

Neural Network 기초 Assignment1

이름: 14기 이지용

Part 1. 함수 (20 points)

1. Sigmoid를 z 에 대해 미분하세요. (2 points)

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

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

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

$$MSE = J(w) = \frac{1}{2} \sum (y_k - o_k)^2, \quad o_k = w_k^T x + b_k$$

$$\frac{\partial J}{\partial w_i} = \frac{\partial J}{\partial o_i} \cdot \frac{\partial o_i}{\partial w_i} = (y_i - o_i) \cdot (-1) \cdot x_i = -(y_i - o_i) x_i$$

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

$$\log likelihood = J(w) = -\sum [y_k \log p_k + (1 - y_k) \log(1 - p_k)],$$

$$p_k = \sigma(z_k), \quad z_k = w_k^T x + b_k$$

$$\begin{aligned} \frac{\partial J}{\partial w_i} &= \frac{\partial J}{\partial p_i} \cdot \frac{\partial p_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_i} = - \left(\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i} \right) \cdot \frac{1}{1+e^{-z_i}} \left(1 - \frac{1}{1+e^{-z_i}} \right) \cdot x_i \\ &\quad \text{변경고} \\ &= - \left(\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i} \right) p_i (1-p_i) x_i = -(y_i - p_i) x_i \end{aligned}$$

4. 다음 식이 올바른 이유를 증명하세요. (5 points)

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

$k=1, -y_1 \log p_1 = -\log p_1 \rightarrow y_1 = 1$

$k=2, -y_1 \log p_1 - y_2 \log p_2 = -\log p_1 \rightarrow (y_1, y_2) = (1, 0)$
 $-y_1 \log p_1 - y_2 \log p_2 = -\log p_2 \rightarrow (y_1, y_2) = (0, 1)$

$k=3, (y_1, y_2, y_3) = (1, 0, 0) \text{ or } (y_1, y_2, y_3) = (0, 1, 0) \text{ or } (y_1, y_2, y_3) = (0, 0, 1)$

k 값에 관계없이 y_k 중 하나만 1이고 나머지는 0이다.

그러므로 위 식은 성립한다.

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

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

$$\frac{\partial CE}{\partial z_i} = \frac{\partial CE}{\partial p_i} \cdot \frac{\partial p_i}{\partial z_i}$$

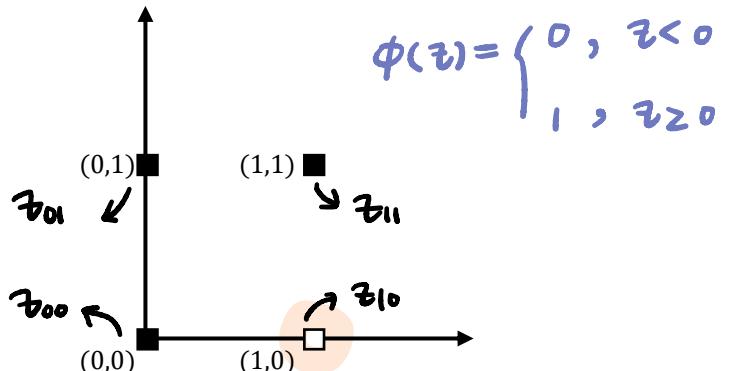
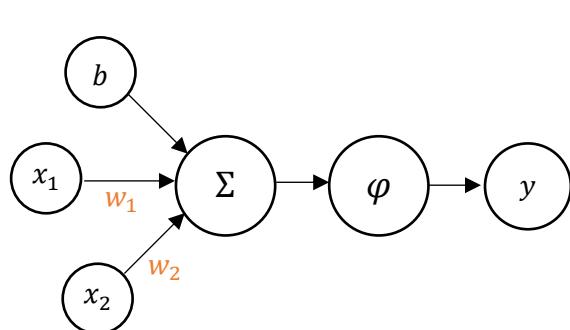
$$\frac{\partial p_i}{\partial z_i} = \frac{\partial \frac{e^{z_i}}{\sum e^{z_k}}}{\partial z_i} / \frac{\partial z_i}{\partial z_i} = \frac{e^{z_i} \sum e^{z_k} - e^{z_i} e^{z_i}}{(\sum e^{z_k})^2} = \frac{e^{z_i}}{\sum e^{z_k}} \left(\frac{\sum e^{z_k} - e^{z_i}}{\sum e^{z_k}} \right)$$

