Neural Network 기초 Assignment1

이름: 정세영

Part 1. 함수 (20 points)

1. Sigmoid를 z에 대해 미분하세요. (2 points)

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

$$\left\{\frac{1}{g(x)}\right\}' = -\frac{g'(x)}{\{g(x)\}^2}$$

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

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

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

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

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

$$J(W) = \frac{1}{2}\sum_{k=1}^{K}(y_k - o_k)^2 = \frac{1}{2}\sum_{k=1}^{K}(y_k - w_k^T x - b_k)^2$$

$$\frac{\partial J(W)}{\partial w_i} = (y_i - w_i^T x - b_i)\cdot(-x)$$

(∵ $k \neq i$ 일때 미분하면 사라지므로 i 항만 남고, $\frac{\partial(w_i^T x + b_i)}{\partial w_i} = \frac{\partial z_i}{\partial w_i} = x$)

$$= -x(y_i - w_i^T x - b_i)$$

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

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

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

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

$$= -\sum_{k=1}^{K}\left\{y_k \log\frac{1/(1+e^{-(w_k^T x+b_k)})}{e^{-(w_k^T x+b_k)}/(1+e^{-(w_k^T x+b_k)})} + \log\frac{e^{-(w_k^T x+b_k)}}{1+e^{-(w_k^T x+b_k)}}\right\}$$

$$= -\sum_{k=1}^{K}\{y_k(w_k^T x+b_k) - \log(1+e^{w_k^T x+b_k})\}$$

$$\frac{\partial J(W)}{\partial w_j} = -\sum_{k=1}^{K}\left\{y_k\cdot\frac{\partial(w_k^T x+b_k)}{\partial w_j} - \frac{1}{1+e^{-(w_k^T x+b_k)}}\cdot\frac{\partial(w_k^T x+b_k)}{\partial w}\right\}$$

(∵ $\frac{\partial(w_i^T x+b_i)}{\partial w_i} = \frac{\partial z_i}{\partial w_i} = x$)

$$= -(y_j x - p_j x) = -(y_j - p_j)x$$

4. 다음 식이 올바른 이유를 증명하세요. (5 points)

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

y는 k개의 class 중 하나만 1이고 나머지는 0이기 때문에

y가 1인 항만 남고 나머지는 사라져 Σ 연산이 필요없다.

즉 $y_i = 1$ 이라면 $y_{k \neq i} = 0$, $-\sum_{k=1}^{K} y_k \log(p_k) = -\log p_i \ (1 \le i \le k)$ 가 성립된다.

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

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

$$\frac{\partial CE}{\partial z_i} = -\sum_{k=1}^{K}\left(y_k \cdot \frac{1}{p_k} \cdot \frac{\partial p_k}{\partial z_i}\right) = -\sum_{k=1}^{K}\left(\frac{y_k}{p_k} \cdot \frac{\partial p_k}{\partial z_i}\right)$$

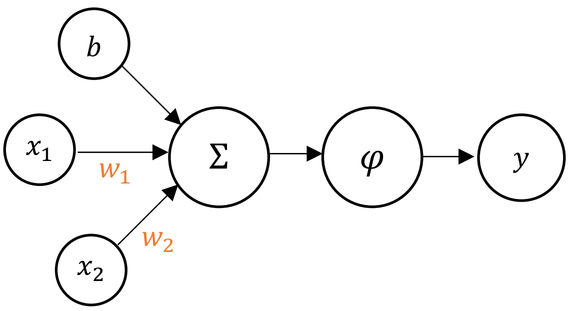
$$\frac{\partial p_k}{\partial z_i} = \begin{cases} \dfrac{e^{z_k}\sum_i e^{z_i} - e^{z_k}e^{z_k}}{(\sum_i e^{z_i})^2} = \dfrac{e^{z_k}}{\sum_i e^{z_i}}\left(1 - \dfrac{e^{z_k}}{\sum_i e^{z_i}}\right) = p_k(1-p_k) & (k = i) \\ -\dfrac{e^{z_k}}{\sum_i e^{z_i}} \dfrac{e^{z_i}}{\sum_i e^{z_i}} = -p_k p_i & (k \neq i) \end{cases}$$

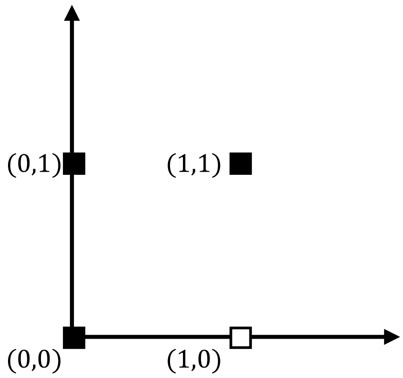
$$\frac{\partial CE}{\partial z_i} = -y_i(1-p_i) + \sum_{k \neq i} y_k \cdot p_i \quad (\because \textstyle\sum y_k = 1)$$

$$= p_i - y_i \quad (= -(y_i - p_i))$$

Part 2. 퍼셉트론 (15 points)

