

# Neural Network Basic Assignment

이름: 이 선민

1. Sigmoid Function을  $z$ 에 대해 미분하세요.

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

① 먼저  $z$ 에 대해 미분합니다.

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

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

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

$$\therefore \sigma'(x) = \frac{e^{-z}}{(1 + e^{-z})^2}$$

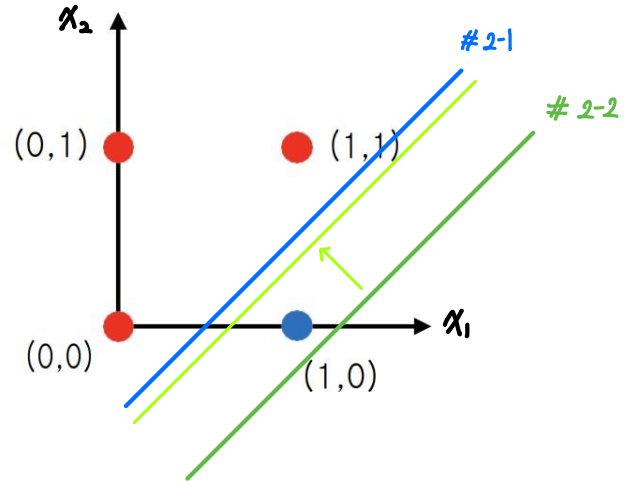
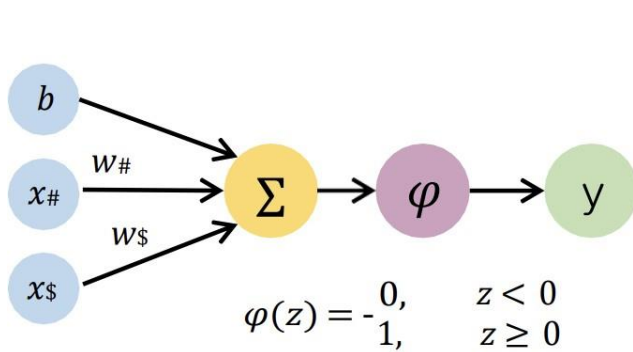
② 다음 식을 정리합니다.

$$\frac{e^{-z}}{(1 + e^{-z})^2} = \underbrace{\frac{1}{(1 + e^{-z})}}_{\sigma(x)} \times \underbrace{\frac{1 + e^{-z} - 1}{(1 + e^{-z})}}_{(1 - \sigma(x))}$$

$$= \sigma(x) \times (1 - \sigma(x))$$

$$\therefore \sigma'(x) = \sigma(x) \times (1 - \sigma(x))$$

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



2-1. ●, ●을 분류하는 임의의  $b, w$ 를 선정하고 분류해보세요.

$$w_1 = -1.0, w_2 = 1.0, b = 0.5$$

$$\textcircled{1} (-1.0 \times 0) + (1.0 \times 0) + (0.5) = 0.5$$

$$\varphi((-1.0 \times 0) + (1.0 \times 0) + (0.5)) = 1$$

$$\textcircled{2} (-1.0 \times 0) + (1.0 \times 1) + (0.5) = 1.5$$

$$\varphi((-1.0 \times 0) + (1.0 \times 1) + (0.5)) = 1$$

$$\textcircled{3} (-1.0 \times 1) + (1.0 \times 1) + (0.5) = 0.5$$

$$\varphi((-1.0 \times 1) + (1.0 \times 1) + (0.5)) = 1$$

$$\textcircled{4} (-1.0 \times 1) + (1.0 \times 0) + (0.5) = -0.5$$

$$\varphi((-1.0 \times 1) + (1.0 \times 0) + (0.5)) = 0$$

제대로 분류되었음을 알 수 있습니다.

| $x_1$ | $x_2$ | $S$  | $y$ |
|-------|-------|------|-----|
| 0     | 0     | -1.5 | 1   |
| 0     | 1     | -2.5 | 1   |
| 1     | 1     | -1.5 | 1   |
| 1     | 0     | -0.5 | 0   |

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

Learning rate가 너무 작으면 학습 속도가 느려지고, 너무 크면 오차를 잡을 수 없으므로, 임의로  $\eta = 0.05$ 로 설정합니다.

