

## Neural Network 기초 Assignment 1

이름: 김상현

## Part 1. 함수 (20 points)

1. Sigmoid를  $z$ 에 대해 미분하세요. (2 points)

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

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

2. Mean Square Error를  $w_i$ 에 대해 편미분하세요. (3 points)

$$MSE = J(W) = \frac{1}{2} \sum_{k=1}^K (y_k - o_k)^2$$

$$o_i = w_i^T x + b_i, (1 \leq i \leq K)$$

$$\frac{\partial MSE}{\partial w_i} = \frac{\partial MSE}{\partial o_i} \cdot \frac{\partial o_i}{\partial w_i} = -(y_i - o_i) \cdot x$$

3. Logistic Regression의 Log Likelihood를  $w_j$ 에 대해 편미분하세요. (3 points)

$$\log likelihood = J(W) = - \sum_{k=1}^K \{y_k \log p_k + (1 - y_k) \log(1 - p_k)\}$$

$$p_j = \sigma(z_j), z_j = w_j^T x + b_j, (1 \leq j \leq K)$$

$$\begin{aligned} \frac{\partial J}{\partial w_j} &= \frac{\partial J}{\partial p_j} \cdot \frac{\partial p_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_j} \\ &= -\left(\frac{y_j}{p_j} + \frac{(y_j - 1)}{(1 - p_j)}\right) \cdot \frac{e^{-z_j}}{(1 + e^{-z_j})^2} \cdot x \\ &= -(y_j - p_j) \cdot x \end{aligned}$$

$$\begin{cases} p_j = \frac{e^{-z_j}}{1 + e^{-z_j}} \\ 1 - p_j = \frac{1}{1 + e^{-z_j}} \end{cases}$$

4. 다음 식이 올바른 이유를 증명하세요. (5 points)

$$-\sum_{k=1}^K y_k \log(p_k) = -\log p_i, (1 \leq i \leq K)$$

$$-\sum_{k=1}^K y_k \log(p_k)$$

$$= -[y_1 \log p_1 + y_2 \log p_2 + \dots + y_K \log(p_K)]$$

$$= -\log p_i \quad (\because Y = (0, 0, \dots, y_i, 0, \dots, 0), y_i = 1).$$

5. Softmax-Cross Entropy를  $z_i$ 에 대해 편미분하세요. (7 points)

$$CE = -\sum_{k=1}^K y_k \log(p_k), \quad p_i = \frac{e^{z_i}}{\sum e^{z_k}}$$

$$\frac{\partial J}{\partial z_i} = -\sum_j \left\{ y_j \frac{\partial}{\partial p_j} \log(p_j) \frac{\partial p_j}{\partial z_i} \right\}$$

$$= -\sum_j \left\{ \frac{y_j}{p_j} \frac{\partial p_j}{\partial z_i} \right\}$$

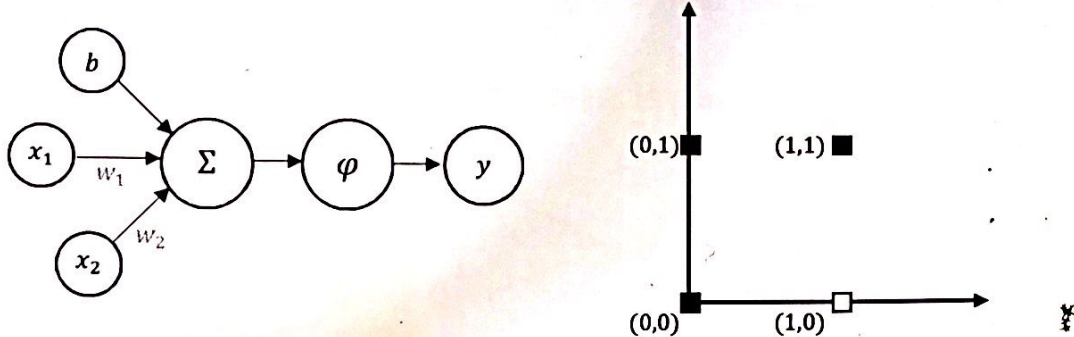
$$= -y_i(1-p_i) + \sum_{j \neq i} y_j p_j$$

$$= p_i - y_i = -(y_i - p_i)$$

$$\begin{cases} \frac{\partial p_i}{\partial z_i} = p_i(1-p_i) \\ \frac{\partial p_i}{\partial z_j} = -p_i p_j \end{cases}$$

## Part 2. 퍼셉트론 (15 points)

다음과 같은 구조의 퍼셉트론과  $\blacksquare (=1)$ ,  $\square (=0)$ 을 평면좌표상에 나타낸 그림이 있습니다.



