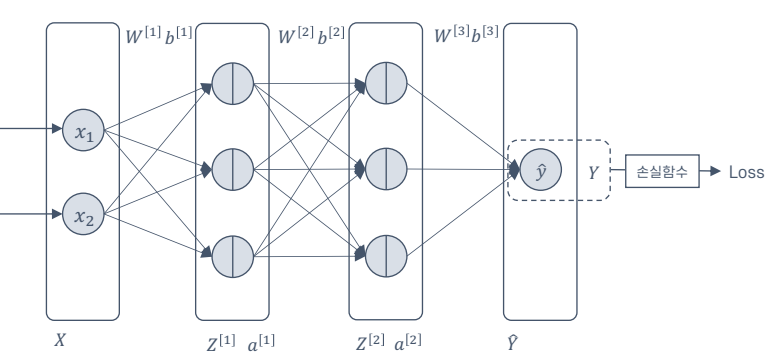


역전파_{Backpropagation} 알고리즘 (1) :

순전파를 하면서 각층의 기울기를 구하기



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

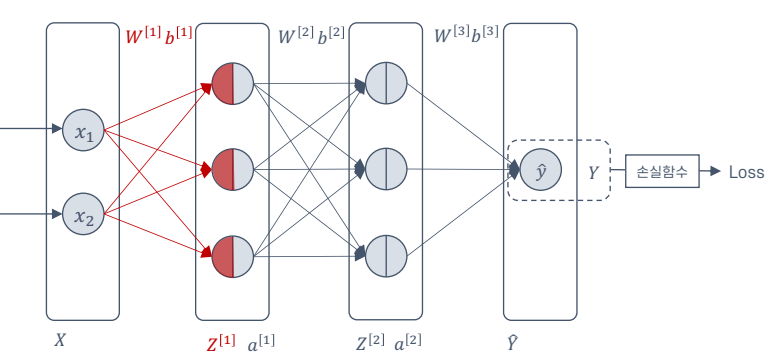
두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

각층의 기울기



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

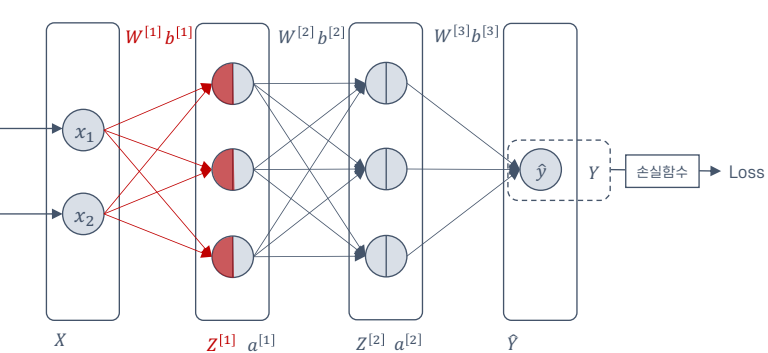
출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

각층의 기울기



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

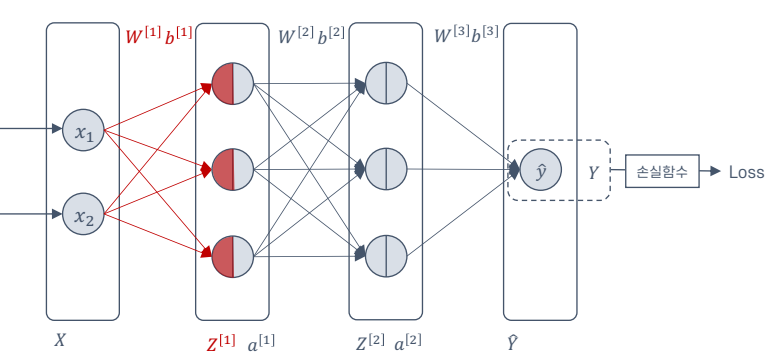
출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{Y}

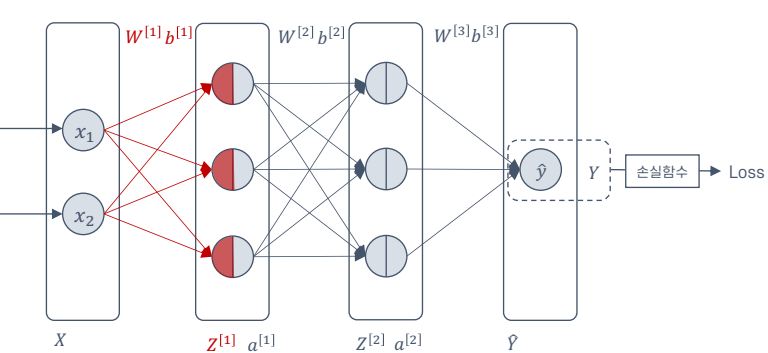
순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{y}

순전파

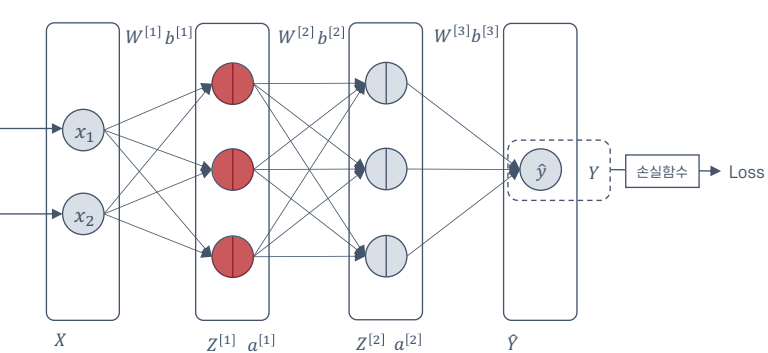
$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $Z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $Z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

