2)} + a_{22}^{(2)}w_{21}^{(2)} + a_{23}^{(2)}w_{31}^{(2)} + b_1^{(2)} & a_{21}^{(2)}w_{12}^{(2)} + a_{22}^{(2)}w_{22}^{(2)} + a_{23}^{(2)}w_{32}^{(2)} + b_2^{(2)} & a_{21}^{(2)}w_{13}^{(2)} + a_{22}^{(2)}w_{23}^{(2)} + a_{23}^{(2)}w_{33}^{(2)} + b_3^{(2)} \\ a_{31}^{(2)}w_{11}^{(2)} + a_{32}^{(2)}w_{21}^{(2)} + a_{33}^{(2)}w_{31}^{(2)} + b_1^{(2)} & a_{31}^{(2)}w_{12}^{(2)} + a_{32}^{(2)}w_{22}^{(2)} + a_{33}^{(2)}w_{32}^{(2)} + b_2^{(2)} & a_{31}^{(2)}w_{13}^{(2)} + a_{32}^{(2)}w_{23}^{(2)} + a_{33}^{(2)}w_{33}^{(2)} + b_3^{(2)} \end{bmatrix}$$

นายพรชัย คำแดง
รหัส 6305000942

คำนวณ Activation $A^{(3)}$ โดยใช้ฟังก์ชันกระตุ้น **tanh**:

$$A^{(3)} = \tanh(Z^{(3)})$$

$$\mathbf{A}^{(3)} = \begin{bmatrix} \tanh(z_{11}^{(3)}) & \tanh(z_{12}^{(3)}) & \tanh(z_{13}^{(3)}) \\ \tanh(z_{21}^{(3)}) & \tanh(z_{22}^{(3)}) & \tanh(z_{23}^{(3)}) \\ \tanh(z_{31}^{(3)}) & \tanh(z_{32}^{(3)}) & \tanh(z_{33}^{(3)}) \end{bmatrix}$$

Forward Propagation: Layer 3 → Layer 4 (Output)

คำนวณค่า pre-activation:

$$\mathbf{Z}^{(4)} = \mathbf{A}^{(3)}\mathbf{W}^{(3)} + \mathbf{b}^{(3)}$$

อินพุตของ Layer 4 คือ $A^{(3)}$ ซึ่งเป็น เมทริกซ์ขนาด (3,3) เนื่องจากมี 3 ตัวอย่าง (samples) และ 3 นิวรอนจาก Layer 3

$$\mathbf{A}^{(3)} = \begin{bmatrix} a_{11}^{(3)} & a_{12}^{(3)} & a_{13}^{(3)} \\ a_{21}^{(3)} & a_{22}^{(3)} & a_{23}^{(3)} \\ a_{31}^{(3)} & a_{32}^{(3)} & a_{33}^{(3)} \end{bmatrix}$$

คำนวณ **W**eighted Sum $Z^{(4)}$

Layer 4 มี 2 **นิวรอน**
ดังนั้น Weight Matrix $W^{(3)}$ **มีขนาด** $(3, 2)$

$$\mathbf{W}^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} \\ w_{31}^{(3)} & w_{32}^{(3)} \end{bmatrix}$$

Bias Vector $\mathbf{b}^{(3)}$ **มีขนาด** $(1, 2)$:

$$\mathbf{b}^{(3)} = \begin{bmatrix} b_1^{(3)} & b_2^{(3)} \end{bmatrix}$$

นายพรชัย คำแดง
รหัส 6305000942

คำนวณ Weighted Sum:

$$\mathbf{Z}^{(4)} = \mathbf{A}^{(3)}\mathbf{W}^{(3)} + \mathbf{b}^{(3)}$$

แทนค่าเมทริกซ์:

$$\begin{bmatrix} a_{11}^{(3)} & a_{12}^{(3)} & a_{13}^{(3)} \\ a_{21}^{(3)} & a_{22}^{(3)} & a_{23}^{(3)} \\ a_{31}^{(3)} & a_{32}^{(3)} & a_{33}^{(3)} \end{bmatrix} \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} \\ w_{31}^{(3)} & w_{32}^{(3)} \end{bmatrix} + \begin{bmatrix} b_1^{(3)} & b_2^{(3)} \end{bmatrix}$$

