Neural Network 기초 Assignment1

이름: 장혜림

Part 1. 함수 (20 points)

1. Sigmoid를 z에 대해 미분하세요. (2 points)

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

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

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

2. Mean Square Error를 $\boldsymbol{w_i}$에 대해 편미분하세요. (3 points)

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

$$o_i = \boldsymbol{w_i^T x} + b_i, (1 \le i \le K)$$

$$\frac{\partial J(W)}{\partial w_i} = \frac{1}{2}\cdot 2\ \Sigma(y_i - o_i)(y_i - o_i)'$$

$$= \Sigma(y_i - o_i)(0 - x - 0)$$

$$= -\Sigma(y_i - o_i)x$$

3. Logistic Regression의 Log Likelihood를 $\boldsymbol{w_j}$에 대해 편미분하세요. (3 points)

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

$$p_j = \sigma(z_i), z_i = \boldsymbol{w_i^T x} + b_i, (1 \le i \le K)$$

$$\frac{\partial J(w)}{\partial w_i} = -\Sigma\left(y_i \frac{\partial}{\partial w_i}\log p_i + (1-y_i)\frac{\partial}{\partial w_i}\log(1-p_i)\right)$$

$$= -\Sigma\left(\frac{y_i p_i(1-p_i)x}{p_i} + \frac{-(1-y_i)p_i(1-p_i)x}{1-p_i}\right)$$

$$= -\Sigma\left(y_i(1-p_i)x - (1-y_i)p_i x\right)$$

$$= -\Sigma(y_i - p_i)x \quad \text{또는} \quad \Sigma(p_i - y_i)x$$

Tobig's 13기 이지용

4. 다음 식이 올바른 이유를 증명하세요. (5 points)

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

k개의 classification 문제일때

ex) k=3 . [1. 0. 0], [0. 1. 0], [0. 0. 1] 처럼 one-hot encoding 가정하면

위 식 성립

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

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

P_i 미분할때 case 분리 (z_i 와 z_j)

i) $i=j$

$$\frac{\partial P_i}{\partial z_i} = \frac{e^{z_i} \cdot \Sigma e^{z_k} - e^{z_i} \cdot e^{z_i}}{(\Sigma e^{z_k})^2} = \frac{e^{z_i}(\Sigma e^{z_k} - e^{z_i})}{(\Sigma e^{z_k})^2} = \frac{\Sigma e^{z_i}}{\Sigma e^{z_k}} \cdot \frac{\Sigma e^{z_k} - e^{z_i}}{\Sigma e^{z_k}} = P_i(1-P_i)$$

ii) $i \neq j$

$$\frac{\partial P_i}{\partial z_j} = \frac{0 - e^{z_i} \cdot e^{z_j}}{(\Sigma e^{z_k})^2} = -\frac{e^{z_i}}{\Sigma e^{z_k}} \cdot \frac{e^{z_j}}{\Sigma e^{z_k}} = -P_i P_j$$

$$\frac{\partial CE}{\partial z_i} = -\Sigma y_k \frac{\frac{\partial}{\partial z_i} P_k}{P_k} = -y_i \frac{\overset{(i=j)}{P_i(1-P_i)}}{P_i} - \sum_{i \neq j} \left(y_j \frac{\overset{(i \neq j)}{-P_i P_j}}{P_j} \right)$$

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

$$= -y_i + \sum_j y_j P_i$$

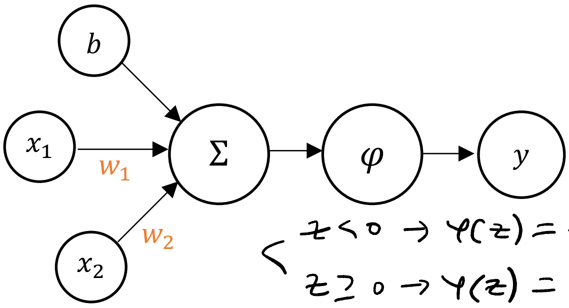
$$= -y_i + P_i \sum_j y_j$$

$$= P_i - y_i \quad (\because \sum_j y_j = 1)$$

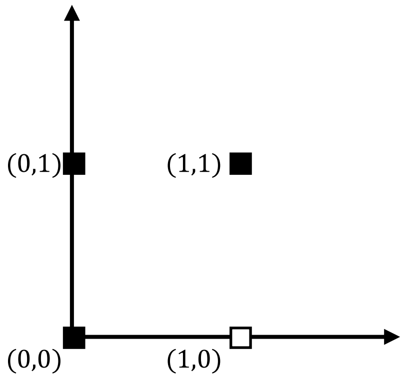
Tobig's 13기 이지용

Part 2. 퍼셉트론 (15 points)

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



$\begin{cases} z < 0 \to \varphi(z) = -1 \\ z \geq 0 \to \varphi(z) = +1 \end{cases}$