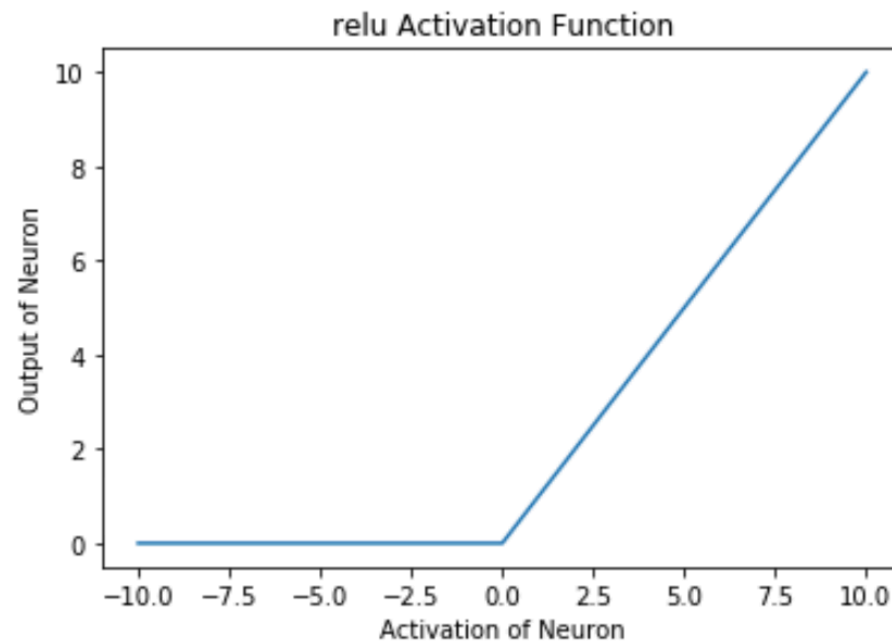
$$a^{[1]} = \text{relu}(Z^{[1]})$$

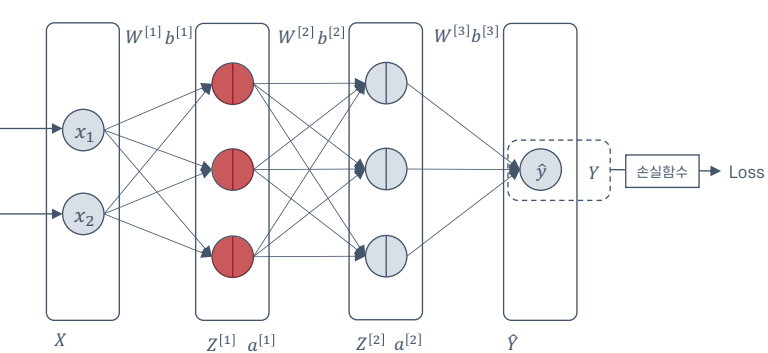
각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$





입력층 : X

첫번째 은닉층의 가중치 : $w^{[1],b[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2],b[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3],b[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

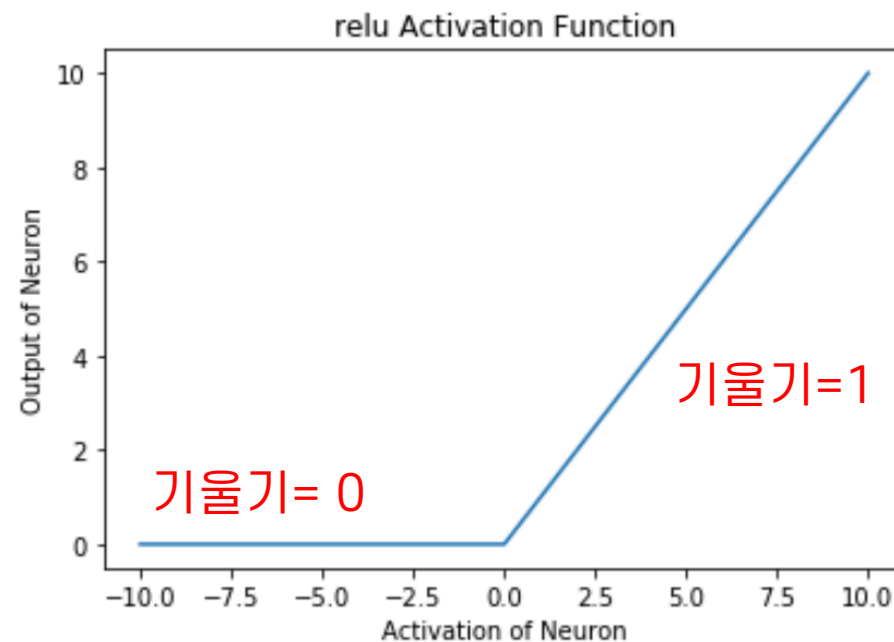
각층의 기울기

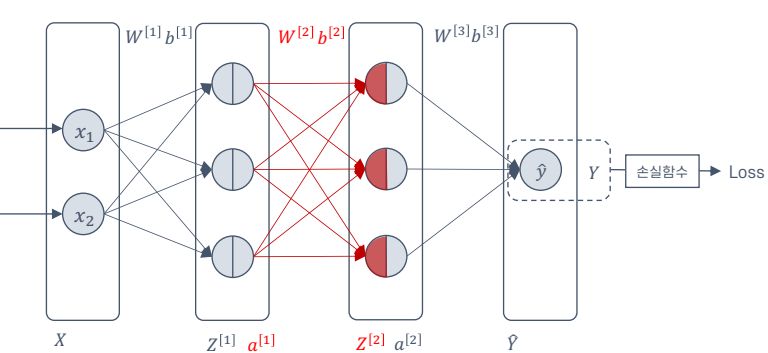
$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$