คำนวณดอทโปรดักต์:

$$\mathbf{Z}^{(4)} = \begin{bmatrix} a_{11}^{(3)}w_{11}^{(3)} + a_{12}^{(3)}w_{21}^{(3)} + a_{13}^{(3)}w_{31}^{(3)} & a_{11}^{(3)}w_{12}^{(3)} + a_{12}^{(3)}w_{22}^{(3)} + a_{13}^{(3)}w_{32}^{(3)} \\ a_{21}^{(3)}w_{11}^{(3)} + a_{22}^{(3)}w_{21}^{(3)} + a_{23}^{(3)}w_{31}^{(3)} & a_{21}^{(3)}w_{12}^{(3)} + a_{22}^{(3)}w_{22}^{(3)} + a_{23}^{(3)}w_{32}^{(3)} \\ a_{31}^{(3)}w_{11}^{(3)} + a_{32}^{(3)}w_{21}^{(3)} + a_{33}^{(3)}w_{31}^{(3)} & a_{31}^{(3)}w_{12}^{(3)} + a_{32}^{(3)}w_{22}^{(3)} + a_{33}^{(3)}w_{32}^{(3)} \end{bmatrix} + \begin{bmatrix} b_1^{(3)} & b_2^{(3)} \end{bmatrix}$$

ผลลัพธ์:

$$\mathbf{Z}^{(4)} = \begin{bmatrix} a_{11}^{(3)}w_{11}^{(3)} + a_{12}^{(3)}w_{21}^{(3)} + a_{13}^{(3)}w_{31}^{(3)} + b_1^{(3)} & a_{11}^{(3)}w_{12}^{(3)} + a_{12}^{(3)}w_{22}^{(3)} + a_{13}^{(3)}w_{32}^{(3)} + b_2^{(3)} \\ a_{21}^{(3)}w_{11}^{(3)} + a_{22}^{(3)}w_{21}^{(3)} + a_{23}^{(3)}w_{31}^{(3)} + b_1^{(3)} & a_{21}^{(3)}w_{12}^{(3)} + a_{22}^{(3)}w_{22}^{(3)} + a_{23}^{(3)}w_{32}^{(3)} + b_2^{(3)} \\ a_{31}^{(3)}w_{11}^{(3)} + a_{32}^{(3)}w_{21}^{(3)} + a_{33}^{(3)}w_{31}^{(3)} + b_1^{(3)} & a_{31}^{(3)}w_{12}^{(3)} + a_{32}^{(3)}w_{22}^{(3)} + a_{33}^{(3)}w_{32}^{(3)} + b_2^{(3)} \end{bmatrix}$$

ค่า Output คือ

$$\mathbf{A}^{(4)} = \mathbf{Z}^{(4)}$$

นายพรชัย คำแดง
รหัส 6305000942

3) Backpropagation:

Layer 4 → Layer 3

นิยามตัวแปร

- $\mathbf{A}^{(4)} = \mathbf{Z}^{(4)}$ **(ไม่มี Activation Function ที่ Layer 4)**
- **Loss Function:** สมมติว่าเราใช้ Mean Squared Error (MSE)

$$L = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{2}(A_{ij}^{(4)} - Y_{ij})^2$$

โดยที่:

- Y เป็นค่าจริง (Ground Truth) ขนาด $(3, 2)$
- $A^{(4)}$ เป็นค่าพยากรณ์ (Prediction) ขนาด $(3, 2)$