위에서 분류가 제대로 되었으므로, 파라미터 업데이트 과정을 공부하기 위해, 분류가 제대로 이루어지지 않은  $w_1, w_2, b$ 를 설정해보았습니다.

$$w_1 = -0.55, w_2 = 0.55, b = 0.55$$

| $x_1$ | $x_2$ | $0$ | $y$ |
|-------|-------|-----|-----|
| 0     | 0     | 1   | 1   |
| 0     | 1     | 1   | 1   |
| 1     | 1     | 1   | 1   |
| 1     | 0     | 1   | 0   |

$$\begin{aligned} b &\leftarrow 0.55 + 0.05(0-1) \times 1 = 0.5 \\ w_1 &\leftarrow -0.55 + 0.05(0-1) \times 1 = -0.6 \\ w_2 &\leftarrow 0.55 + 0.05(0-1) \times 0 = 0.55 \end{aligned}$$

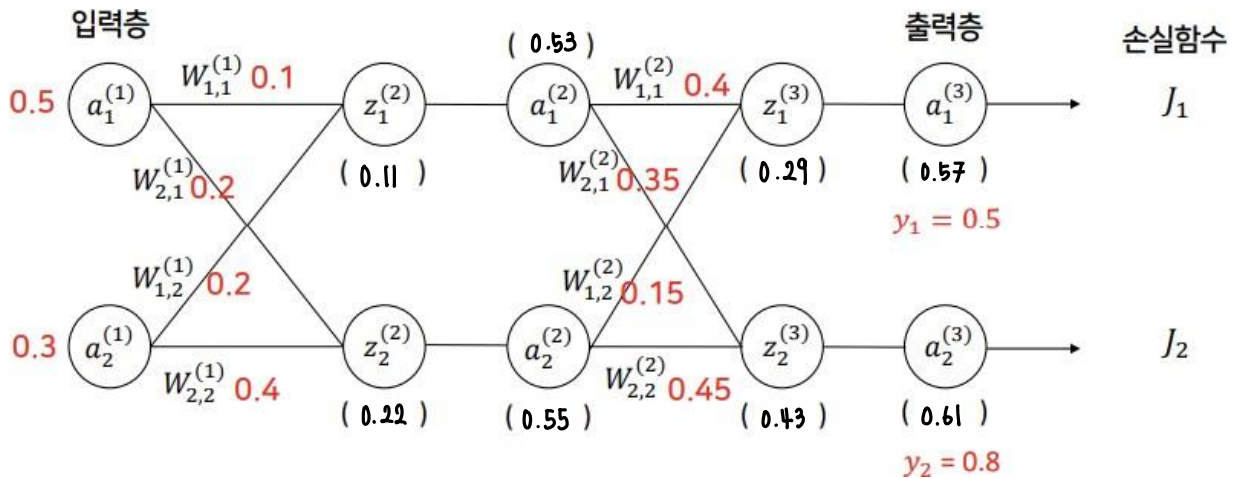
$b, w_1, w_2$  1회 업데이트

$$\therefore (-0.6 \times 1) + (0.55 \times 0) + (0.5) = -0.1$$

$$y = \varphi((-0.6 \times 1) + (0.55 \times 0) + (0.5)) = 0$$

$\Rightarrow$  가중치를 조정하여 파라미터를 업데이트하여 제대로 분류되도록 하였습니다.

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.)

Activation function 이 sigmoid 함수이므로,  $\phi(z) = \frac{1}{1+e^{-z}}$  입니다.

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

$$\textcircled{2} a_1^{(2)} = \phi(z_1^{(2)}) \\ = \frac{1}{1+e^{-0.11}} \\ = 0.53$$

$$\textcircled{3} z_2^{(2)} = W_{2,1}^{(1)} a_1^{(1)} + W_{2,2}^{(1)} a_2^{(1)} \\ = 0.2 \times 0.5 + 0.4 \times 0.3 \\ = 0.22$$

$$\textcircled{4} a_2^{(2)} = \phi(z_2^{(2)}) \\ = \frac{1}{1+e^{-0.22}} \\ = 0.55$$

$$\textcircled{5} z_1^{(3)} = W_{1,1}^{(2)} a_1^{(2)} + W_{1,2}^{(2)} a_2^{(2)} \\ = 0.4 \times 0.53 + 0.15 \times 0.55 \\ = 0.29$$

$$\textcircled{6} a_1^{(3)} = \phi(z_1^{(3)}) \\ = \frac{1}{1+e^{-0.29}} \\ = 0.57$$

$$\textcircled{7} z_2^{(3)} = W_{2,1}^{(2)} a_1^{(2)} + W_{2,2}^{(2)} a_2^{(2)} \\ = 0.35 \times 0.53 + 0.45 \times 0.55 \\ = 0.43$$

$$\textcircled{8} a_2^{(3)} = \phi(z_2^{(3)}) \\ = \frac{1}{1+e^{-0.43}} \\ = 0.61$$