입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

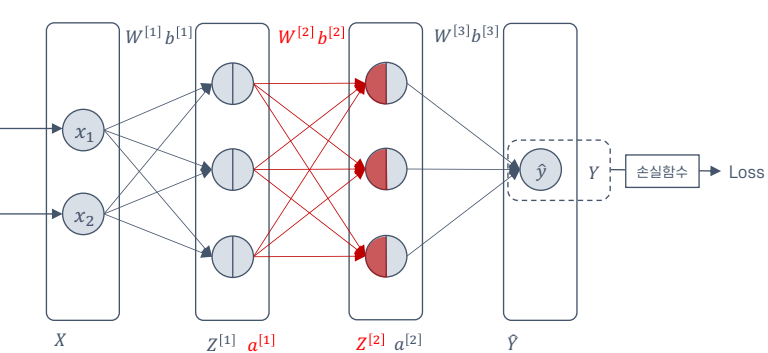
$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \quad \frac{dZ^{[1]}}{db^{[1]}} = 1 \quad \frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

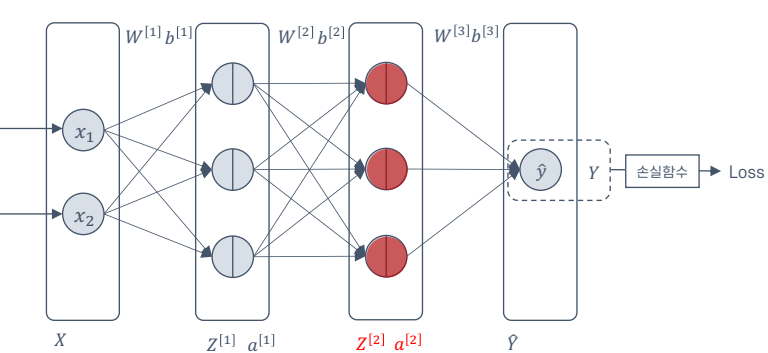
$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \quad \frac{dZ^{[1]}}{db^{[1]}} = 1 \quad \frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]} \quad \frac{dZ^{[2]}}{db^{[2]}} = 1 \quad \frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

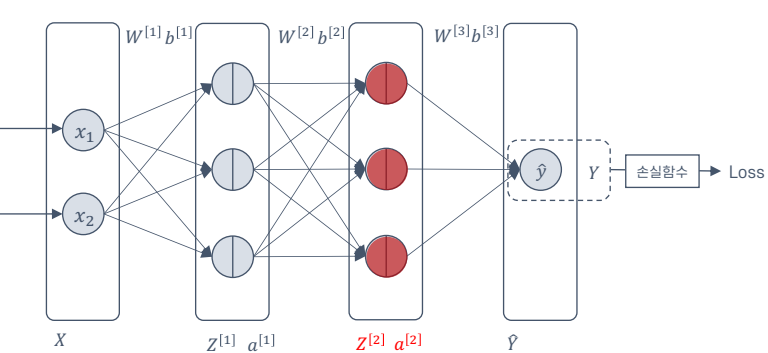
$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

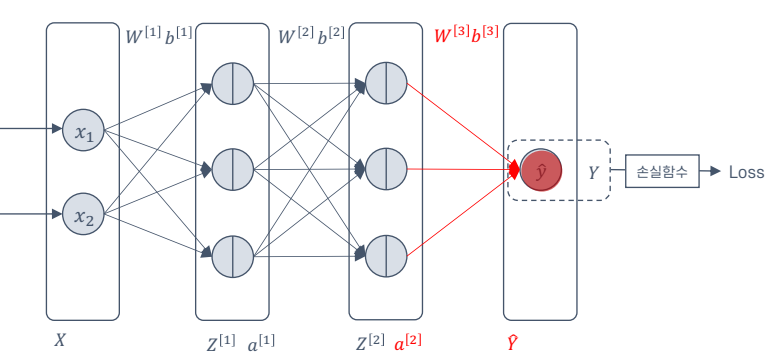
각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \quad \frac{dZ^{[1]}}{db^{[1]}} = 1 \quad \frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]} \quad \frac{dZ^{[2]}}{db^{[2]}} = 1 \quad \frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{y} = a^{[2]}W^{[3]} + b^{[3]}$$

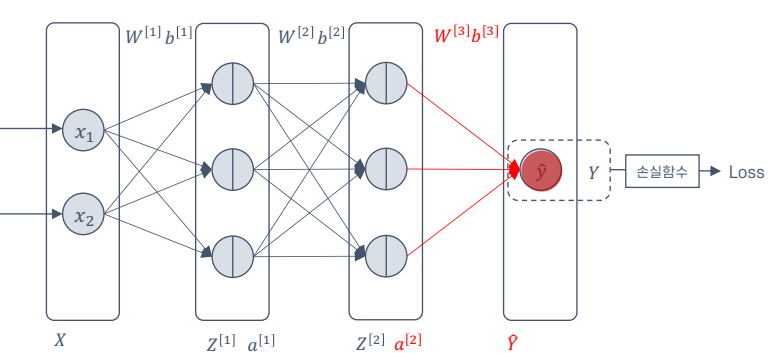
각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \quad \frac{dZ^{[1]}}{db^{[1]}} = 1 \quad \frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]} \quad \frac{dZ^{[2]}}{db^{[2]}} = 1 \quad \frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

각층의 기울기

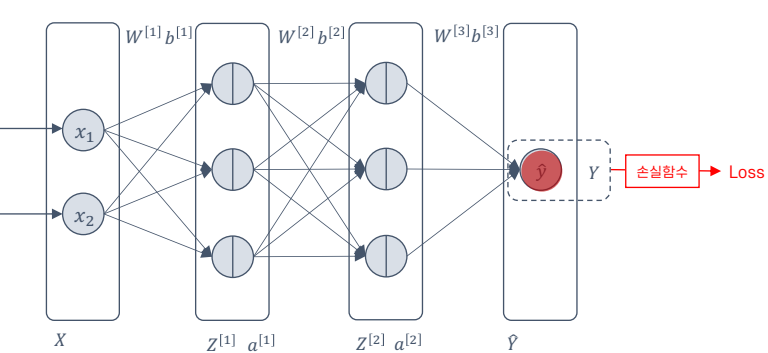
$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \quad \frac{dZ^{[1]}}{db^{[1]}} = 1 \quad \frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]} \quad \frac{dZ^{[2]}}{db^{[2]}} = 1 \quad \frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]} \quad \frac{d\hat{Y}}{db^{[3]}} = 1 \quad \frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $Z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $Z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

