투빅스 20기 정규세션3주차

Neural Network Basic Assignment 1

이름: 황태연

1. Sigmoid Function을 z에 대해 미분하세요.

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

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

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

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

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

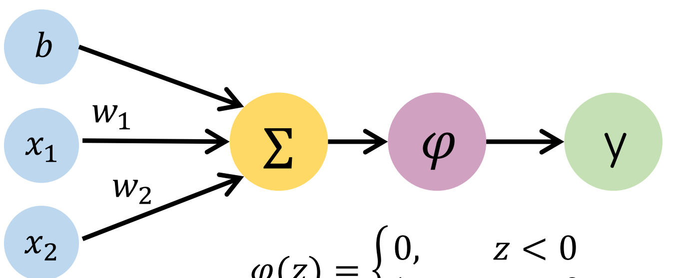
$= \frac{1+e^{-z}-1}{(1+e^{-z})^2}$

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

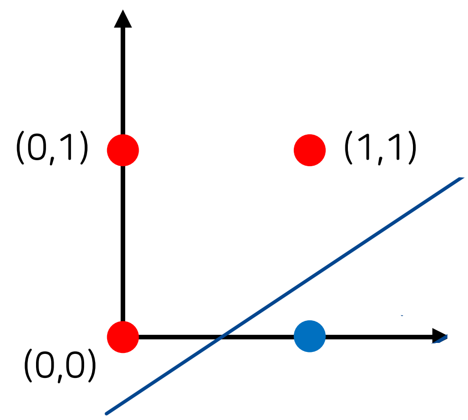
$= \sigma(z)\cdot(1-\sigma(z)).$

$\therefore \frac{d\sigma}{dz} = \sigma(z)\cdot(1-\sigma(z)).$ □

2. 다음과 같은 구조의 Perceptron과 ●(=1), ●(=0)을 평면좌표상에 나타낸 그림이 있습니다.



$$\varphi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$



2-1. ●, ● 를 분류하는 임의의 b, w를 선정하고 분류해보세요.

$y = \varphi(w_1x_1 + w_2x_2 + b)$ $= 0$ $x_1 - 1.1$

Take $w_1 = -1.0$, $w_2 = 1.0$, $b = -0.1$

Then

$x_1=0, x_2=0 \Rightarrow y = \varphi((-1)\cdot 0 + 1\cdot 0 - 0.1) = \varphi(-0.1) = 0$

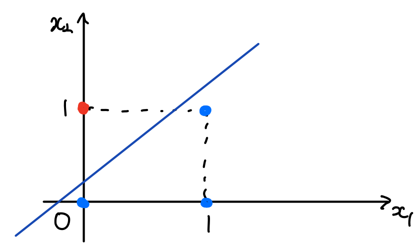
$x_1=0, x_2=1 \Rightarrow y = \varphi((-1)\cdot 0 + 1\cdot 1 - 0.1) = \varphi(0.9) = 1$

$x_1=1, x_2=0 \Rightarrow y = \varphi((-1)\cdot 1 + 1\cdot 0 - 0.1) = \varphi(-1.1) = 0$

$x_1=1, x_2=1 \Rightarrow y = \varphi((-1)\cdot 1 + 1\cdot 1 - 0.1) = \varphi(-0.1) = 0$

We can get the result:

x_1	x_2	y
0	0	0
0	1	1
1	0	0
1	1	0



2-2. Perceptron 학습 규칙에 따라 임의의 학습률을 정하고 b, w를 1회 업데이트 해주세요.

Take $\eta = 0.1$

Then

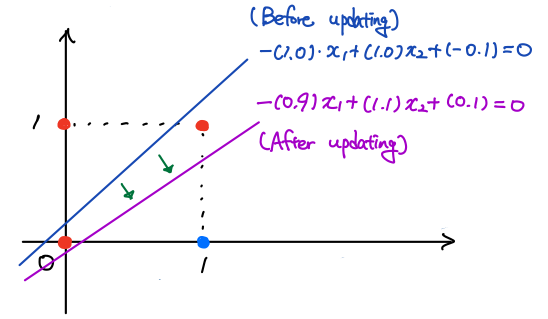
x_1	x_2	output o	target y	
0	0	0	1	①
0	1	1	1	
1	0	0	0	
1	1	0	1	②