1.  $\blacksquare$ ,  $\square$ 를 분류하는 임의의  $b$ ,  $w$ 를 선정하고 분류하는 과정을 보이세요. (5 points)

$$b=1, w_0=0.4, w_1=-0.6, w_2=0.6$$

$$\varphi(0.4 + (-0.6) \cdot 0 + (0.6) \cdot 0) = 1$$

$$\varphi(0.4 + (-0.6) \cdot 1 + (0.6) \cdot 0) = -1$$

$$\varphi(0.4 + (-0.6) \cdot 0 + (0.6) \cdot 1) = 1$$

$$\varphi(0.4 + (-0.6) \cdot 1 + (0.6) \cdot 1) = 1$$

2. Perceptron 학습 규칙에 따라 임의의 학습률  $\eta$ 을 정하고  $b$ ,  $w$ 를 한 번 업데이트해 주세요. (5 points)

$$\eta=0.05 \quad w_0=0.4, w_1=-0.6, w_2=0.6$$

$$w_0 \leftarrow w_0 + 0.05((1) - (1)) \cdot 1$$

$$w_1 \leftarrow w_1 + 0.05((1) - (1)) \cdot 0 \quad (0,0)$$

$$w_2 \leftarrow w_2 + 0.05((1) - (1)) \cdot 0$$

$$\Rightarrow w_0=0.4, w_1=-0.6, w_2=0.6$$

3. Adaline Gradient Descent에 따라 임의의 학습률  $\eta$ 을 정하고  $b$ ,  $w$ 를 한 번 업데이트해 주세요. (5 points)

$$\eta=0.05 \quad W^T = (0.4 \ -0.6 \ 0.6) \quad x' = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$W^T x' = 0.4 \quad y' - W^T x' = 0.6$$

$$w_0 \leftarrow w_0 + 0.05 \cdot 0.6 \cdot 1 = 0.43$$

$$w_1 \leftarrow w_1 + 0.05 \cdot 0.6 \cdot 0 = w_1$$

$$w_2 \leftarrow w_2 + 0.05 \cdot 0.6 \cdot 0 = w_2$$

Tobig's 13기 이지용

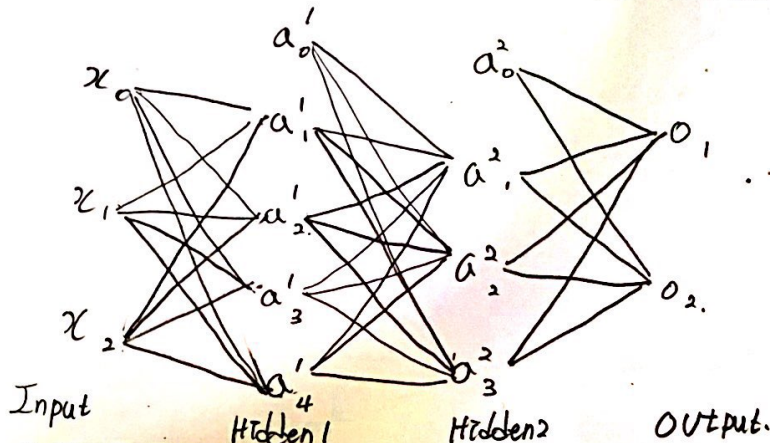
$$\Rightarrow w_0=0.43, w_1=-0.6, w_2=0.6$$



## Part 3. 다층 퍼셉트론 (30 points)

Input Layer가 2차원, 첫 번째 Hidden Layer가 4차원 첫 번째 활성화 함수가 ReLU, 두 번째 Hidden Layer가 3차원, 두 번째 활성화 함수가 Sigmoid, Output Layer가 2차원인 다층 퍼셉트론 구조의 신경망이 있습니다.

1. 위 신경망의 구조를 간략하게 그림으로 그리세요. (5 points)



2. Bias를 포함하여 각 Layer에 존재하는 Weight(Parameter)의 개수와 전체 Weight의 개수를 구하세요. (10 points)

$$\text{Input} \rightarrow \text{Hidden1} : 12$$

$$\text{Hidden1} \rightarrow \text{Hidden2} : 15$$

$$\text{Hidden2} \rightarrow \text{Output} : 8$$

$\Rightarrow 35 \text{ 개}$

3. 위 신경망을 식으로 나타낼 때 필요한 함수, 벡터와 행렬을 정의하고 순전파 과정을 행렬 식으로 표현하세요. (ex) input:  $x = (x_1, x_2, \dots)^T$ ,  $x$ 는 4x1차원 (15 points)

$$x = (x_1, x_2)^T \quad W^1 = \begin{pmatrix} w_{11}^1 & w_{12}^1 & w_{13}^1 & w_{14}^1 \\ w_{21}^1 & w_{22}^1 & w_{23}^1 & w_{24}^1 \end{pmatrix}^T \quad B^1 = (b_1^1, b_2^1, b_3^1, b_4^1)^T$$

$$a^1 = f(W^1 x + B^1)$$

$$W^2 = \begin{pmatrix} w_{11}^2 & w_{12}^2 & w_{13}^2 \\ w_{21}^2 & w_{22}^2 & w_{23}^2 \\ w_{31}^2 & w_{32}^2 & w_{33}^2 \end{pmatrix}^T \quad B^2 = (b_1^2, b_2^2, b_3^2)^T$$