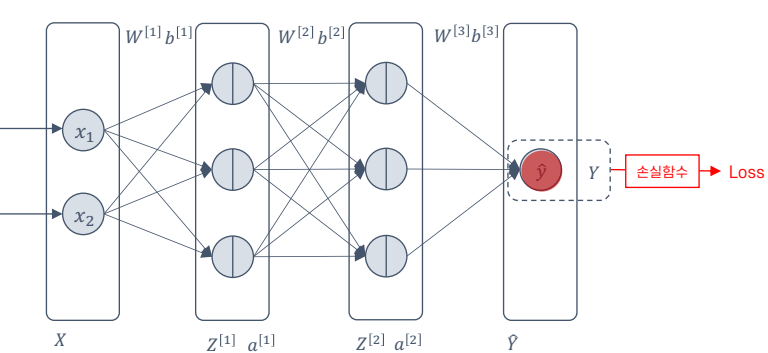
$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \quad \frac{dZ^{[1]}}{db^{[1]}} = 1 \quad \frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]} \quad \frac{dZ^{[2]}}{db^{[2]}} = 1 \quad \frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]} \quad \frac{d\hat{Y}}{db^{[3]}} = 1 \quad \frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1],b^{[1]}}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2],b^{[2]}}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3],b^{[3]}}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$Loss = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

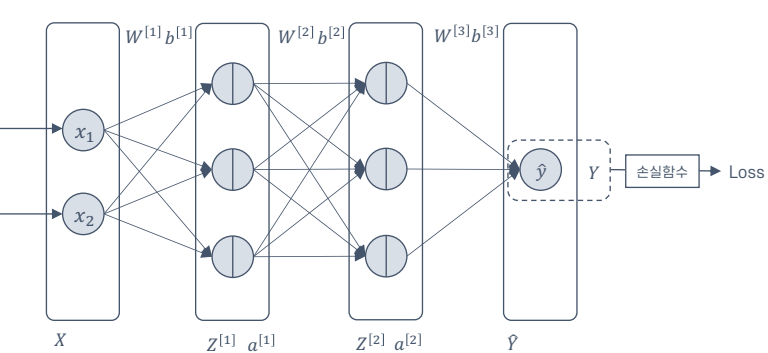
$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1],b[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2],b[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3],b[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

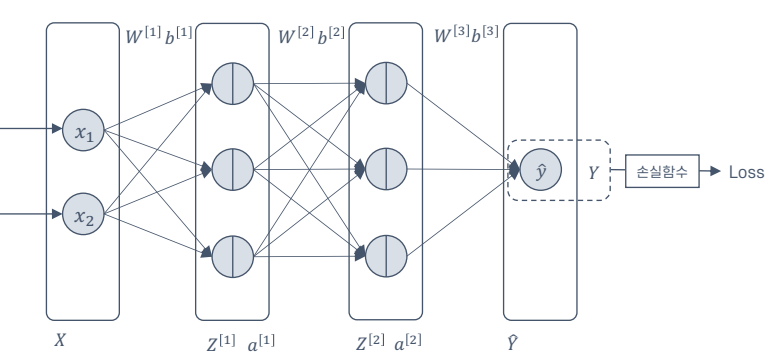
학습에 필요한 그래디언트

$$\frac{\partial \text{Loss}}{\partial W^{[1]}} , \frac{\partial \text{Loss}}{\partial W^{[2]}} , \frac{\partial \text{Loss}}{\partial W^{[3]}}$$

$$\frac{\partial \text{Loss}}{\partial b^{[1]}} , \frac{\partial \text{Loss}}{\partial b^{[2]}} , \frac{\partial \text{Loss}}{\partial b^{[3]}}$$

역전파_{Backpropagation} 알고리즘 (2) :

각 가중치의 손실에 대한 기울기를 구하기



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1],b^{[1]}}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2],b^{[2]}}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3],b^{[3]}}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

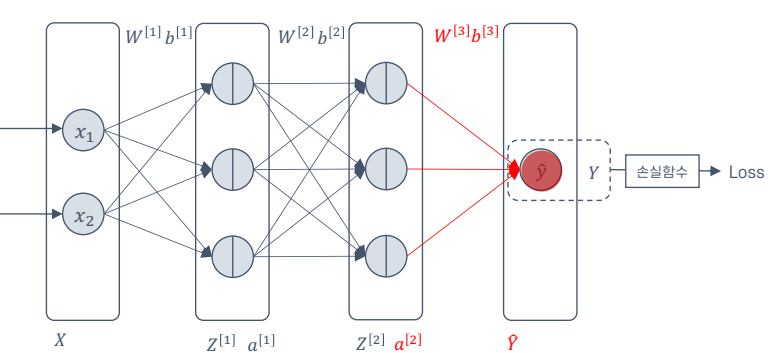
$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

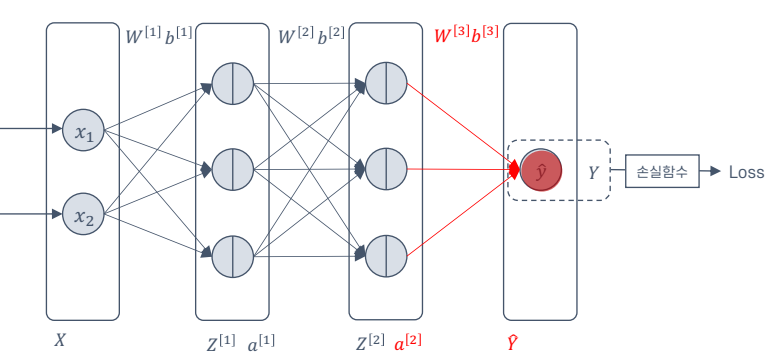
$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