Using: $b \leftarrow b + \eta(y - o)\cdot 1$ (output)
$w_i \leftarrow w_i + \eta(y-o)x_i$

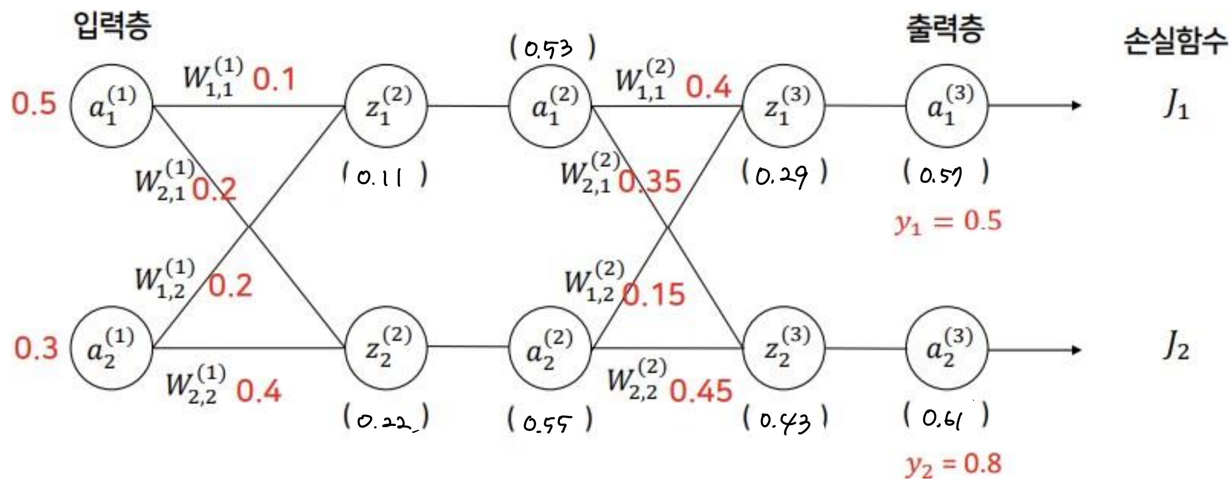
① $b \leftarrow (-0.1) + 0.1(1-0)\cdot 1$ $\therefore b = 0.0$
$w_1 \leftarrow (-1.0) + 0.1(1-0)\cdot 0$ $\therefore w_1 = -1.0$
$w_2 \leftarrow (1.0) + 0.1(1-0)\cdot 0$ $\therefore w_2 = 1.0$

② $b \leftarrow (0.0) + 0.1(1-0)\cdot 1$ $\therefore b = 0.1$
$w_1 \leftarrow (-1.0) + 0.1(1-0)\cdot 1$ $\therefore w_1 = -0.9$
$w_2 \leftarrow (1.0) + 0.1(1-0)\cdot 1$ $\therefore w_2 = 1.1$

$\therefore w_1 = -0.9,\ w_2 = 1.1,\ b = 0.1$



3. 다음과 같이 입력과 가중치가 주어진 퍼셉트론이 있을 때, 아래의 물음에 답해주세요. 모든 문제는 풀이과정을 자세하게 적어주세요! (3-3까지 있습니다.)



