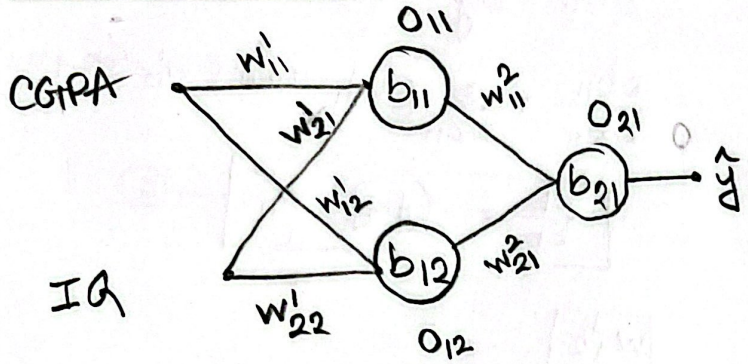
Backpropagation (Regression)

Neural Network:

CGPA	IQ	Package
3.3	8	20
2.9	9	50



Activation function: Linear

Regression → (Loss function) MSE, $L = \frac{1}{2}(y-\hat{y})^2$

(ypred) $\hat{y} = o_{21} = w^2_{11} \cdot o_{11} + w^2_{21} \cdot o_{12} + b_{21}$

Output Layer

$$w^2_{11new} = w^2_{11new} - \eta \frac{\partial L}{\partial w^2_{11old}}$$

$$w^2_{21new} = w^2_{11old} - \eta \frac{\partial L}{\partial w^2_{21old}}$$

$$b_{21new} = b_{21new} - \eta \frac{\partial L}{\partial b_{21}}$$

Derivatives terms: (9)

$\frac{\partial L}{\partial w^1_{11}}, \frac{\partial L}{\partial w^1_{12}}, \frac{\partial L}{\partial b_{11}}$

$\frac{\partial L}{\partial w^1_{21}}, \frac{\partial L}{\partial w^1_{22}}, \frac{\partial L}{\partial b_{12}}$

$\frac{\partial L}{\partial w^2_{11}}, \frac{\partial L}{\partial w^2_{21}}, \frac{\partial L}{\partial b_{21}}$

w^2_{11}

$$\therefore \frac{\partial L}{\partial w^2_{11}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w^2_{11}}$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}\left(\frac{1}{2}(y-\hat{y})^2\right) = -(y-\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w^2_{11}} = \frac{\partial}{\partial w^2_{11}}\left(w^2_{11} \cdot o_{11} + w^2_{21} o_{12} + b_{21}\right) = o_{11}$$

$$\therefore \boxed{\frac{\partial L}{\partial w^2_{11}} = -(y-\hat{y}) \cdot o_{11}} \quad -(1)$$

w^2_{21}

$$\frac{\partial L}{\partial w^2_{21}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w^2_{21}}$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}\left(\frac{1}{2}(y-\hat{y})^2\right) = -(y-\hat{y})$$

$$\frac{\partial \hat{y}}{\partial w^2_{21}} = \frac{\partial}{\partial w^2_{21}}\left(w^2_{11} \cdot o_{11} + w^2_{21} \cdot o_{12} + b_{21}\right) = o_{12}$$

$$\therefore \boxed{\frac{\partial L}{\partial w^2_{21}} = -(y-\hat{y}) \cdot o_{12}} \quad -(2)$$

b_{21}

$$\frac{\partial L}{\partial b_{21}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b_{21}}$$

$$\frac{\partial L}{\partial \hat{y}} = -(y - \hat{y})$$

$$\frac{\partial \hat{y}}{\partial b_{21}} = \frac{\partial}{\partial b_{21}}\left(w_{11}^2 o_{11} + w_{21}^2 o_{12} + b_{21}\right) = 1$$

$$\therefore \boxed{\frac{\partial L}{\partial b_{21}} = -(y - \hat{y})} \quad \text{---} (3)$$

w_{11}^1

$$\frac{\partial L}{\partial w_{11}^1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial o_{11}} \cdot \frac{\partial o_{11}}{\partial w_{11}^1}$$

$$\frac{\partial \hat{y}}{\partial o_{11}} = \frac{\partial}{\partial o_{11}}\left(w_{11}^2 \cdot o_{11} + w_{21}^2 o_{12} + b_{21}\right) = w_{11}^2$$

$$\frac{\partial o_{11}}{\partial w_{11}^1} = \frac{\partial}{\partial w_{11}^1}\left(w_{11}^1 \cdot \text{CGPA} + w_{12}^1 \cdot \text{IQ} + b_{11}\right) = \text{CGPA}$$