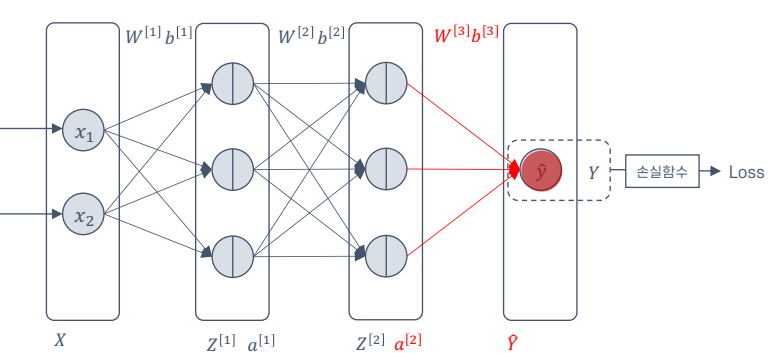
$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1],b[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2],b[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3],b[3]}$

출력층 : \hat{y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

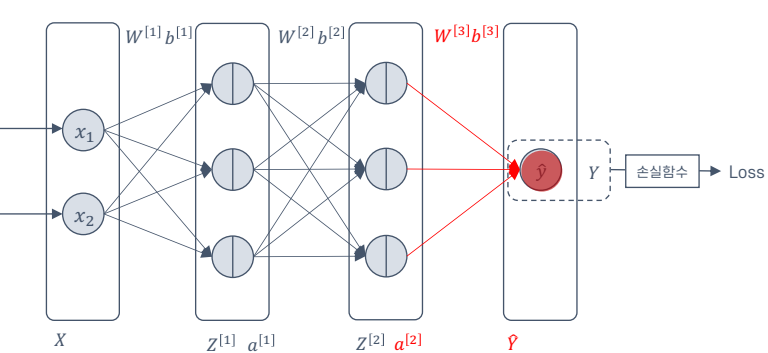
$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1],b^{[1]}}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2],b^{[2]}}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3],b^{[3]}}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

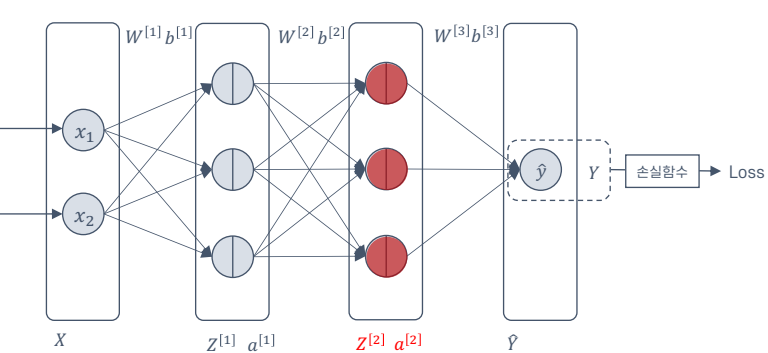
$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1],b^{[1]}}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2],b^{[2]}}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3],b^{[3]}}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

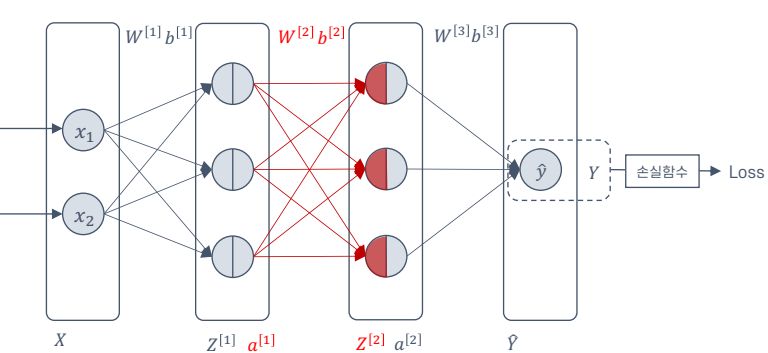
역전파

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1], b^{[1]}}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2], b^{[2]}}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3], b^{[3]}}$

출력층 : \hat{y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}}$$

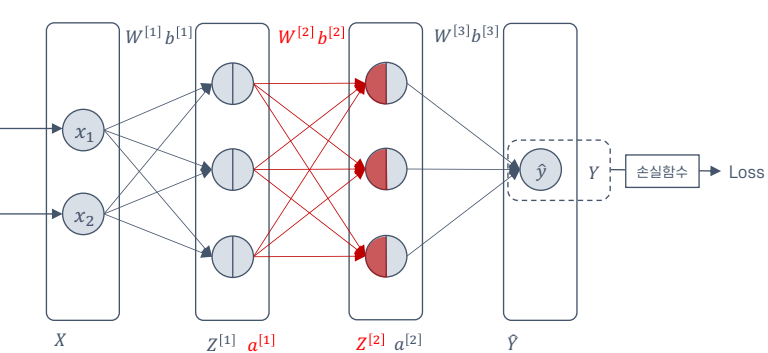
$$\frac{dL}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}}$$