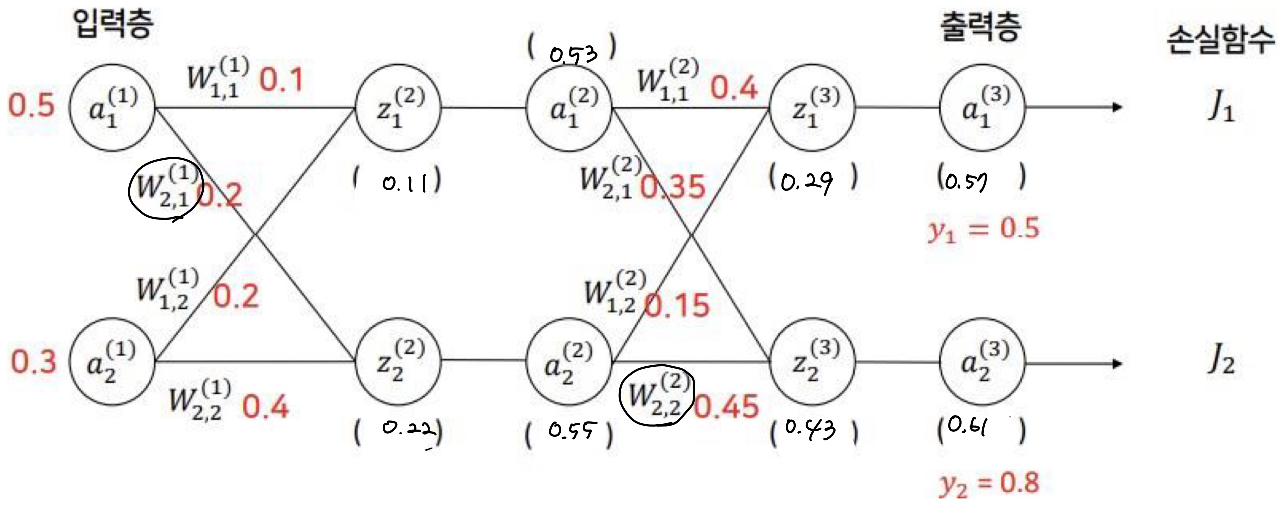
3-1. FeedForward가 일어날 때, 각 노드가 갖는 값을 빈칸에 써주세요. 단, 활성화함수는 sigmoid 함수입니다. (모든 계산의 결과는 소수점 셋째자리에서 반올림하여 둘째자리까지만 써주세요.)

$$z_1^{(2)} = w_{1,1}^{(1)} a_1^{(1)} + w_{1,2}^{(1)} a_2^{(1)} = 0.1 \cdot 0.5 + 0.2 \cdot 0.3 = 0.05 + 0.06 = 0.11$$
$$z_2^{(2)} = w_{2,1}^{(1)} a_1^{(1)} + w_{2,2}^{(1)} a_2^{(1)} = 0.2 \cdot 0.5 + 0.4 \cdot 0.3 = 0.1 + 0.12 = 0.22$$
$$a_1^{(2)} = \sigma(z_1^{(2)}) = \frac{1}{1+e^{-0.11}} \approx 0.53$$
$$a_2^{(2)} = \sigma(z_2^{(2)}) = \frac{1}{1+e^{-0.22}} \approx 0.55$$
$$z_1^{(3)} = w_{1,1}^{(2)} a_1^{(2)} + w_{1,2}^{(2)} a_2^{(2)} = 0.4 \cdot 0.53 + 0.15 \cdot 0.55 \approx 0.29$$
$$z_2^{(3)} = w_{2,1}^{(2)} a_1^{(2)} + w_{2,2}^{(2)} a_2^{(2)} = 0.35 \cdot 0.53 + 0.45 \cdot 0.55 \approx 0.43$$
$$a_1^{(3)} = \sigma(z_1^{(3)}) = \frac{1}{1+e^{-0.29}} \approx 0.57$$
$$a_2^{(3)} = \sigma(z_2^{(3)}) = \frac{1}{1+e^{-0.43}} \approx 0.61 \quad \square$$

3-2. 3-1에서 구한 값을 이용하여 손실함수 J_1과 J_2의 값을 구해주세요. (J_1과 J_2는 반올림하지 말고 써주세요.)

We use MSE as the loss function.

$$J_i = \frac{1}{2}(a_i^{(3)} - y_i)^2$$
$$\therefore J_1 = \frac{1}{2}(a_1^{(3)} - y_1)^2 = \frac{1}{2}(0.57 - 0.5)^2 = 0.00245$$
$$J_2 = \frac{1}{2}(a_2^{(3)} - y_2)^2 = \frac{1}{2}(0.61 - 0.8)^2 = 0.01805 \quad \square$$



3-3. 위에서 구한 값을 토대로, BackPropagation이 일어날 때 $W_{2,2}^{(2)}$과 $W_{2,1}^{(1)}$의 조정된 값을 구해주세요. 단, learning rate는 0.1입니다. (계산 과정에서 소수점 넷째자리에서 반올림하여 셋째자리까지만 써주시고, 마지막 결과인 $W_{2,1}^{(1)}$과 $W_{2,2}^{(2)}$의 값만 반올림하지 말고 써주세요.)