$$= \frac{e^{z_i}}{\sum e^{z_k}} \left(1 - \frac{e^{z_i}}{\sum e^{z_k}} \right) = p_i (1 - p_i)$$

$$\frac{\partial CE}{\partial z_i} = \frac{\partial CE}{\partial p_i} \cdot \frac{\partial p_i}{\partial z_i} = -\frac{y_i}{p_i} \cdot p_i (1 - p_i) = -y_i (1 - p_i)$$

Part 2. 퍼셉트론 (15 points)

다음과 같은 구조의 퍼셉트론과 ■(=1), □(=0)을 평면좌표상에 나타낸 그림이 있습니다.



- , □를 분류하는 임의의 b , w 를 선정하고 분류하는 과정을 보이세요. (5 points)

$$(w_0, w_1, w_2) = (0.4, 0.1, -0.2) / b = -1$$

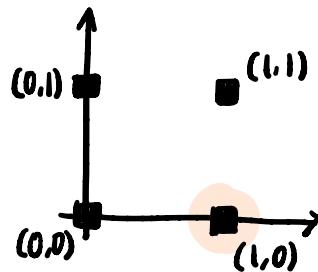
$$z = w_0 + w_1 x_1 + w_2 x_2 = 0.4 + 0.1 x_1 - 0.2 x_2$$

$$z_{00} = 0.4 \rightarrow \phi(z_{00}) = 1$$

$$z_{01} = 0.2 \rightarrow \phi(z_{01}) = 1$$

$$z_{10} = 0.5 \rightarrow \phi(z_{10}) = 1$$

$$z_{11} = 0.3 \rightarrow \phi(z_{11}) = 1$$



- Perceptron 학습 규칙에 따라 임의의 학습률 η 을 정하고 b , w 를 한 번 업데이트해 주세요. (5 points)

$$\hookrightarrow w_j \leftarrow w_j + \eta (y - \phi)x, \quad \eta = 0.05$$

$$\text{ex. } x = (-1, 1, 0) \quad \text{target}$$

$$w_0 + \eta (y - \phi)x_0 = 0.4 + 0.05(0 - 1) \cdot (-1) = 0.45 \rightarrow w_0$$

$$w_1 + \eta (y - \phi)x_1 = 0.1 + 0.05(0 - 1) \cdot 1 = 0.05 \rightarrow w_1$$

$$w_2 + \eta (y - \phi)x_2 = -0.2 + 0.05(0 - 1) \cdot 0 = -0.2 \rightarrow w_2$$

$$\therefore (w_0, w_1, w_2) = (0.45, 0.05, -0.2)$$

- Adaline Gradient Descent에 따라 임의의 학습률 η 을 정하고 b , w 를 한 번 업데이트해 주세요. (5 points)

$$\hookrightarrow w_j \leftarrow w_j + \eta (y^{(i)} - w^T x^{(i)}) x_j^{(i)}, \quad \eta = 0.05$$

$$\text{ex. } x = (-1, 1, 0)$$

$$w_0 + \eta (y^{(i)} - w^T x^{(i)}) x_0^{(i)} = 0.4 + 0.05(0 - (-0.3))(-1) = 0.4 - 0.015 = 0.385 \rightarrow w_0$$

$$w_1 + \eta (y^{(i)} - w^T x^{(i)}) x_1^{(i)} = 0.1 + 0.05(0 - (-0.3)) \cdot 1 = 0.1 + 0.015 = 0.115 \rightarrow w_1$$

$$w_2 + \eta (y^{(i)} - w^T x^{(i)}) x_2^{(i)} = -0.2 + 0.05(0 - (-0.3)) \cdot 0 = -0.2 \rightarrow w_2$$

