투빅스 18기 정규세션 3주차

Neural Network Basic Assignment

이름: 강효은

1. Sigmoid Function을 z에 대해 미분하세요.

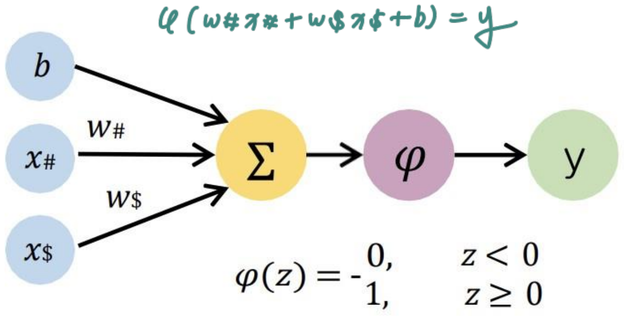
$$\sigma(z) = \frac{1}{1+e^{-z}}$$

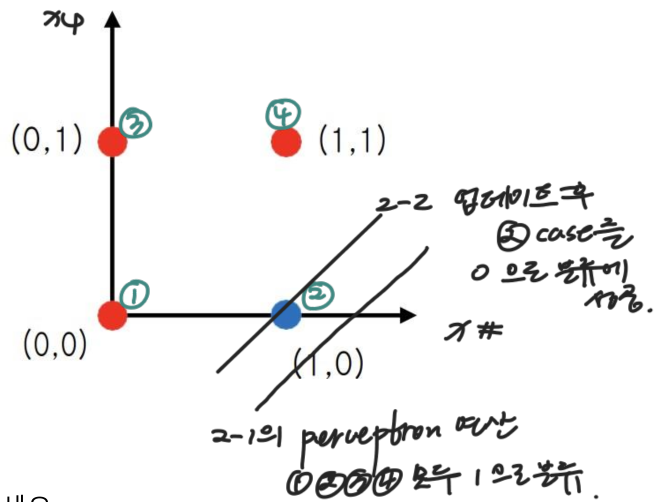
$$\frac{d}{dz}\sigma(z) = \frac{-(-1)\times e^{-z}}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1-1+e^{-z}}{(1+e^{-z})^2}$$

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

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

$\varphi(w_\# x_\# + w_\$ x_\$ + b) = y$





가중치 조정식 $w_i \leftarrow w_i + \eta(y-o)x_i$

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

$w_\# = -1.0$ $w_\$ = 1.0$ $b = 1.2$ 라는 임의의 b, w 값을 설정한다.

① $(-1.0\times0)+(1.0\times0)+1.2 = 1.2$ $\varphi(1.2) = 1$

③ $(-1.0\times0)+(1.0\times1)+1.2 = 2.2$ $\varphi(2.2) = 1$

② $(-1.0\times1)+(1.0\times0)+1.2 = 0.2$ $\varphi(0.2) = 1$

④ $(-1.0\times1)+(1.0\times1)+1.2 = 1.2$ $\varphi(1.2) = 1$

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

임의의 학습률 0.15 설정. 잘못분류된 ②번 케이스에 대하여 업데이트 진행.

$b \leftarrow b + 0.15(0-1)\times1.0$

$w_\# \leftarrow w_\# + 0.15(0-1)\times1.0$

$w_\$ \leftarrow w_\$ + 0.15(0-1)\times0$

* 업데이트 후 분류

① $(-1.15\times0)+(1.0\times0)+1.05 = 1.05$, $\varphi(1.05) = 1$

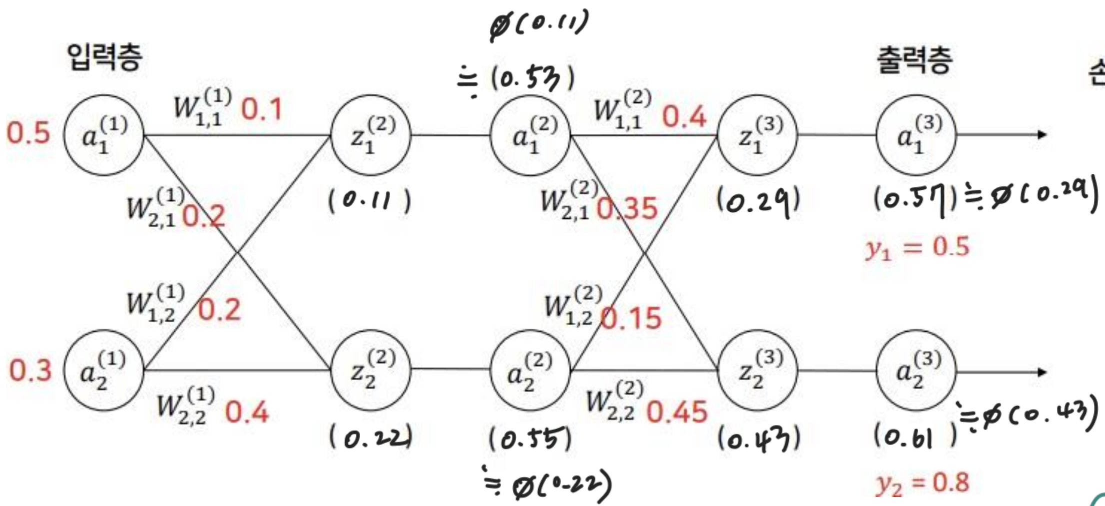
② $(-1.15\times1)+(1.0\times0)+1.05 = -0.1$, $\varphi(-0.1) = 0$