คำนวณ Gradient ของ Loss ต่อ $Z^{(4)}$

เนื่องจาก Output Layer **ไม่มีฟังก์ชันกระตุ้น**

$$\frac{\partial L}{\partial Z^{(4)}} = \frac{\partial L}{\partial A^{(4)}}$$

จาก MSE:

$$\frac{\partial L}{\partial A^{(4)}} = \frac{2}{N}(A^{(4)} - Y)$$

$$\delta^{(4)} = \frac{\partial L}{\partial Z^{(4)}} = \frac{2}{N}(A^{(4)} - Y)$$

$$\delta^{(4)} = \begin{bmatrix} \frac{2}{N}(a_{11}^{(4)} - y_{11}) & \frac{2}{N}(a_{12}^{(4)} - y_{12}) \\ \frac{2}{N}(a_{21}^{(4)} - y_{21}) & \frac{2}{N}(a_{22}^{(4)} - y_{22}) \\ \frac{2}{N}(a_{31}^{(4)} - y_{31}) & \frac{2}{N}(a_{32}^{(4)} - y_{32}) \end{bmatrix}$$

นายพรชัย คำแดง
รหัส 6305000942

คำนวณ Gradient ของ $W^{(3)}$ และ $b^{(3)}$

Gradient **ของ** $W^{(3)}$

$$\frac{\partial L}{\partial W^{(3)}} = A^{(3)T}\delta^{(4)}$$

โดยที่:

- $A^{(3)}$ มีขนาด $(3,3)$
- $\delta^{(4)}$ มีขนาด $(3,2)$
- $W^{(3)}$ มีขนาด $(3,2)$

เมทริกซ์ของ Gradient:

$$\begin{bmatrix} a_{11}^{(3)} & a_{21}^{(3)} & a_{31}^{(3)} \\ a_{12}^{(3)} & a_{22}^{(3)} & a_{32}^{(3)} \\ a_{13}^{(3)} & a_{23}^{(3)} & a_{33}^{(3)} \end{bmatrix} \begin{bmatrix} \delta_{11}^{(4)} & \delta_{12}^{(4)} \\ \delta_{21}^{(4)} & \delta_{22}^{(4)} \\ \delta_{31}^{(4)} & \delta_{32}^{(4)} \end{bmatrix}$$

ผลลัพธ์เป็นเมทริกซ์ขนาด $(3,2)$:

$$\frac{\partial L}{\partial W^{(3)}} = \begin{bmatrix} a_{11}^{(3)}\delta_{11}^{(4)} + a_{21}^{(3)}\delta_{21}^{(4)} + a_{31}^{(3)}\delta_{31}^{(4)} & a_{11}^{(3)}\delta_{12}^{(4)} + a_{21}^{(3)}\delta_{22}^{(4)} + a_{31}^{(3)}\delta_{32}^{(4)} \\ a_{12}^{(3)}\delta_{11}^{(4)} + a_{22}^{(3)}\delta_{21}^{(4)} + a_{32}^{(3)}\delta_{31}^{(4)} & a_{12}^{(3)}\delta_{12}^{(4)} + a_{22}^{(3)}\delta_{22}^{(4)} + a_{32}^{(3)}\delta_{32}^{(4)} \\ a_{13}^{(3)}\delta_{11}^{(4)} + a_{23}^{(3)}\delta_{21}^{(4)} + a_{33}^{(3)}\delta_{31}^{(4)} & a_{13}^{(3)}\delta_{12}^{(4)} + a_{23}^{(3)}\delta_{22}^{(4)} + a_{33}^{(3)}\delta_{32}^{(4)} \end{bmatrix}$$

Gradient **ของ** $b^{(3)}$

$$\frac{\partial L}{\partial b^{(3)}} = \sum_{i=1}^{3} \delta^{(4)}$$

$$\frac{\partial L}{\partial b^{(3)}} = \begin{bmatrix} \delta_{11}^{(4)} + \delta_{21}^{(4)} + \delta_{31}^{(4)} & \delta_{12}^{(4)} + \delta_{22}^{(4)} + \delta_{32}^{(4)} \end{bmatrix}$$