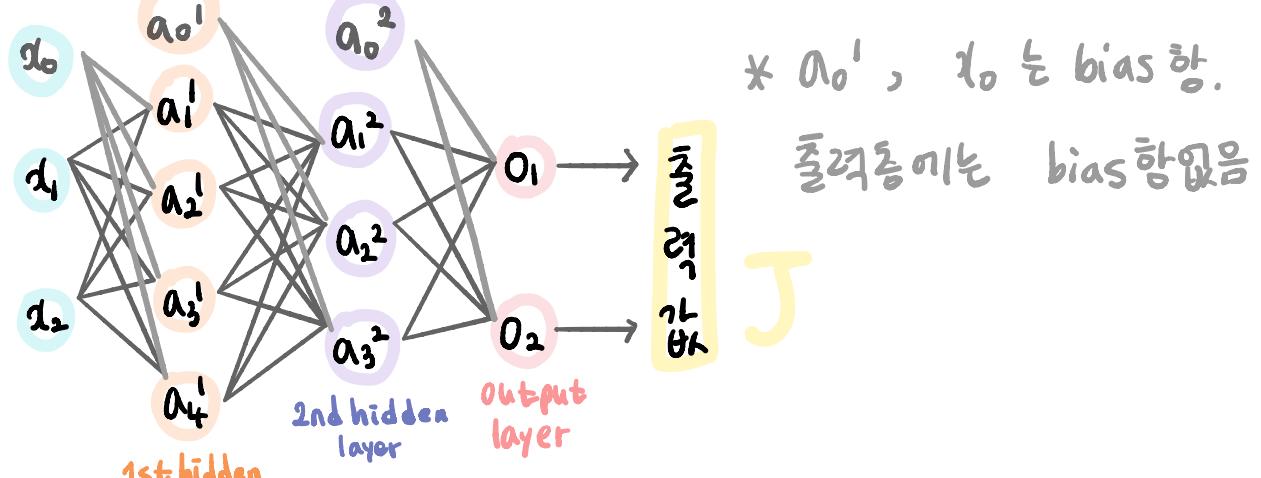
Tobig's 13기 이지용

$$\therefore (w_0, w_1, w_2) = (0.385, 0.115, -0.2)$$

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.

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



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

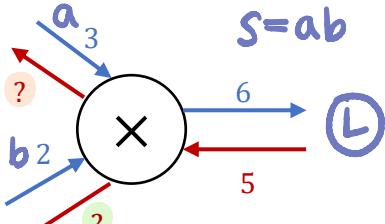
$$\begin{array}{ll} \text{1st hidden layer : } 12개 & \text{총 } 33\text{개} \\ \text{2nd hidden layer : } 15개 & \\ \text{output layer : } 6개 & \end{array}$$

- 위 신경망을 식으로 나타날 때 필요한 함수, 벡터와 행렬을 정의하고 순전파 과정을 행렬식으로 표현하세요. (ex) input: $x^i = (x_1^{(i)}, x_2^{(i)}, \dots)^T$, x 는 4×1 차원) (15 points)

$$\begin{aligned} X &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & b^1 &= \begin{pmatrix} b_1^1 \\ b_2^1 \\ b_3^1 \\ b_4^1 \end{pmatrix} & W^2 &= \begin{pmatrix} w_{11}^2 & w_{12}^2 & w_{13}^2 & w_{14}^2 \\ w_{21}^2 & w_{22}^2 & w_{23}^2 & w_{24}^2 \\ w_{31}^2 & w_{32}^2 & w_{33}^2 & w_{34}^2 \end{pmatrix} & \text{* bias 항은 따로 'b' *} \\ W^1 &= \begin{pmatrix} w_{11}^1 & w_{12}^1 \\ w_{21}^1 & w_{22}^1 \\ w_{31}^1 & w_{32}^1 \\ w_{41}^1 & w_{42}^1 \end{pmatrix} & & & & & \text{* matrix로 표현} \\ z^1 &= W^1 X + b^1 & & & b^2 &= (b_1^2, b_2^2, b_3^2)^T & \vec{z}^3 = W^3 a^2 \\ 4 \times 1 & & & & & & 2 \times 1 \\ a^1 &= \text{ReLU}(z^1) & & & z^2 &= W^2 a^1 + b^2 & g: \text{output activation fn} \\ 4 \times 1 & & & & 3 \times 1 & & \\ & & & & & a^2 &= \sigma(z^2) \\ & & & & & 3 \times 1 & \\ & & & & & & \\ & & & & & & 0 = g(\vec{z}^3) \\ & & & & & & 2 \times 1 \end{aligned}$$