③ $(-1.15\times0)+(1.0\times1)+1.05 = 2.05$, $\varphi(2.05) = 1$

④ $(-1.15\times1)+(1.0\times1)+1.05 = 0.9$, $\varphi(0.9) = 1$

$\Rightarrow$ $b = 1.05$, $w_\# = -1.15$, $w_\$ = 1.0$ 으로 업데이트. ▶ 성공적으로 분류되는것 확인할수 있다.

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



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

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

$z_1^{(2)} = W_{1,1}^{(1)} a_1^{(1)} + W_{1,2}^{(1)} a_2^{(1)} = 0.1 \times 0.5 + 0.2 \times 0.3 = 0.11$ $\quad a_1^{(2)} = \phi(z_1^{(2)}) = \frac{1}{1+e^{0.11}}$

$z_2^{(2)} = W_{2,1}^{(1)} a_1^{(1)} + W_{2,2}^{(1)} a_2^{(1)} = 0.2 \times 0.5 + 0.4 \times 0.3 = 0.22$ $\quad a_2^{(2)} = \phi(z_2^{(2)}) = \frac{1}{1+e^{-0.22}}$

$z_1^{(3)} = W_{1,1}^{(2)} a_1^{(2)} + W_{1,2}^{(2)} a_2^{(2)} = 0.4 \times 0.53 + 0.15 \times 0.55 \fallingdotseq 0.29$ $\quad a_1^{(3)} = \phi(z_1^{(3)}) = \frac{1}{1+e^{-0.29}}$

$z_2^{(3)} = W_{2,1}^{(2)} a_1^{(2)} + W_{2,2}^{(2)} a_2^{(2)} = 0.35 \times 0.53 + 0.45 \times 0.55 \fallingdotseq 0.43$ $\quad a_2^{(3)} = \phi(z_2^{(3)}) = \frac{1}{1+e^{-0.43}}$

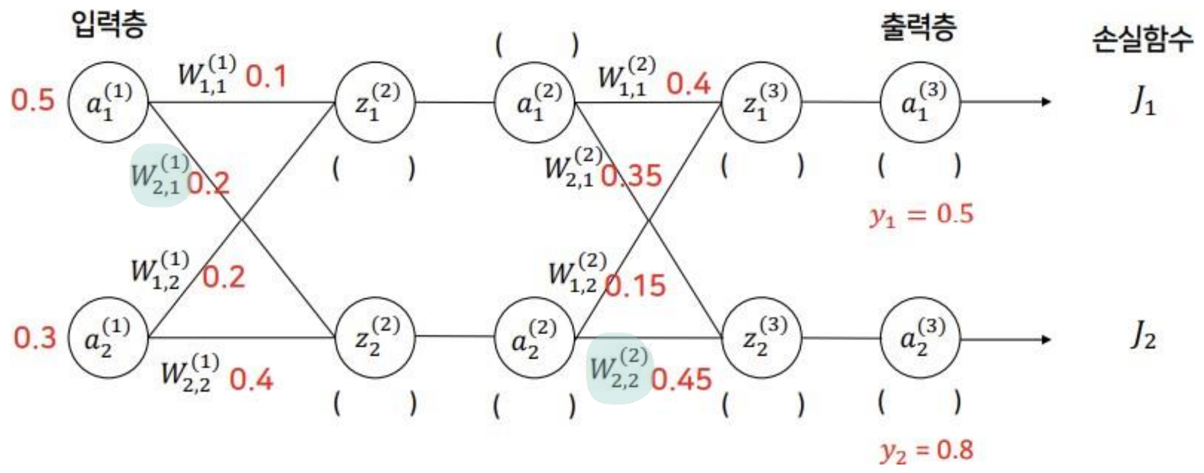
3-2. 3-1에서 구한 값을 이용하여 손실함수 J_1과 J_2의 값을 구해주세요. (J_1과 J_2는 반올림하지 말고 써주세요.)

손실함수 $MSE = \frac{1}{2N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$

$J_1 = \frac{1}{2}(a_1^{(3)} - y_1)^2 = \frac{1}{2}(0.57 - 0.5)^2 = 0.00245$

$J_2 = \frac{1}{2}(a_2^{(3)} - y_2)^2 = \frac{1}{2}(0.61 - 0.8)^2 = 0.01805$

가중치 업데이트식 $w_j = w_j - \eta \frac{\partial J_{total}}{\partial w_j}$



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.)