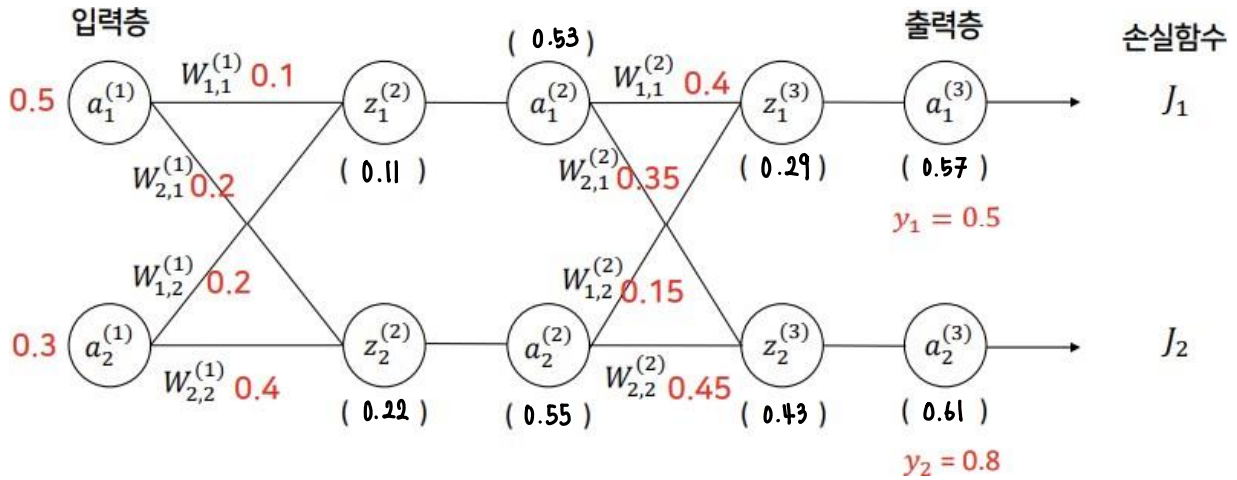
- 3-2. 3-1에서 구한 값을 이용하여 손실함수  $J_1$ 과  $J_2$ 의 값을 구해주세요. ( $J_1$ 과  $J_2$ 는 반올림하지 말고 써주세요.)

Cost function 이 MSE 이므로, 식은  $\frac{1}{2N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$  입니다.

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

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

$\therefore J_1 = 0.00245, J_2 = 0.01805$ 의 loss 값을 갖습니다.



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

Learning rate = 0.1 이므로,  $w_j = w_j - 0.1 \times \frac{\Delta J_{total}}{\Delta w_j}$  입니다.

①  $\delta_2^{(3)}$  을 먼저 구해봅시다.

$$\begin{aligned} \delta_2^{(3)} &= \frac{\Delta J_2}{\Delta z_2^{(3)}} \\ &= (a_2^{(3)} - y_2) \times a_2^{(3)} \times (1 - a_2^{(3)}) \\ &= (0.61 - 0.8) \times 0.61 \times (1 - 0.61) \\ &= -0.045 \end{aligned}$$

②  $\delta_2^{(3)}$  을 이용해서  $w_{2,1}^{(2)}$  을 구해봅시다.

$$\begin{aligned} w_{2,1}^{(2)} &= w_{2,1}^{(2)} - \delta_2^{(3)} a_1^{(2)} \\ &= 0.35 - (-0.045) \times 0.53 \\ &= 0.37385 \end{aligned}$$

③  $\delta_2^{(3)}$  을 이용해서  $w_{2,2}^{(2)}$  을 구해봅시다.

$$\begin{aligned} w_{2,2}^{(2)} &= w_{2,2}^{(2)} - \delta_2^{(3)} a_2^{(2)} \\ &= 0.45 - (-0.045) \times 0.55 \\ &= 0.47475 \end{aligned}$$

$\therefore$  Back propagation이 일어날 때, 조정된  $w_{2,1}^{(2)}$ 의 값은 0.37385이고, 조정된  $w_{2,2}^{(2)}$ 의 값은 0.47475입니다.