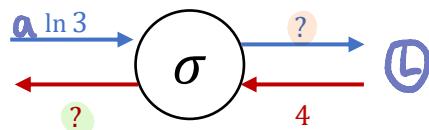
Part 4. 역전파 (35 points)

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



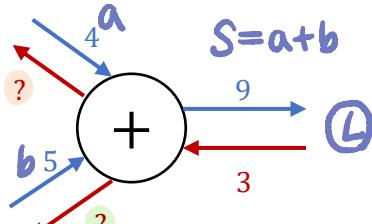
$$\frac{\partial L}{\partial a} = \frac{\partial L}{\partial S} \cdot \frac{\partial S}{\partial a} = 5 \cdot b = 5 \cdot 2 = 10$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial S} \cdot \frac{\partial S}{\partial b} = 5 \cdot a = 5 \cdot 3 = 15$$



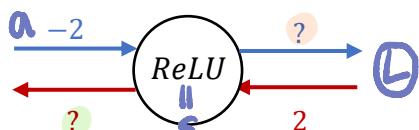
$$\sigma(\ln 3) = \frac{1}{1 + e^{-\ln 3}} = \frac{3}{4}$$

$$\frac{\partial L}{\partial a} = \frac{\partial L}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial a} = 4 \cdot \left[\frac{1}{1 + e^{-a}} \left(1 - \frac{1}{1 + e^{-a}} \right) \Big|_{a=\ln 3} \right] = \frac{3}{4}$$



$$\frac{\partial L}{\partial a} = \frac{\partial L}{\partial S} \cdot \frac{\partial S}{\partial a} = 3 \cdot 1 = 3$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial S} \cdot \frac{\partial S}{\partial b} = 3 \cdot 1 = 3$$



$$S(-2) = 0$$

$$\frac{\partial L}{\partial a} = \frac{\partial L}{\partial S} \cdot \frac{\partial S}{\partial a} = 2 \cdot 0 = 0$$

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

$$\delta^3 = \frac{\partial J}{\partial z^3} = \begin{pmatrix} \frac{\partial J}{\partial z_1^3} \\ \frac{\partial J}{\partial z_2^3} \end{pmatrix} = \begin{pmatrix} \delta_1^3 \\ \delta_2^3 \end{pmatrix}, \quad \delta^3(a^2) = \begin{pmatrix} \delta_1^3 \\ \delta_2^3 \end{pmatrix} (a_1^2 \ a_2^2 \ a_3^2)$$

$$W^3 \leftarrow W^3 - \eta \delta^3(a^2)^T$$

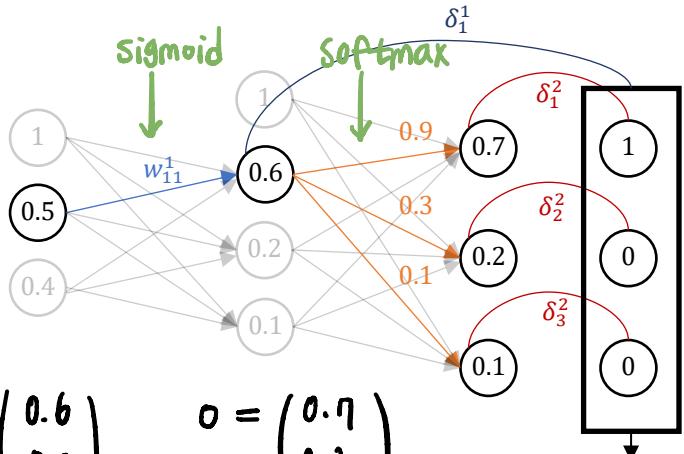
$$\delta^2 = \frac{\partial J}{\partial z^2} = \frac{\partial a^2}{\partial z^2} \frac{\partial z^2}{\partial a^2} \cdot \frac{\partial J}{\partial z^3} = \begin{pmatrix} \sigma'(a_1^2) \\ \sigma'(a_2^2) \\ \sigma'(a_3^2) \end{pmatrix} \odot \underbrace{\begin{pmatrix} w_{11}^3 & w_{21}^3 & \\ w_{12}^3 & w_{22}^3 & \\ w_{13}^3 & w_{23}^3 & \end{pmatrix}}_{(W^3)^T} \begin{pmatrix} \delta_1^3 \\ \delta_2^3 \end{pmatrix} = \begin{pmatrix} \delta_1^2 \\ \delta_2^2 \\ \delta_3^2 \end{pmatrix}$$