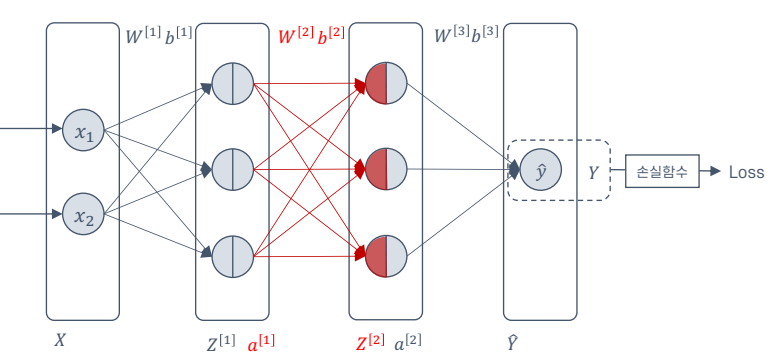
$$\frac{dL}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $Z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $Z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

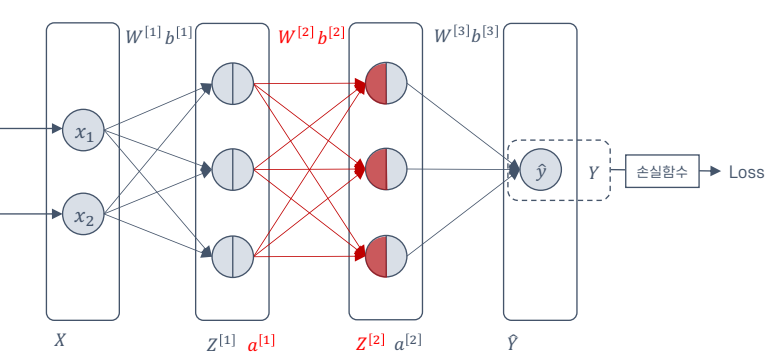
$$\frac{dL}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

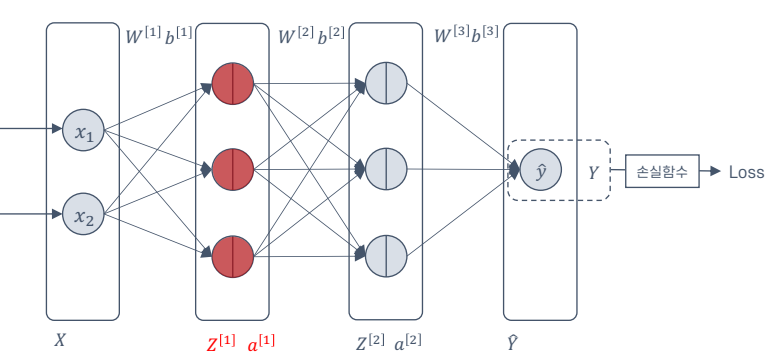
$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1],b[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2],b[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3],b[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

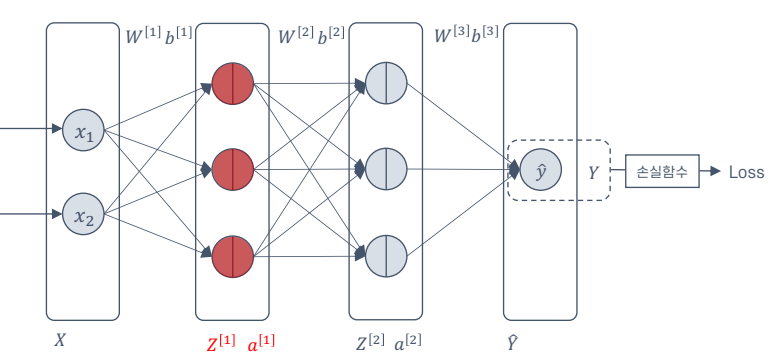
$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1], b^{[1]}}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2], b^{[2]}}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3], b^{[3]}}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

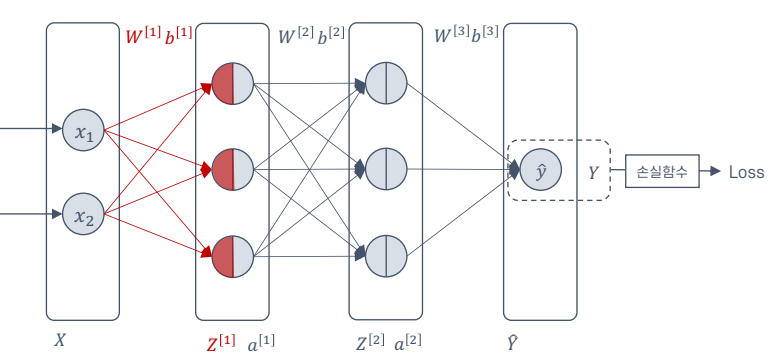
$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $Z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $Z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}}$$

$$\frac{dL}{dX}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

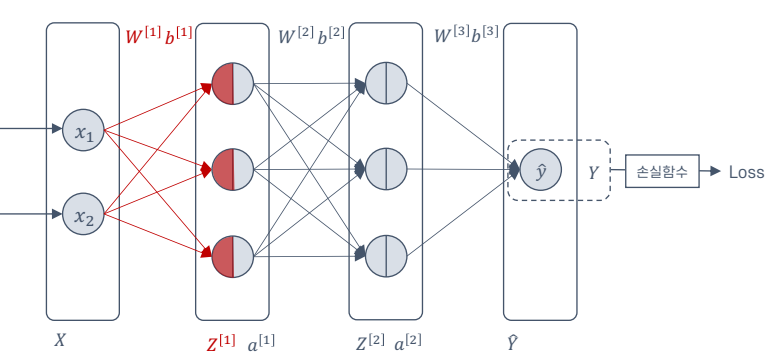
$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}}$$

$$\frac{dL}{dX}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

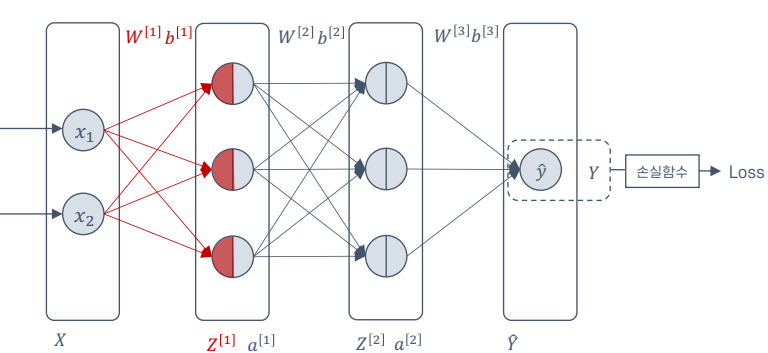
$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$$