นายพรชัย คำแดง
รหัส 6305000942

คำนวณ Error ที่ส่งกลับไป Layer 3

$$\delta^{(3)} = \delta^{(4)} W^{(3)T} \odot \tanh'(Z^{(3)})$$

โดยที่:

- $W^{(3)T}$ มีขนาด $(2,3)$
- $\delta^{(4)}$ มีขนาด $(3,2)$
- $\delta^{(3)}$ คำนวณออกมาจะได้ขนาด $(3,3)$

$$\begin{bmatrix} \delta_{11}^{(4)} & \delta_{12}^{(4)} \\ \delta_{21}^{(4)} & \delta_{22}^{(4)} \\ \delta_{31}^{(4)} & \delta_{32}^{(4)} \end{bmatrix} \begin{bmatrix} w_{11}^{(3)} & w_{21}^{(3)} & w_{31}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(3)} & w_{32}^{(3)} \end{bmatrix}$$

ผลลัพธ์:

$$\delta^{(3)} = \begin{bmatrix} \delta_{11}^{(4)} w_{11}^{(3)} + \delta_{12}^{(4)} w_{12}^{(3)} & \delta_{11}^{(4)} w_{21}^{(3)} + \delta_{12}^{(4)} w_{22}^{(3)} & \delta_{11}^{(4)} w_{31}^{(3)} + \delta_{12}^{(4)} w_{32}^{(3)} \\ \delta_{21}^{(4)} w_{11}^{(3)} + \delta_{22}^{(4)} w_{12}^{(3)} & \delta_{21}^{(4)} w_{21}^{(3)} + \delta_{22}^{(4)} w_{22}^{(3)} & \delta_{21}^{(4)} w_{31}^{(3)} + \delta_{22}^{(4)} w_{32}^{(3)} \\ \delta_{31}^{(4)} w_{11}^{(3)} + \delta_{32}^{(4)} w_{12}^{(3)} & \delta_{31}^{(4)} w_{21}^{(3)} + \delta_{32}^{(4)} w_{22}^{(3)} & \delta_{31}^{(4)} w_{31}^{(3)} + \delta_{32}^{(4)} w_{32}^{(3)} \end{bmatrix} \odot (1 - A^{(3)} \odot A^{(3)})$$

Backpropagation: Layer 3 → Layer 2

นิยามตัวแปร

จาก Forward Propagation ก่อนหน้า:

- $A^{(3)} = \tanh(Z^{(3)})$
- $Z^{(3)} = A^{(2)} W^{(2)} + b^{(2)}$
- $A^{(2)} = \tanh(Z^{(2)})$

จาก Backpropagation ของ Layer 4 → Layer 3 เราคำนวณได้ว่า:

$$\delta^{(3)} = \left(\delta^{(4)} W^{(3)T}\right) \odot \left(1 - A^{(3)} \odot A^{(3)}\right)$$

นายพรชัย คำแดง
รหัส 6305000942

คำนวณ Gradient ของ $W^{(2)}$ และ $b^{(2)}$

Gradient ของ $W^{(2)}$

$$\frac{\partial L}{\partial W^{(2)}} = A^{(2)T}\delta^{(3)}$$

โดยที่:

- $A^{(2)}$ มีขนาด $(3,3)$
- $\delta^{(3)}$ มีขนาด $(3,3)$
- $W^{(2)}$ มีขนาด $(3,3)$

$$\begin{bmatrix} a_{11}^{(2)} & a_{21}^{(2)} & a_{31}^{(2)} \\ a_{12}^{(2)} & a_{22}^{(2)} & a_{32}^{(2)} \\ a_{13}^{(2)} & a_{23}^{(2)} & a_{33}^{(2)} \end{bmatrix} \begin{bmatrix} \delta_{11}^{(3)} & \delta_{12}^{(3)} & \delta_{13}^{(3)} \\ \delta_{21}^{(3)} & \delta_{22}^{(3)} & \delta_{23}^{(3)} \\ \delta_{31}^{(3)} & \delta_{32}^{(3)} & \delta_{33}^{(3)} \end{bmatrix}$$