1. ■, □를 분류하는 임의의 b, $\boldsymbol{w}$를 선정하고 분류하는 과정을 보이세요. (5 points)

$b = +1$, $w_1 = -2$, $w_2 = +2$

$(0.0) \to \varphi(1) = 1 \to \hat{y} = 1$
$(0.1) \to \varphi(1+2) = 1 \to \hat{y} = 1$
$(1.0) \to \varphi(1-2) = -1 \to \hat{y} = 0$
$(1,1) \to \varphi(1-2+2) = 1 \to \hat{y} = 1$

2. Perceptron 학습 규칙에 따라 임의의 학습률 η을 정하고 b, $\boldsymbol{w}$를 한 번 업데이트해 주세요. (5 points)

$\eta = 0.01$

$b \leftarrow b + 0.01 \times 0$
$w_1 \leftarrow w_1 + 0.01 \times 0$
$w_2 \leftarrow w_2 + 0.01 \times 0$

$b = +1$. $w_1 = -2$, $w_2 = +2$

3. Adaline Gradient Descent에 따라 임의의 학습률 η을 정하고 b, $\boldsymbol{w}$를 한 번 업데이트해 주세요. (5 points)

x_0	x_1	x_2	y	w^Tx	$y-w^Tx$
1	0	0	1	1	0
1	0	1	1	3	-2
1	1	0	0	-1	1
1	1	1	1	1	0

$\eta = 0.01$

$b \leftarrow b + 0.01 \times (-3)$
$w_1 \leftarrow w_1 + 0.01 \times 1$
$w_2 \leftarrow w_2 + 0.01 \times (-2)$

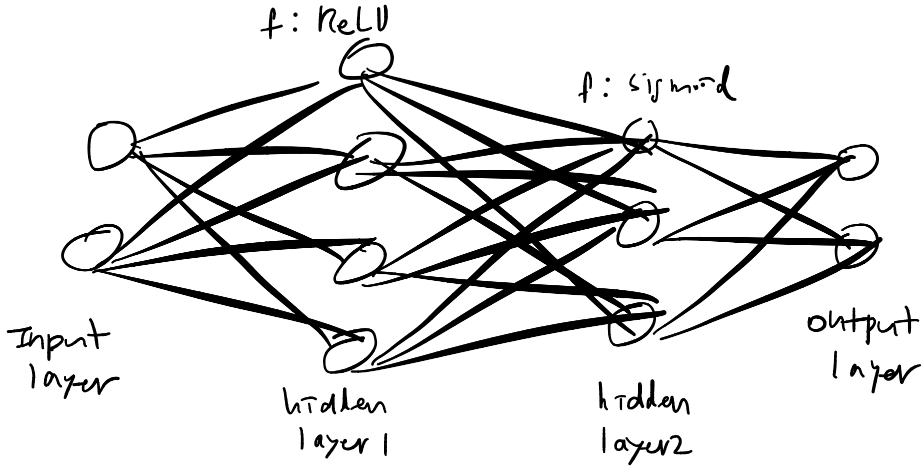
$b = 0.97$, $w_1 = -1.99$, $w_2 = 1.98$

Tobig's 13기 이지용

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)

input layer : 3
hidden layer 1 : 5
hidden layer 2 : 4
output layer : 2

total = 3×5×4×2 = 120

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

$$x = (x_1, x_2)^T,\quad x \in \mathbb{R}^{2\times 1}$$

$$W^1 = \begin{pmatrix} W_{11}^1 & W_{21}^1 & W_{31}^1 & W_{41}^1 \\ W_{12}^1 & W_{22}^1 & W_{32}^1 & W_{42}^1 \end{pmatrix}^T,\quad W^1 \in \mathbb{R}^{4\times 2}$$

$$b^1 = (b_1^1, b_2^1, b_3^1, b_4^1)^T,\quad b^1 \in \mathbb{R}^{4\times 1}$$

$$W^2 = \begin{pmatrix} W_{11}^2 & W_{21}^2 & W_{31}^2 \\ W_{12}^2 & W_{22}^2 & W_{32}^2 \\ W_{13}^2 & W_{23}^2 & W_{33}^2 \\ W_{14}^2 & W_{24}^2 & W_{34}^2 \end{pmatrix}^T,\quad W^2 \in \mathbb{R}^{3\times 4}$$