$$\frac{dL}{dX}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

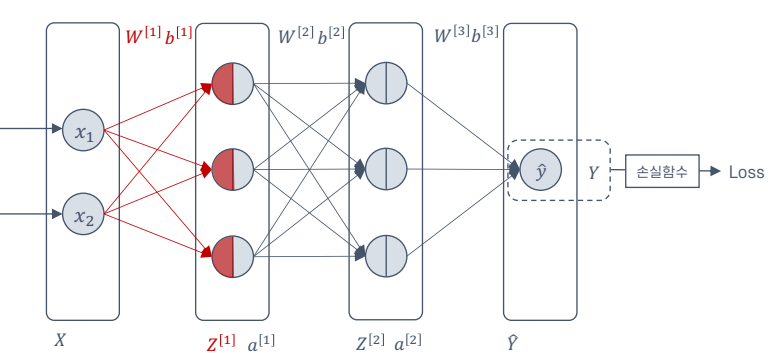
$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1], b^{[1]}}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $w^{[2], b^{[2]}}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3], b^{[3]}}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$$

$$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

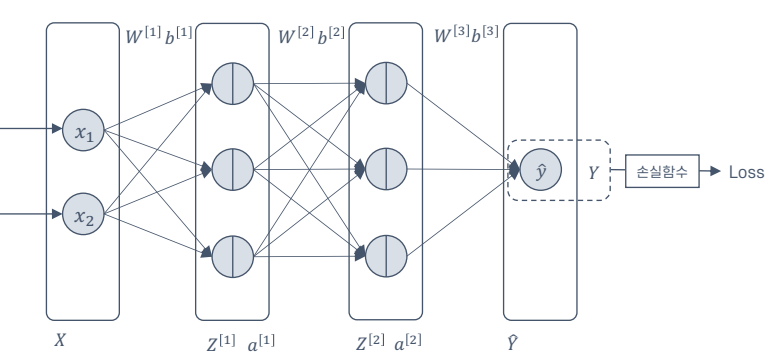
$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $Z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $Z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$$

$$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

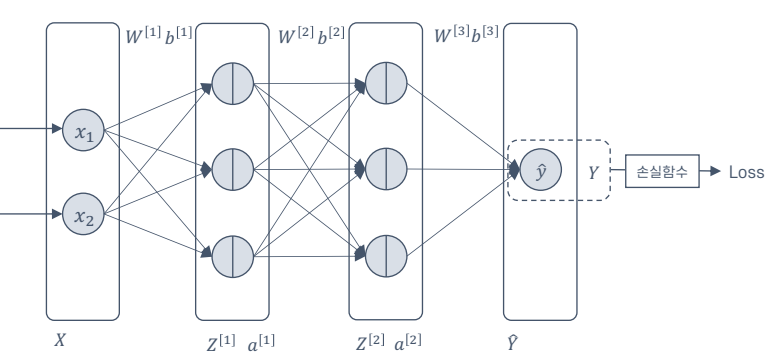
$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

인공신경망의 학습에 필요한 그래디언트 계산

역전파 알고리즘 개괄



입력층 : X

첫번째 은닉층의 가중치 : $w^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

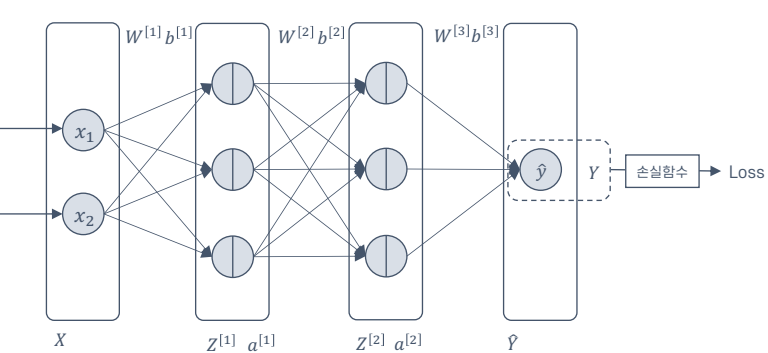
두번째 은닉층의 가중치 : $w^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $w^{[3]}, b^{[3]}$

출력층 : \hat{Y}



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $Z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $Z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T \quad \frac{dZ^{[1]}}{db^{[1]}} = 1 \quad \frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

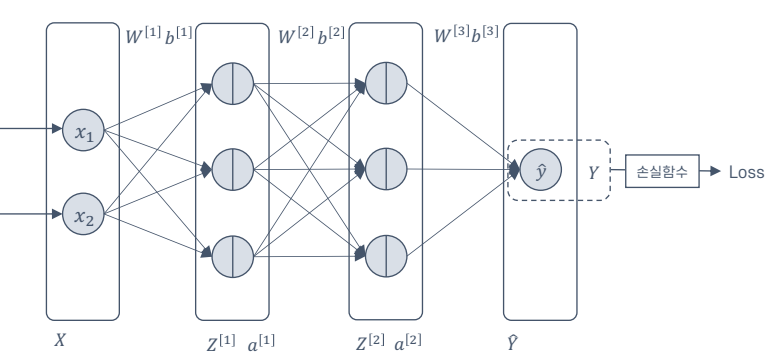
$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]} \quad \frac{dZ^{[2]}}{db^{[2]}} = 1 \quad \frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]} \quad \frac{d\hat{Y}}{db^{[3]}} = 1 \quad \frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