ผลลัพธ์เป็นเมทริกซ์ขนาด $(3,3)$:

$$\frac{\partial L}{\partial W^{(2)}} = \begin{bmatrix} a_{11}^{(2)}\delta_{11}^{(3)} + a_{21}^{(2)}\delta_{21}^{(3)} + a_{31}^{(2)}\delta_{31}^{(3)} & a_{11}^{(2)}\delta_{12}^{(3)} + a_{21}^{(2)}\delta_{22}^{(3)} + a_{31}^{(2)}\delta_{32}^{(3)} & a_{11}^{(2)}\delta_{13}^{(3)} + a_{21}^{(2)}\delta_{23}^{(3)} + a_{31}^{(2)}\delta_{33}^{(3)} \\ a_{12}^{(2)}\delta_{11}^{(3)} + a_{22}^{(2)}\delta_{21}^{(3)} + a_{32}^{(2)}\delta_{31}^{(3)} & a_{12}^{(2)}\delta_{12}^{(3)} + a_{22}^{(2)}\delta_{22}^{(3)} + a_{32}^{(2)}\delta_{32}^{(3)} & a_{12}^{(2)}\delta_{13}^{(3)} + a_{22}^{(2)}\delta_{23}^{(3)} + a_{32}^{(2)}\delta_{33}^{(3)} \\ a_{13}^{(2)}\delta_{11}^{(3)} + a_{23}^{(2)}\delta_{21}^{(3)} + a_{33}^{(2)}\delta_{31}^{(3)} & a_{13}^{(2)}\delta_{12}^{(3)} + a_{23}^{(2)}\delta_{22}^{(3)} + a_{33}^{(2)}\delta_{32}^{(3)} & a_{13}^{(2)}\delta_{13}^{(3)} + a_{23}^{(2)}\delta_{23}^{(3)} + a_{33}^{(2)}\delta_{33}^{(3)} \end{bmatrix}$$

Gradient ของ $b^{(2)}$

$$\frac{\partial L}{\partial b^{(2)}} = \sum_{i=1}^{3} \delta^{(3)}$$

$$\frac{\partial L}{\partial b^{(2)}} = \begin{bmatrix} \delta_{11}^{(3)} + \delta_{21}^{(3)} + \delta_{31}^{(3)} \\ \delta_{12}^{(3)} + \delta_{22}^{(3)} + \delta_{32}^{(3)} \\ \delta_{13}^{(3)} + \delta_{23}^{(3)} + \delta_{33}^{(3)} \end{bmatrix}$$

คำนวณ Error ที่ส่งกลับไป Layer 2

$$\delta^{(2)} = \delta^{(3)}W^{(2)T} \odot \tanh'(Z^{(2)})$$

โดยที่:

- $W^{(2)T}$ มีขนาด $(3,3)$
- $\delta^{(3)}$ มีขนาด $(3,3)$
- $\delta^{(2)}$ คำนวณออกมาได้ขนาด $(3,3)$

$$\begin{bmatrix} \delta_{11}^{(3)} & \delta_{12}^{(3)} & \delta_{13}^{(3)} \\ \delta_{21}^{(3)} & \delta_{22}^{(3)} & \delta_{23}^{(3)} \\ \delta_{31}^{(3)} & \delta_{32}^{(3)} & \delta_{33}^{(3)} \end{bmatrix} \begin{bmatrix} w_{11}^{(2)} & w_{21}^{(2)} & w_{31}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} & w_{32}^{(2)} \\ w_{13}^{(2)} & w_{23}^{(2)} & w_{33}^{(2)} \end{bmatrix}$$