Tobig's 13기 이지용

$$b^2 = (b_1^2\ \ b_2^2\ \ b_3^2)^T,\quad b^2 \in \mathbb{R}^{3\times 1}$$

$$W^3 = \begin{pmatrix} W_{11}^3 & W_{21}^3 \\ W_{12}^3 & W_{22}^3 \\ W_{13}^3 & W_{23}^3 \end{pmatrix}^T,\quad W^3 \in \mathbb{R}^{2\times 3}$$

$$b^3 = (b_1^3\ \ b_2^3)^T,\quad b^3 \in \mathbb{R}^{2\times 1}$$

$$z^1 = W^1 x + b^1$$

$$a^1 = \mathrm{ReLU}(z^1)$$

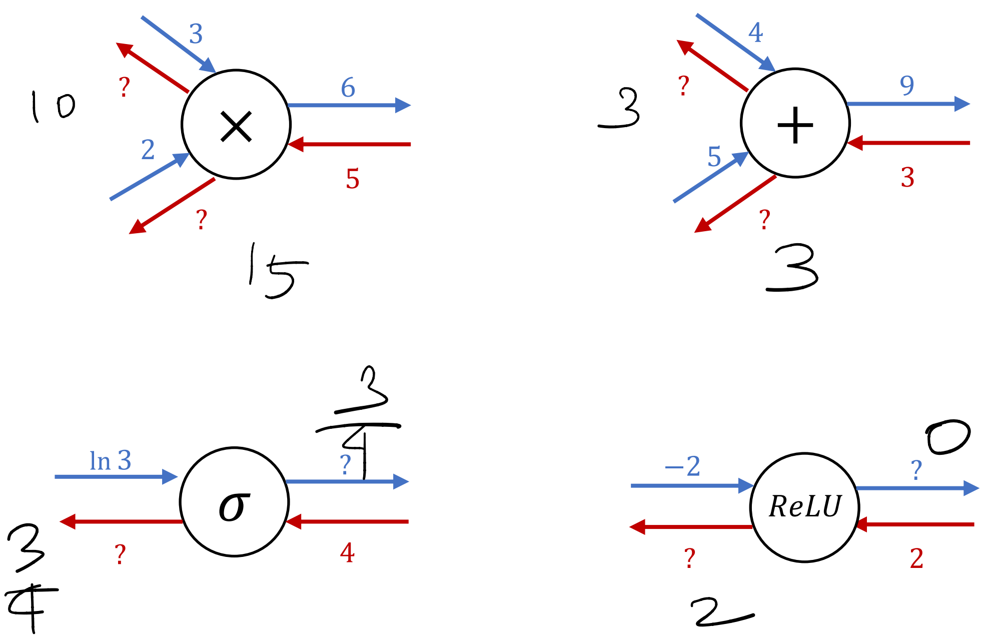
$$z^2 = W^2 a^1 + b^2$$

$$a^2 = \mathrm{sigmoid}(z^2)$$

$$z^3 = W^3 a^2 + b^3$$

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)

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

$$\delta^2 = \frac{\partial J}{\partial z^2} = f'(z^2)\,\Sigma W^3 \delta^3$$

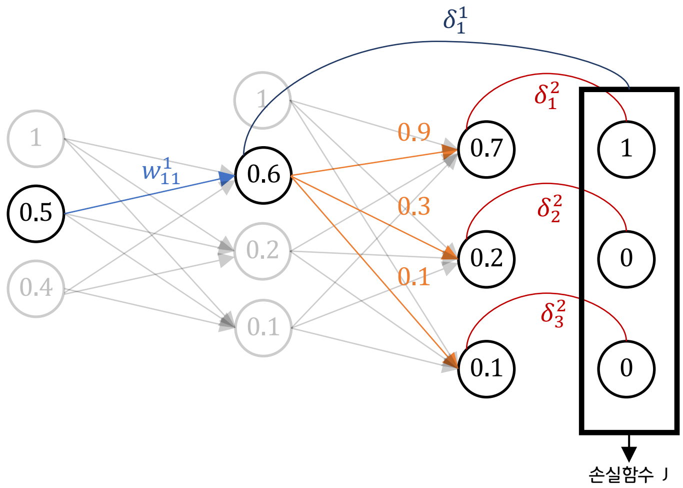
$$\delta^1 = \frac{\partial J}{\partial z^1} = f'(z^1)\,\Sigma W^2 \delta^2$$

$$b^3 \leftarrow b^3 - \eta\,\delta^3$$

$$b^2 \leftarrow b^2 - \eta\,\delta^2$$

$$b^1 \leftarrow b^1 - \eta\,\delta^1$$

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



$$\delta_1^2 = 1-0.7 = 0.3$$

$$\delta_2^2 = 0-0.2 = -0.2$$

$$\delta_3^2 = 0-0.1 = -0.1$$

$$\delta_1^1 = f'(0.1)(0.9\delta_1^2 + 0.3\delta_2^2 + 0.1\delta_3^2)$$

$$= 0.6(1-0.6) \times 0.2$$

$$= 0.048$$

$$\Delta w_{11}^1 = 0.05 \times \delta_1^1 x_1$$

$$= 0.05 \times 0.048 \times 0.5$$

$$= 0.0012$$

고생하셨습니다~

Tobig's 13기 이지용