$$\boxed{\frac{\partial L}{\partial w_{11}^1} = -(y - \hat{y}) \cdot w_{11}^2 \cdot \text{CGPA}} \quad \text{---} (4)$$

w_{21}^1

$$\frac{\partial L}{\partial w_{12}^1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial o_{12}} \cdot \frac{\partial o_{12}}{\partial w_{12}^1}$$

$$\frac{\partial \hat{y}}{\partial o_{11}} = w_{11}^2$$

$$\frac{\partial o_{12}}{\partial w_{21}^1} = \frac{\partial}{\partial w_{21}^1}\left(w_{11}^1 \text{CGPA} + w_{21}^1 \cdot \text{IQ} + b_{11}\right) = \text{IQ}$$

$$\boxed{\frac{\partial L}{\partial w_{12}^1} = -(y - \hat{y}) \cdot w_{11}^2 \cdot \text{IQ}} \quad \text{---} (5)$$

b_{11}

$$\frac{\partial L}{\partial b_{11}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial o_{11}} \cdot \frac{\partial o_{11}}{\partial b_{11}}$$

$$\frac{\partial o_{11}}{\partial b_{11}} = \frac{\partial}{\partial b_{11}}\left(w_{11}^1 \cdot \text{CGPA} + w_{12}^1 \cdot \text{IQ} + b_{11}\right) = 1$$

$$\therefore \boxed{\frac{\partial L}{\partial b_{11}} = -(y - \hat{y}) \cdot w_{11}^2} \quad \text{---} (6)$$

w_{12}^1

$$\frac{\partial L}{\partial w_{12}^1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial o_{12}} \cdot \frac{\partial o_{12}}{\partial w_{12}^1}$$

$$\frac{\partial \hat{y}}{\partial o_{12}} = \frac{\partial}{\partial o_{12}}\left(w_{11}^2 \cdot o_{11} + w_{21}^2 \cdot o_{12} + b_{21}\right) = w_{21}^2$$

$$\frac{\partial o_{12}}{\partial w_{12}^1} = \frac{\partial}{\partial w_{12}^1}\left(w_{12}^1 \cdot \text{CGPA} + w_{22}^1 \cdot \text{IQ} + b_{12}\right) = \text{CGPA}$$

$$\boxed{\frac{\partial L}{\partial w_{12}^1} = -(y - \hat{y}) \cdot w_{21}^2 \cdot \text{CGPA}} \quad \text{---} (7)$$

w_{22}^1

$$\frac{\partial L}{\partial w_{22}^1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial o_{12}} \cdot \frac{\partial o_{12}}{\partial w_{22}^1}$$

$$\frac{\partial o_{12}}{\partial w_{22}^1} = \frac{\partial}{\partial w_{22}^1}\left(w_{12}^1 \cdot \text{CGPA} + w_{22}^1 \cdot \text{IQ} + b_{12}\right) = \text{IQ}$$

$$\boxed{\frac{\partial L}{\partial w_{22}} = -(y - \hat{y}) \cdot w_{21}^2 \cdot \text{IQ}} \quad \text{---} (8)$$

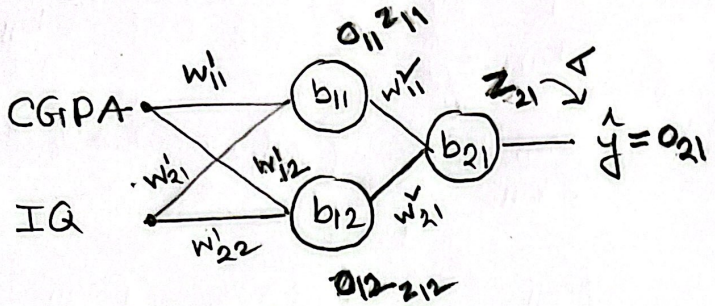
b_{12}

$$\frac{\partial L}{\partial b_{12}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial o_{12}} \cdot \frac{\partial o_{12}}{\partial b_{12}}$$

$$\frac{\partial o_{12}}{\partial b_{12}} = \frac{\partial}{\partial b_{12}}\left(w_{12}^1 \text{CGPA} + w_{22}^1 \cdot \text{IQ} + b_{12}\right) = 1$$

$$\boxed{\frac{\partial L}{\partial b_{12}} = -(y - \hat{y}) \cdot w_{21}^2} \quad \text{---} (9)$$

Backpropagation (classification)



Activation = Sigmoid

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

$z \to 0$

(Loss function)