ผลลัพธ์:

$$\delta^{(2)} = \begin{bmatrix} \delta_{11}^{(3)}w_{11}^{(2)} + \delta_{12}^{(3)}w_{12}^{(2)} + \delta_{13}^{(3)}w_{13}^{(2)} & \delta_{11}^{(3)}w_{21}^{(2)} + \delta_{12}^{(3)}w_{22}^{(2)} + \delta_{13}^{(3)}w_{23}^{(2)} & \delta_{11}^{(3)}w_{31}^{(2)} + \delta_{12}^{(3)}w_{32}^{(2)} + \delta_{13}^{(3)}w_{33}^{(2)} \\ \delta_{21}^{(3)}w_{11}^{(2)} + \delta_{22}^{(3)}w_{12}^{(2)} + \delta_{23}^{(3)}w_{13}^{(2)} & \delta_{21}^{(3)}w_{21}^{(2)} + \delta_{22}^{(3)}w_{22}^{(2)} + \delta_{23}^{(3)}w_{23}^{(2)} & \delta_{21}^{(3)}w_{31}^{(2)} + \delta_{22}^{(3)}w_{32}^{(2)} + \delta_{23}^{(3)}w_{33}^{(2)} \\ \delta_{31}^{(3)}w_{11}^{(2)} + \delta_{32}^{(3)}w_{12}^{(2)} + \delta_{33}^{(3)}w_{13}^{(2)} & \delta_{31}^{(3)}w_{21}^{(2)} + \delta_{32}^{(3)}w_{22}^{(2)} + \delta_{33}^{(3)}w_{23}^{(2)} & \delta_{31}^{(3)}w_{31}^{(2)} + \delta_{32}^{(3)}w_{32}^{(2)} + \delta_{33}^{(3)}w_{33}^{(2)} \end{bmatrix} \odot (1 - A^{(2)} \odot A^{(2)})$$

นายพรชัย คำแดง
รหัส 6305000942

Backpropagation: Layer 2 → Layer 1

คำนวณ δ(1)

จากขั้นตอนของ Chain Rule, เราต้องคำนวณค่า error $\delta^{(1)}$ ซึ่งจะถูกใช้ในการคำนวณ Gradients ของ $W^{(1)}$ และ $b^{(1)}$

$$\delta^{(1)} = \left(\delta^{(2)} W^{(1)T}\right) \odot \tanh'(Z^{(1)})$$

โดยที่:

- $\delta^{(2)}$ คือ **delta ของ Layer 2** (ขนาด 3×3)
- $W^{(1)T}$ คือการ Transpose ของน้ำหนัก $W^{(1)}$ จาก Layer 1 (ขนาด 3×2)
- $\tanh'(Z^{(1)})$ คืออนุพันธ์ของฟังก์ชันกระตุ้น $\tanh$ ใน Layer 1 ซึ่งมีขนาด 3×3

ขั้นที่ 1: คำนวณ $\delta^{(1)}$

1. $\delta^{(2)}$ คำนวณจากการส่งผ่าน error มาจาก Layer 3 และเป็นเมทริกซ์ขนาด $(3, 3)$
2. $W^{(1)T}$ คือการ Transpose ของน้ำหนัก $W^{(1)}$ ซึ่งมีขนาด 3×2
3. $\tanh'(Z^{(1)})$: เราต้องคำนวณอนุพันธ์ของ $\tanh(Z^{(1)})$ ที่มีขนาด 3×3

สมมติว่า:

- $\delta^{(2)} = \begin{bmatrix} \delta_{11}^{(2)} & \delta_{12}^{(2)} & \delta_{13}^{(2)} \\ \delta_{21}^{(2)} & \delta_{22}^{(2)} & \delta_{23}^{(2)} \\ \delta_{31}^{(2)} & \delta_{32}^{(2)} & \delta_{33}^{(2)} \end{bmatrix}$
- $W^{(1)} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix}$
- $\tanh'(Z^{(1)}) = \begin{bmatrix} \tanh'(z_1^{(1)}) & \tanh'(z_2^{(1)}) & \tanh'(z_3^{(1)}) \\ \tanh'(z_4^{(1)}) & \tanh'(z_5^{(1)}) & \tanh'(z_6^{(1)}) \\ \tanh'(z_7^{(1)}) & \tanh'(z_8^{(1)}) & \tanh'(z_9^{(1)}) \end{bmatrix}$