$\underline{J_{total} = \sum_i J_i}$ $\quad \therefore J_{total} = J_1 + J_2$ $\quad$ cf. $(MSE) = \frac{1}{2}\sum_i (a_i^{(3)} - y_i)^2$

i) Update $w_{2,2}^{(2)}$

$$\frac{\partial J_{total}}{\partial w_{2,2}^{(2)}} = \frac{\partial J_{total}}{\partial J_2} \cdot \frac{\partial J_2}{\partial a_2^{(3)}} \cdot \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \cdot \frac{\partial z_2^{(3)}}{\partial w_{2,2}^{(2)}}$$

Since

$$\frac{\partial J_{total}}{\partial J_2} = \frac{\partial (J_1 + J_2)}{\partial J_2} = 1, \quad \frac{\partial J_2}{\partial a_2^{(3)}} = \frac{\partial}{\partial a_2^{(3)}}\left(\frac{1}{2}(a_2^{(3)} - y_2)^2\right) = a_2^{(3)} - y_2,$$

$$\frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} = \frac{\partial}{\partial z_2^{(3)}}\left(\sigma(z_2^{(3)})\right) = \sigma(z_2^{(3)})(1 - \sigma(z_2^{(3)})) = a_2^{(3)}(1 - a_2^{(3)}), \quad (\because \text{Problem } 1)$$

$$\frac{\partial z_2^{(3)}}{\partial w_{2,2}^{(2)}} = \frac{\partial}{\partial w_{2,2}^{(2)}}\left(w_{2,1}^{(2)} a_1^{(2)} + w_{2,2}^{(2)} a_2^{(2)}\right) = a_2^{(2)},$$

$\frac{\partial J_{total}}{\partial w_{2,2}^{(2)}}$ is equal to

$$1 \cdot (a_2^{(3)} - y_2) \cdot a_2^{(3)}(1 - a_2^{(3)}) \cdot a_2^{(2)}.$$

$$= 1 \cdot (0.61 - 0.8) \cdot 0.61 \cdot (1 - 0.61) \cdot 0.55$$

$$\approx -0.025$$

$$\therefore \quad w_{2,2}^{(2)} = w_{2,2}^{(2)} - \eta \frac{\partial J_{total}}{\partial w_{2,2}^{(2)}}$$

$$= 0.45 - 0.1 \cdot (-0.025)$$

$$= 0.45 + 0.0025$$

$$= \underline{0.4525}$$

(Continue on the next page.)

ii) Update $w_{2,1}^{(1)}$

$$\begin{tikzcd}[column sep=small] & J_{total} \arrow[dl, no head] \arrow[dr, no head] & \\ J_1 \arrow[d, dashed, no head] & & J_2 \arrow[d, dashed, no head] \\ w_{2,1}^{(1)} & & w_{2,1}^{(1)} \end{tikzcd}$$

$$\frac{\partial J_{total}}{\partial w_{2,1}^{(1)}} = \frac{\partial J_{total}}{\partial J_1}\frac{\partial J_1}{\partial w_{2,1}^{(1)}} + \frac{\partial J_{total}}{\partial J_2}\frac{\partial J_2}{\partial w_{2,1}^{(1)}}$$

$$= 1\cdot\frac{\partial J_1}{\partial w_{2,1}^{(1)}} + 1\cdot\frac{\partial J_2}{\partial w_{2,1}^{(1)}}$$

$$= 1\cdot\underbrace{\frac{\partial J_1}{\partial a_1^{(3)}}\cdot\frac{\partial a_1^{(3)}}{\partial z_1^{(3)}}}_{\delta_1^{(3)}}\cdot\underbrace{\frac{\partial z_1^{(3)}}{\partial a_2^{(2)}}}_{w_{1,2}^{(2)}}\cdot\underline{\frac{\partial a_2^{(2)}}{\partial z_2^{(2)}}\cdot\frac{\partial z_2^{(2)}}{\partial w_{2,1}^{(1)}}}$$