$\eta = 0.1$

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

$$\frac{\partial J_{total}}{\partial w_{2,2}^{(2)}} = \frac{\partial J_2}{\partial a_2^{(3)}} \times \frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} \times \frac{\partial z_2^{(3)}}{\partial w_{2,2}^{(2)}} \qquad (\because \text{chain rule})$$

i) $\frac{\partial J_2}{\partial a_2^{(3)}} = \frac{1}{2} \frac{\partial}{\partial a_2^{(3)}} (a_2^{(3)} - y_1)^2 = a_2^{(3)} - y_1 = 0.61 - 0.8 = -0.19$

ii) $\frac{\partial a_2^{(3)}}{\partial z_2^{(3)}} = \frac{\partial \phi(z_2^{(3)})}{\partial z_2^{(3)}} = \phi(z_2^{(3)})(1 - \phi(z_2^{(3)})) = a_2^{(3)}(1 - a_2^{(3)}) = 0.61 \times 0.39 \fallingdotseq 0.238$

$(\because \frac{d}{dz}\phi(z) = \phi(z)(1-\phi(z))$, ※1에서 증명$)$

iii) $\frac{\partial z_2^{(3)}}{\partial w_{2,2}^{(2)}} = \frac{\partial (w_{2,1}^{(2)} a_1^{(2)} + w_{2,2}^{(2)} a_2^{(2)})}{\partial w_{2,2}^{(2)}} = a_2^{(2)} = 0.55$

$\Rightarrow w_{2,2}^{(2)} = w_{2,2}^{(2)} - 0.1 \times (-0.19) \times 0.238 \times 0.55$

$= 0.45 + 0.024871 = 0.474871$

$$\boxed{\therefore w_{2,2}^{(2)} = 0.474871}$$

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

$$\frac{\partial J_{total}}{\partial w_{2,1}^{(1)}} = \frac{\partial J_{total}}{\partial a_2^{(2)}} \times \frac{\partial a_2^{(2)}}{\partial z_2^{(2)}} \times \frac{\partial z_2^{(2)}}{\partial w_{2,1}^{(1)}}$$

i) $$\frac{\partial J_{total}}{\partial a_2^{(2)}} = \frac{\partial J_1}{\partial a_2^{(2)}} + \frac{\partial J_2}{\partial a_2^{(2)}} = \frac{\partial J_1}{\partial z_1^{(3)}} \times \frac{\partial z_1^{(3)}}{\partial a_2^{(2)}} + \frac{\partial J_2}{\partial z_2^{(3)}} \times \frac{\partial z_2^{(3)}}{\partial a_2^{(2)}}$$

$$= \delta_1^{(3)} \times \frac{\partial (w_{1,1}^{(2)} a_1^{(2)} + w_{1,2}^{(2)} a_2^{(2)})}{\partial a_2^{(2)}} + \delta_2^{(3)} \times \frac{\partial (w_{2,1}^{(2)} a_1^{(2)} + w_{2,2}^{(2)} a_2^{(2)})}{\partial a_2^{(2)}}$$

$$= \delta_1^{(3)} \times w_{1,2}^{(2)} + \delta_2^{(3)} \times w_{2,2}^{(2)}$$

이때, $$\delta_1^{(3)} = \frac{\partial J_1}{\partial z_1^{(3)}} = (a_1^{(3)} - y_1) \times a_1^{(3)}(1 - a_1^{(3)}) = 0.07 \times 0.57 \times 0.43 \fallingdotseq 0.017$$

$$\delta_2^{(3)} = \frac{\partial J_2}{\partial z_2^{(3)}} = (a_2^{(3)} - y_2) \times a_2^{(3)}(1 - a_2^{(3)}) = 0.19 \times 0.43 \times 0.57 \fallingdotseq 0.047$$

ii) $$\frac{\partial a_2^{(2)}}{\partial z_2^{(2)}} = a_2^{(2)}(1 - a_2^{(2)}) = 0.53 \times 0.47 \fallingdotseq 0.249$$

iii) $$\frac{\partial z_2^{(2)}}{\partial w_{2,1}^{(1)}} = \frac{\partial (w_{2,1}^{(1)} a_1^{(1)} + w_{2,2}^{(1)} a_2^{(1)})}{\partial w_{2,1}^{(1)}} = a_1^{(1)} = 0.5$$

$$\Rightarrow \frac{\partial J_{total}}{\partial w_{2,1}^{(1)}} = (\delta_1^{(3)} \times w_{1,2}^{(2)} + \delta_2^{(3)} \times w_{2,2}^{(2)}) \times a_2^{(2)}(1 - a_2^{(2)}) \times a_1^{(1)}$$

$$= (0.017 \times 0.15 + 0.047 \times 0.45) \times 0.249 \times 0.5$$

$$= 0.00295065 \fallingdotseq 0.003$$

$$\rightarrow w_{2,1}^{(1)} = w_{2,1}^{(1)} - 0.1 \times 0.003 = 0.2 - 0.0003 = 0.1997$$

$$\boxed{\therefore w_{2,1}^{(1)} = 1.997}$$