Binary Cross Entropy, $(BCE)L = -y \log \hat{y} + (1-y)\log(1-\hat{y})$

$\hat{y} = \sigma(z)$; $z_{21} = w^2_{11} \cdot o_{11} + w^2_{21} \cdot o_{12} + b_{21}$

trainable parameters = 9

$\frac{\partial L}{\partial w^2_{11}}, \frac{\partial L}{\partial w^2_{21}}, \frac{\partial L}{\partial b_{21}}, \mid \frac{\partial L}{\partial w^1_{11}}, \frac{\partial L}{\partial w^1_{21}}, \frac{\partial L}{\partial b_{11}} \mid \frac{\partial L}{\partial w_{12}}, \frac{\partial L}{\partial w_{22}}, \frac{\partial L}{\partial b_{12}}$

w^2_{11}

$$\frac{\partial L}{\partial w^2_{11}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_{21}} \cdot \frac{\partial z_{21}}{\partial w^2_{11}}$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}} \left(-y \log \hat{y} - (1-y)\log(1-\hat{y})\right)$$

$$= -y \cdot \frac{1}{\hat{y}} + (1-y)\frac{1}{(1-\hat{y})}$$

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

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

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

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

$$\frac{\partial L}{\partial \hat{y}} = -\frac{(y-\hat{y})}{\hat{y}(1-\hat{y})}$$

$$\frac{\partial \hat{y}}{\partial z} = \sigma'(z_{21}) = \sigma(z_{21}) - (1 - \sigma(z_{21}))$$

$$= \hat{y} \cdot (1-\hat{y})$$

$$\frac{\partial z_{21}}{\partial w^2_{11}} = \frac{\partial}{\partial w^2_{11}} \cdot \left(w^2_{11} \cdot o_{11} + w^2_{21} \cdot o_{12} + b_{21}\right)$$

$$\frac{\partial z_{21}}{\partial w^2_{11}} = o_{11}$$

$$\therefore \frac{\partial L}{\partial w^2_{11}} = -\frac{y-\hat{y}}{\hat{y}(1-\hat{y})} \cdot \hat{y}(1-\hat{y}) \cdot o_{11}$$

$$\boxed{\frac{\partial L}{\partial w^2_{11}} = -(y-\hat{y}) \cdot o_{11}} \quad -(1)$$

$$\frac{\partial L}{\partial w^2_{21}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_{21}} \cdot \frac{\partial z_{21}}{\partial w^2_{21}}$$

$$\Rightarrow \frac{\partial z_{21}}{\partial w^2_{21}} = \frac{\partial}{\partial w^2_{21}} \left(w^2_{11} \cdot o_{11} + w^2_{21} \cdot o_{12} + b_{21}\right)$$

$$= o_{12}$$

$$\therefore \boxed{\frac{\partial L}{\partial w^2_{21}} = -(y-\hat{y}) \cdot o_{12}} \quad -(2)$$

b_{21}

$$\frac{\partial L}{\partial b_{21}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z_{21}}{\partial b_{21}}$$

$$\frac{\partial z_{21}}{\partial b_{21}} = \frac{\partial}{\partial b_{21}} \cdot \left(w_{11}^{2} \cdot o_{11} + w_{12}^{2} \cdot o_{12} + b_{21} \right) = 1$$

$$\boxed{\frac{\partial L}{\partial b_{21}} = -(y - \hat{y})} \quad -③$$

w_{11}^{1}

$$\frac{\partial L}{\partial w_{11}^{1}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_{21}} \cdot \frac{\partial z_{21}}{\partial o_{11}} \cdot \frac{\partial o_{11}}{\partial z_{11}} \cdot \frac{\partial z_{11}}{\partial w_{11}^{1}}$$

$$\frac{\partial z_{21}}{\partial o_{11}} = \frac{\partial}{\partial o_{11}} \left(w_{11}^{2} \cdot o_{11} + w_{21}^{2} \cdot o_{12} + b_{21} \right)$$

$$= w_{11}^{2}$$

$$\frac{\partial o_{11}}{\partial z_{11}} = \sigma'(o_{11}) \quad (z_{11} = o_{11} \cdot CGPA + w_{21}^{1} \cdot IQ + b_{11})$$

$$= o_{11}(1 - o_{11})$$

$$\frac{\partial z_{11}}{\partial w_{11}^{1}} = \frac{\partial}{\partial w_{11}^{1}} \left(w_{11}^{1} \cdot CGPA + w_{21}^{1} IQ + b_{11} \right)$$