$$+1\cdot\underbrace{\frac{\partial J_2}{\partial a_2^{(3)}}\cdot\frac{\partial a_2^{(3)}}{\partial z_2^{(3)}}}_{\delta_2^{(3)}}\cdot\underbrace{\frac{\partial z_2^{(3)}}{\partial a_2^{(2)}}}_{w_{2,2}^{(2)}}\cdot\underline{\frac{\partial a_2^{(2)}}{\partial z_2^{(2)}}\cdot\frac{\partial z_2^{(2)}}{\partial w_{2,1}^{(1)}}}$$

$$= \left(\delta_1^{(3)} w_{1,2}^{(2)} + \delta_2^{(3)} w_{2,2}^{(2)}\right)\frac{\partial a_2^{(2)}}{\partial z_2^{(2)}}\cdot\frac{\partial z_2^{(2)}}{\partial w_{2,1}^{(1)}}$$

Since

$$\delta_1^{(3)} = \frac{\partial J_1}{\partial a_1^{(3)}}\cdot\frac{\partial a_1^{(3)}}{\partial z_1^{(3)}}$$

$$= \frac{\partial}{\partial a_1^{(3)}}\left(\frac{1}{2}\cdot(a_1^{(3)} - y_1)^2\right)\cdot\frac{\partial}{\partial z_1^{(3)}}\left(\sigma(z_1^{(3)})\right)$$

$$= (a_1^{(3)} - y_1)\cdot \sigma(z_1^{(3)})\cdot(1-\sigma(z_1^{(3)}))$$

$$= (a_1^{(3)} - y_1)\cdot a_1^{(3)}\cdot(1-a_1^{(3)}) = (0.57-0.5)\cdot 0.57\cdot(1-0.57) \approx 0.017,$$

$$\delta_2^{(3)} = \frac{\partial J_2}{\partial a_2^{(3)}}\cdot\frac{\partial a_2^{(3)}}{\partial z_2^{(3)}}$$

$$= (a_2^{(3)} - y_2)\cdot a_2^{(3)}(1-a_2^{(3)}) = (0.61-0.8)\cdot 0.61(1-0.61) \approx -0.045$$

$$\frac{\partial a_2^{(2)}}{\partial z_2^{(2)}}\cdot\frac{\partial z_2^{(2)}}{\partial w_{2,1}^{(1)}} = \frac{\partial}{\partial z_2^{(2)}}\left(\sigma(z_2^{(2)})\right)\cdot\frac{\partial}{\partial w_{2,1}^{(1)}}\left(w_{2,1}^{(1)}a_1^{(1)} + w_{2,2}^{(1)}a_2^{(1)}\right)$$

$$= a_2^{(2)}\cdot(1-a_2^{(2)})\cdot a_1^{(1)} \approx 0.55\cdot(1-0.55)\cdot 0.5 \approx 0.124$$

$\dfrac{\partial J_{total}}{\partial w_{2,1}^{(1)}}$ is equal to

$$\left(\delta_1^{(3)} w_{1,2}^{(2)} + \delta_2^{(3)} w_{2,2}^{(2)}\right)\frac{\partial a_2^{(2)}}{\partial z_2^{(2)}}\cdot\frac{\partial z_2^{(2)}}{\partial w_{2,1}^{(1)}}$$

$$= \left(0.017\cdot w_{1,2}^{(2)} + (-0.045)\cdot w_{2,2}^{(2)}\right)\cdot 0.124$$

$$= (0.017\cdot 0.15 + (-0.045)\cdot 0.45)\cdot 0.124$$

$$\approx -0.002$$

$$\therefore\quad w_{2,1}^{(1)} = w_{2,1}^{(1)} - \eta\,\frac{\partial J_{total}}{\partial w_{2,1}^{(1)}}$$

$$= 0.2 - 0.1\cdot(-0.002)$$

$$= 0.2 + 0.0002$$

$$= \underline{0.2002}$$

$\square$