ตอนนี้ **คำนวณ** $\delta^{(1)}$:

$$\delta^{(1)} = \left(\delta^{(2)} W^{(1)T}\right) \odot \tanh'(Z^{(1)})$$

คำนวณ $\delta^{(2)} W^{(1)T}$:

การคูณของ $\delta^{(2)}$ กับ $W^{(1)T}$:

$$\delta^{(2)} W^{(1)T} = \begin{bmatrix} \delta_{11}^{(2)} & \delta_{12}^{(2)} & \delta_{13}^{(2)} \\ \delta_{21}^{(2)} & \delta_{22}^{(2)} & \delta_{23}^{(2)} \\ \delta_{31}^{(2)} & \delta_{32}^{(2)} & \delta_{33}^{(2)} \end{bmatrix} \times \begin{bmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \end{bmatrix}$$

ซึ่งจะให้ผลลัพธ์เป็น:

$$\delta^{(2)}W^{(1)T} = \begin{bmatrix} \delta_{11}^{(2)}w_{11} + \delta_{12}^{(2)}w_{12} + \delta_{13}^{(2)}w_{13} & \delta_{11}^{(2)}w_{21} + \delta_{12}^{(2)}w_{22} + \delta_{13}^{(2)}w_{23} & \delta_{11}^{(2)}w_{31} + \delta_{12}^{(2)}w_{32} + \delta_{13}^{(2)}w_{33} \\ \delta_{21}^{(2)}w_{11} + \delta_{22}^{(2)}w_{12} + \delta_{23}^{(2)}w_{13} & \delta_{21}^{(2)}w_{21} + \delta_{22}^{(2)}w_{22} + \delta_{23}^{(2)}w_{23} & \delta_{21}^{(2)}w_{31} + \delta_{22}^{(2)}w_{32} + \delta_{23}^{(2)}w_{33} \\ \delta_{31}^{(2)}w_{11} + \delta_{32}^{(2)}w_{12} + \delta_{33}^{(2)}w_{13} & \delta_{31}^{(2)}w_{21} + \delta_{32}^{(2)}w_{22} + \delta_{33}^{(2)}w_{23} & \delta_{31}^{(2)}w_{31} + \delta_{32}^{(2)}w_{32} + \delta_{33}^{(2)}w_{33} \end{bmatrix}$$

ขั้นที่ 2: คูณด้วย $\tanh'(Z^{(1)})$

หลังจากที่เราได้ $\delta^{(2)}W^{(1)T}$, เราจะคูณมันกับ $\tanh'(Z^{(1)})$ แบบจุด (element-wise multiplication):

$$\delta^{(1)} = \left(\delta^{(2)}W^{(1)T}\right) \odot \tanh'(Z^{(1)})$$

โดยที่:

$$\delta^{(1)} = \begin{bmatrix} \delta_{11}^{(2)}w_{11} + \delta_{12}^{(2)}w_{12} + \delta_{13}^{(2)}w_{13} & \delta_{11}^{(2)}w_{21} + \delta_{12}^{(2)}w_{22} + \delta_{13}^{(2)}w_{23} & \delta_{11}^{(2)}w_{31} + \delta_{12}^{(2)}w_{32} + \delta_{13}^{(2)}w_{33} \\ \delta_{21}^{(2)}w_{11} + \delta_{22}^{(2)}w_{12} + \delta_{23}^{(2)}w_{13} & \delta_{21}^{(2)}w_{21} + \delta_{22}^{(2)}w_{22} + \delta_{23}^{(2)}w_{23} & \delta_{21}^{(2)}w_{31} + \delta_{22}^{(2)}w_{32} + \delta_{23}^{(2)}w_{33} \\ \delta_{31}^{(2)}w_{11} + \delta_{32}^{(2)}w_{12} + \delta_{33}^{(2)}w_{13} & \delta_{31}^{(2)}w_{21} + \delta_{32}^{(2)}w_{22} + \delta_{33}^{(2)}w_{23} & \delta_{31}^{(2)}w_{31} + \delta_{32}^{(2)}w_{32} + \delta_{33}^{(2)}w_{33} \end{bmatrix}$$