$$= CGPA$$

$$\boxed{\frac{\partial L}{\partial w_{11}^{1}} = -(y - \hat{y}) \cdot w_{11}^{2} \cdot o_{12}(1 - o_{11}) \cdot CGPA} \quad -④$$

w_{21}^{1}

$$\frac{\partial L}{\partial w_{21}^{1}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{z_{21}} \cdot \frac{\partial z_{21}}{\partial o_{11}} \cdot \frac{\partial z_{11}}{\partial w_{21}^{1}}$$

$$\frac{\partial z_{11}}{\partial w_{21}^{1}} = \frac{\partial}{\partial w_{21}^{1}} \left(w_{11}^{1} \cdot CGPA + w_{21}^{1} \cdot IQ + b_{11} \right)$$

$$= IQ$$

$$\boxed{\frac{\partial L}{\partial w_{21}^{1}} = -(y - \hat{y}) w_{11}^{2} \cdot o_{11}(1 - o_{11}) \cdot IQ} \quad -⑤$$

b_{11}

$$\frac{\partial L}{\partial b_{11}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_{21}} \cdot \frac{\partial z_{21}}{\partial o_{11}} \cdot \frac{\partial o_{11}}{\partial z_{11}} \cdot \frac{\partial z_{11}}{\partial b_{11}}$$

$$\frac{\partial z_{11}}{\partial b_{11}} = \frac{\partial}{\partial b_{11}} \left(w_{11}^{2} \cdot CGPA + w_{21}^{1} \cdot IQ + b_{11} \right)$$

$$= 1$$

$$\boxed{\frac{\partial L}{\partial b_{11}} = -(y - \hat{y}) \cdot o_{11} \cdot (1 - o_{11})} \quad -⑥$$

w_{12}^{1}

$$\frac{\partial L}{\partial w_{12}^{1}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_{21}} \cdot \frac{\partial z_{21}}{\partial o_{12}} \cdot \frac{\partial o_{12}}{\partial z_{12}} \cdot \frac{\partial z_{12}}{\partial w_{12}^{1}}$$

$$\frac{\partial z_{21}}{\partial o_{12}} = \frac{\partial}{\partial o_{11}} \left(w_{11}^{2} \cdot o_{11} + w_{21}^{2} \cdot o_{12} + b_{21} \right)$$

$$= w_{21}^{2}$$

$$\frac{\partial o_{12}}{\partial z_{12}} = \sigma'(o_{12}) = o_{12} \cdot (1 - o_{12})$$

$$\frac{\partial z_{12}}{\partial w_{12}^{1}} = \frac{\partial}{\partial w_{12}^{1}} \left(w_{12}^{1} \cdot CGPA + w_{22}^{1} \cdot IQ + b_{12} \right)$$

$$= CGPA$$

$$\boxed{\frac{\partial L}{\partial w_{12}^{1}} = -(y - \hat{y}) \cdot w_{21}^{2} \cdot o_{12}(1 - o_{12}) \cdot CGPA} \quad -⑦$$

w_{22}^{1}

$$\frac{\partial L}{\partial w_{22}^{1}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_{21}} \cdot \frac{\partial z_{21}}{\partial o_{12}} \cdot \frac{\partial o_{12}}{\partial z_{12}} \cdot \frac{\partial z_{12}}{\partial w_{22}^{1}}$$

$$\frac{\partial z_{12}}{\partial w_{22}^{1}} = \frac{\partial}{\partial w_{22}^{1}} \left(w_{12}^{1} \cdot CGPA + w_{22}^{1} IQ + b_{12} \right)$$

$$= IQ$$

$$\boxed{\frac{\partial L}{\partial w_{22}} = -(y - \hat{y}) \cdot w_{21}^{2} \cdot o_{12}(1 - o_{12}) \cdot IQ} \quad -⑧$$

b_{12}

$$\frac{\partial L}{\partial b_{12}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_{21}} \cdot \frac{\partial z_{21}}{\partial o_{12}} \cdot \frac{\partial o_{12}}{\partial z_{12}} \cdot \frac{\partial z_{12}}{\partial b_{11}}$$

$$\frac{\partial z_{12}}{\partial b_{11}} = \frac{\partial}{\partial b_{12}} \left(w_{12}^{1} \cdot CGPA + w_{22}^{1} \cdot IQ + b_{12} \right) = 1$$

$$\boxed{\frac{\partial L}{\partial b_{12}} = -(y - \hat{y}) \cdot w_{21}^{2} \cdot o_{12}(1 - o_{12})} \quad -⑨$$

C1

Classification (chain)