(1) 순전파를 진행하면서 필요한 미분값 및 손실값을 구한 후



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $Z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $Z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$$

$$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

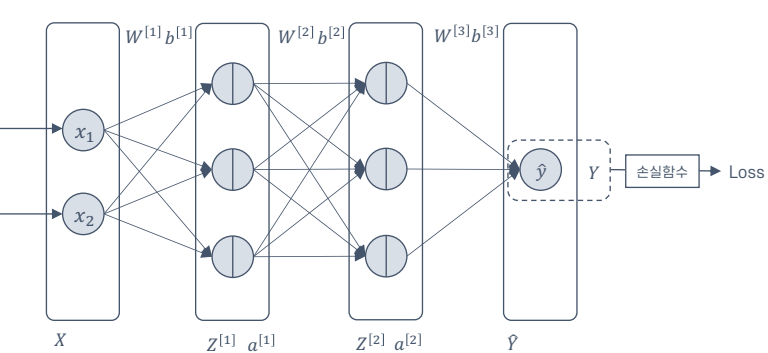
$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

(2) 미리 계산된 미분값과 오차신호로 학습에 필요한 그래디언트 계산

역전파를 통해 알아야 하는 점 (1) 학습 순서



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $Z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $Z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$$

$$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

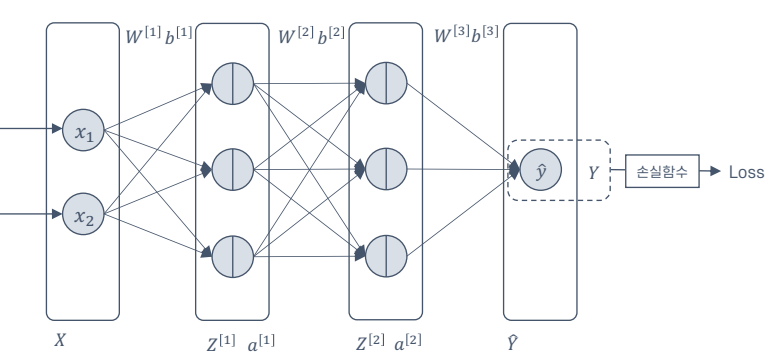
$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

1. 순전파

2. 역전파

역전파를 통해 알아야 하는 점 (2) 메모리



입력층 : X

첫번째 은닉층의 가중치 : $W^{[1]}, b^{[1]}$

첫번째 은닉층의 로짓 값: $Z^{[1]}$

첫번째 은닉층의 활성화 값: $a^{[1]}$

두번째 은닉층의 가중치 : $W^{[2]}, b^{[2]}$

두번째 은닉층의 로짓 값: $Z^{[2]}$

두번째 은닉층의 활성화 값: $a^{[2]}$

출력층의 가중치 : $W^{[3]}, b^{[3]}$

출력층 : \hat{Y}

순전파

$$Z^{[1]} = XW^{[1]} + b^{[1]}$$

$$a^{[1]} = \text{relu}(Z^{[1]})$$

$$Z^{[2]} = a^{[1]}W^{[2]} + b^{[2]}$$

$$a^{[2]} = \text{relu}(Z^{[2]})$$

$$\hat{Y} = a^{[2]}W^{[3]} + b^{[3]}$$

$$\text{Loss} = \frac{1}{2}(\hat{y} - y)^2$$

각층의 기울기

$$\frac{dZ^{[1]}}{dW^{[1]}} = X^T$$

$$\frac{dZ^{[1]}}{db^{[1]}} = 1$$

$$\frac{dZ^{[1]}}{dX} = W^{T[1]}$$

$$\frac{da^{[1]}}{dZ^{[1]}} = \begin{cases} 0, & z^{[1]} < 0 \\ 1, & z^{[1]} \geq 0 \end{cases}$$

$$\frac{dZ^{[2]}}{dW^{[2]}} = a^{T[1]}$$

$$\frac{dZ^{[2]}}{db^{[2]}} = 1$$

$$\frac{dZ^{[2]}}{da^{[1]}} = W^{T[2]}$$

$$\frac{da^{[2]}}{dZ^{[2]}} = \begin{cases} 0, & z^{[2]} < 0 \\ 1, & z^{[2]} \geq 0 \end{cases}$$

$$\frac{d\hat{Y}}{dW^{[3]}} = a^{T[2]}$$

$$\frac{d\hat{Y}}{db^{[3]}} = 1$$

$$\frac{d\hat{Y}}{da^{[2]}} = W^{T[3]}$$

$$\frac{dL}{d\hat{Y}} = \hat{Y} - Y$$

역전파

$$\frac{dL}{dW^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dW^{[1]}}$$

$$\frac{dL}{db^{[1]}} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{db^{[1]}}$$

$$\frac{dL}{dX} = \frac{dL}{dZ^{[1]}} \frac{dZ^{[1]}}{dX}$$

$$\frac{dL}{dZ^{[1]}} = \frac{dL}{da^{[1]}} \frac{da^{[1]}}{dZ^{[1]}}$$

$$\frac{dL}{dW^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{dW^{[2]}}$$

$$\frac{dL}{db^{[2]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{db^{[2]}}$$

$$\frac{dL}{da^{[1]}} = \frac{dL}{dZ^{[2]}} \frac{dZ^{[2]}}{da^{[1]}}$$

$$\frac{dL}{dZ^{[2]}} = \frac{dL}{da^{[2]}} \frac{da^{[2]}}{dZ^{[2]}}$$

$$\frac{dL}{dW^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{dW^{[3]}}$$

$$\frac{dL}{db^{[3]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{db^{[3]}}$$

$$\frac{dL}{da^{[2]}} = \frac{dL}{d\hat{Y}} \frac{d\hat{Y}}{da^{[2]}}$$

각 층에서의 기울기 정보(특히 **출력 값**)을 저장해야 함 -> 메모리 많이 필요