$$W^2 \leftarrow W^2 - \eta \delta^2(a^1)^T$$

$$\delta^1 = \frac{\partial J}{\partial z^1} = \frac{\partial a^1}{\partial z^1} \cdot \frac{\partial z^2}{\partial a^1} \cdot \frac{\partial J}{\partial z^3} = \begin{pmatrix} \text{ReLU}'(a_1^1) \\ \text{ReLU}'(a_2^1) \\ \text{ReLU}'(a_3^1) \\ \text{ReLU}'(a_4^1) \end{pmatrix} \odot \begin{pmatrix} w_{11}^2 & w_{21}^2 & w_{31}^2 \\ w_{12}^2 & w_{22}^2 & w_{32}^2 \\ w_{13}^2 & w_{23}^2 & w_{33}^2 \\ w_{14}^2 & w_{24}^2 & w_{34}^2 \end{pmatrix} \begin{pmatrix} \delta_1^2 \\ \delta_2^2 \\ \delta_3^2 \end{pmatrix} = \begin{pmatrix} \delta_1^1 \\ \delta_2^1 \\ \delta_3^1 \\ \delta_4^1 \end{pmatrix}$$

$$W^1 \leftarrow W^1 - \eta \delta^1 X^T$$

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



$$X = \begin{pmatrix} 0.5 \\ 0.4 \end{pmatrix} \quad a^1 = \begin{pmatrix} 0.6 \\ 0.2 \\ 0.1 \end{pmatrix} \quad o = \begin{pmatrix} 0.9 \\ 0.3 \\ 0.1 \end{pmatrix} \quad w_{.1}^2 - \eta \delta^2(a^1)$$

$$(w_{11}^{12} \ w_{21}^{12} \ w_{31}^{12})^T = (0.9 \ 0.3 \ 0.1)^T = w_{.1}^{12}$$

$$\delta^2 = \frac{\partial J}{\partial z^2} = \begin{pmatrix} -1(1-0.9) \\ 0(1-0.3) \\ 0(1-0.1) \end{pmatrix} = \begin{pmatrix} -0.3 \\ 0 \\ 0 \end{pmatrix}$$

$$w_{.1}^{12} - \eta \delta^2 a^1 = \begin{pmatrix} 0.9 \\ 0.3 \\ 0.1 \end{pmatrix} - 0.05 \begin{pmatrix} -0.3 \\ 0 \\ 0 \end{pmatrix} \cdot 0.6 = \begin{pmatrix} 0.909 \\ 0 \\ 0 \end{pmatrix}$$

$\rightarrow w_{.1}^{12}$

$$\delta^1 = \frac{\partial J}{\partial z^1} = \frac{\partial a^1}{\partial z^1} \cdot \frac{\partial z^2}{\partial a^1} \cdot \frac{\partial o}{\partial z^2} \cdot \frac{\partial J}{\partial o} = \begin{pmatrix} \sigma'(z_1^1) \\ \sigma'(z_2^1) \\ \sigma'(z_3^1) \end{pmatrix} \Theta(w_{.1}^{12})^T \delta^2 = \begin{pmatrix} \sigma'(z_1^1) \cdot (-0.2727) \\ 0 \\ 0 \end{pmatrix}$$

$$\delta_1^1 = -0.2727 \sigma'(z_1^1)$$

$$\begin{aligned} w_{11}^1 &\leftarrow w_{11}^1 + \Delta w_{11}^1 = w_{11}^1 - \eta \delta_1^1 d_1 \\ &= w_{11}^1 - 0.05 \times (-0.2727 \sigma'(z_1^1)) \times 0.5 \\ &\approx w_{11}^1 + 0.00682 \sigma'(z_1^1) \end{aligned}$$

$$\therefore \Delta w_{11}^1 = 0.00682 \sigma'(z_1^1)$$

고생하셨습니다~