คูณกับ $\tanh'(Z^{(1)})$ ตามแต่ละตำแหน่งในเมทริกซ์

คำนวณ Gradients ของ $W^{(1)}$ และ $b^{(1)}$

Gradient ของ $W^{(1)}$:

$$\frac{\partial L}{\partial W^{(1)}} = A^{(0)T}\delta^{(1)}$$

โดยที่ $A^{(0)}$ คือ input ของ Layer 1 ซึ่งมีขนาด $(3, 2)$ ดังนั้นการคูณของ $A^{(0)T}$ และ $\delta^{(1)}$ จะให้ผลลัพธ์เป็น 2×3, ซึ่งตรงกับขนาดของ $W^{(1)}$

Gradient ของ $b^{(1)}$:

$$\frac{\partial L}{\partial b^{(1)}} = \sum_{i=1}^{3} \delta_i^{(1)}$$

โดยการ Sum ตาม Sample (ตามแถว) จะให้ผลลัพธ์เป็น 3×1 ซึ่งตรงกับขนาดของ $b^{(1)}$ ที่มีขนาด 1×3

นายพรชัย คำแดง
รหัส 6305000942

4) Gradient Descent

หลังจากที่คำนวณ **gradients** สำหรับแต่ละชั้น (Layer 1, Layer 2, และ Layer 3) แล้ว ขั้นตอนถัดไปคือการใช้ **Gradient Descent** เพื่ออัปเดตพารามิเตอร์ต่างๆ ในโมเดลของเรา (เช่น น้ำหนัก W และ b) โดยที่เราจะทำการอัปเดตตามสูตรของ **Gradient Descent**:

$$W^{(l)} = W^{(l)} - \alpha \frac{\partial L}{\partial W^{(l)}}$$

$$b^{(l)} = b^{(l)} - \alpha \frac{\partial L}{\partial b^{(l)}}$$

โดยที่:

- α คือ learning rate
- $\frac{\partial L}{\partial W^{(l)}}$ คือ gradient ของน้ำหนักในชั้นที่ l
- $\frac{\partial L}{\partial b^{(l)}}$ คือ gradient ของ bias ในชั้นที่ l
- $W^{(l)}$ คือน้ำหนักในชั้นที่ l
- $b^{(l)}$ คือ bias ในชั้นที่ l

อัปเดตน้ำหนักและ bias

หลังจากที่คำนวณ Gradients สำหรับทุกชั้นแล้ว, การอัปเดตน้ำหนักและ bias จะเป็นไปตามสูตรที่กล่าวไว้:

สำหรับ Layer 3 (output layer):

$$W^{(3)} = W^{(3)} - \alpha \frac{\partial L}{\partial W^{(3)}}$$

$$b^{(3)} = b^{(3)} - \alpha \frac{\partial L}{\partial b^{(3)}}$$

สำหรับ Layer 2:

$$W^{(2)} = W^{(2)} - \alpha \frac{\partial L}{\partial W^{(2)}}$$

$$b^{(2)} = b^{(2)} - \alpha \frac{\partial L}{\partial b^{(2)}}$$

สำหรับ Layer 1:

$$W^{(1)} = W^{(1)} - \alpha \frac{\partial L}{\partial W^{(1)}}$$

$$b^{(1)} = b^{(1)} - \alpha \frac{\partial L}{\partial b^{(1)}}$$