$$a^2 = \sigma(W^2 a^1 + B^2)$$

$$W^3 = \begin{pmatrix} w_{11}^3 & w_{12}^3 \\ w_{21}^3 & w_{22}^3 \\ w_{31}^3 & w_{32}^3 \end{pmatrix}^T \quad B^3 = (b_1^3, b_2^3)^T$$

$$O = W^3 a^2 + B^3$$

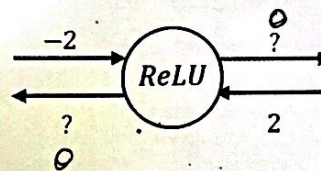
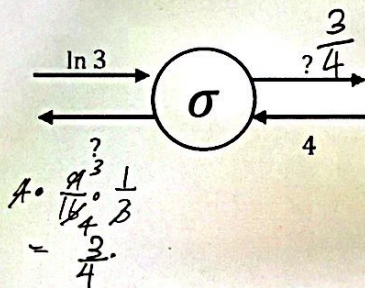
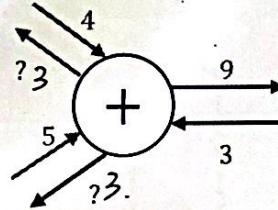
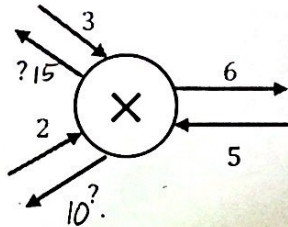
$$f(x_i) = \begin{cases} 0, & x_i < 0 \\ x_i, & x_i \geq 0 \end{cases} \quad x_i \in X$$

$$\sigma(x_i) = \frac{1}{1 + e^{-x_i}} \quad x_i \in X$$



## Part 4. 역전파 (35 points)

1. 다음 그림들의 물음표에 들어갈 숫자를 구하세요. (5 points)



2. 3-3에서 정의한 함수, 벡터와 행렬로 각 Layer에 존재하는 Bias값들의 업데이트를 행렬식으로 표현하세요. (15 points)

$$\delta^3 = \frac{\partial J}{\partial z^3}$$

$$B^3 = w_o^3 a_o^2$$

$$w_o^3 = \begin{pmatrix} w_{o1}^3 \\ w_{o2}^3 \end{pmatrix}$$

$$B^3 \leftarrow B^3 - \eta \delta^3 a_o^2$$

$$\delta^2 = \sigma'(a^1) \cdot (w^3)^T \delta^3$$

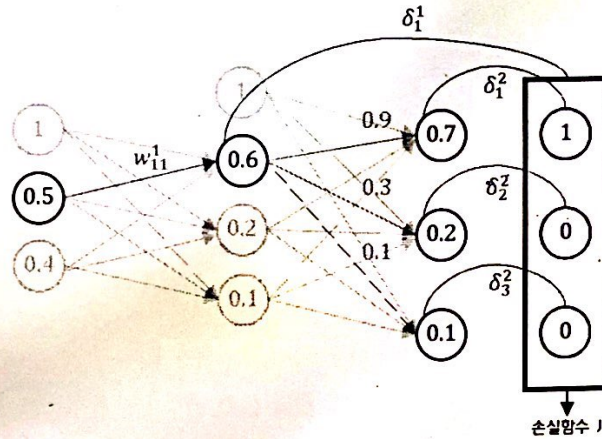
$$B^2 \leftarrow B^2 - \eta \delta^2 a_o^1$$

$$\delta^1 = f'(x) \cdot (w^2)^T \delta^2$$

$$B^1 \leftarrow B^1 - \eta \delta^1 x_o$$



3. 다음 그림에서 Loss Function은 Cross Entropy, Output Layer의 Activation Function은 Softmax, Hidden Layer의 Activation Function은 Sigmoid이고 Learning Rate는 0.05일 때, 각  $\delta$ 의 값과  $w_{11}^1$ 의 변화량을 구하세요. (각 노드의 숫자는 활성화 이후의 숫자입니다.) (15 points)



$$\delta_1^2 = \frac{\partial J}{\partial z_1^2} = -(1 - 0.7) = -0.3$$

$$\delta_2^2 = 0.2$$

$$\delta_3^2 = 0.1$$

$$\delta_1' = \frac{\partial J}{\partial z_1^1} = f'(z_1^1) \sum w_{k1}^2 \delta_k^2$$

$$f'(z_1^1) = f(z_1^1)(1 - f(z_1^1)) = 0.6(1 - 0.6) = 0.24$$

$$\begin{aligned} \therefore \delta_1' &= 0.24 \cdot (-0.3 \cdot 0.9 + 0.2 \cdot 0.3 + 0.1 \cdot 0.1) \\ &= -0.0336 \end{aligned}$$

$$\Delta w_{11}^1 = \eta \delta_1' x_1 = 0.05 \cdot (-0.0336) \cdot 0.5 = -0.00084$$

고생하셨습니다~