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





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

$b = 0.5$, $w_1 = -0.5$, $w_2 = 0.5$ 로 임의로 초기화한다. $\left(\varphi(z) = \begin{cases} 0, & z<0 \\ 1, & z \geq 0 \end{cases} \right)$

$\varphi(0.5) = 1$ (0,0)
$\varphi(1) = 1$ (0,1)
$\varphi(0) = 1$ (1,0)
$\varphi(0.5) = 1$ (1,1)

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

(1,0)에 대해서만 업데이트가 필요하다. 학습률 η은 0.05로 둔다.

$b \leftarrow b + 0.05(0-1)$
$w_1 \leftarrow w_1 + 0.05(0-1)$
$w_2 \leftarrow w_2 + 0.05(0-1)$

$b = 0.45$, $w_1 = -0.55$, $w_2 = 0.45$

$\Rightarrow$

$\varphi(0.45) = 1$ (0,0)
$\varphi(0.9) = 1$ (0,1)
$\varphi(-0.1) = 0$ (1,0)
$\varphi(0.35) = 1$ (1,1)

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	$w^T x$	$y - w^T x$
1	0	0	1	0.5	0.5
1	0	1	1	1	0
1	1	0	0	0	0
1	1	1	1	0.5	0.5

$b \leftarrow b + 0.05(1)$
$w_1 \leftarrow w_1 + 0.05(0.5)$
$w_2 \leftarrow w_2 + 0.05(0.5)$

$b = 0.55$, $w_1 = -0.475$, $w_2 = 0.525$

$\Downarrow$

$\varphi(0.55) = 1$ (0,0)
$\varphi(1.075) = 1$ (0,1)
$\varphi(0.075) = 1$ (1,0)
$\varphi(0.6) = 1$ (1,1)

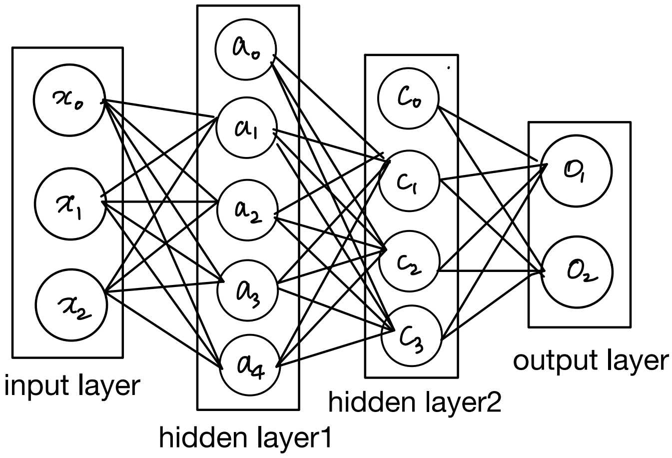
한번 업데이트로는 Perceptron 학습 결과가 더 좋다...

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 : 3개, hidden1 : 5개, hidden2 : 4개, output : 2개 (bias 포함)

layer1 : $3 \times 4 = 12$
layer2 : $5 \times 3 = 15$ $\quad 12 + 15 + 8 = 35$개
layer3 : $4 \times 2 = 8$

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

input : $x = (x_1, x_2, \ldots)^T$, x는 2×1 차원

weights : $W_i^L = (W_{i1}^L, W_{i2}^L, \ldots)^T$, W_i^1은 2×1 차원, W_i^2은 4×1 차원, W_i^3은 3×1 차원

$W^L = (W_1^L, W_2^L, \ldots)^T$, W^1은 4×2 차원, W^2은 3×4 차원, W^3은 2×3 차원

bias : $b^L = (b_1^L, b_2^L, \ldots)^T$, b^1은 4×1 차원, b^2은 3×1 차원, b^3은 2×1 차원

activation function : $\underset{(\text{ReLU})}{f(x)} = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$, $\underset{(\text{softmax})}{g(x)} = \dfrac{e^{z_x}}{\sum_i e^{z_i}}$

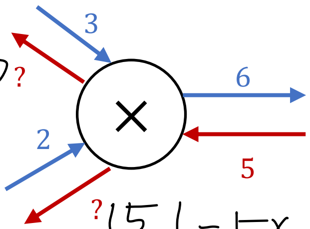
$$\Rightarrow \begin{aligned} z^1 &= W^1 x + b^1, \quad a = f(z^1) \\ z^2 &= W^2 a + b^2, \quad c = f(z^2) \\ z^3 &= W^3 c + b^3, \quad o = g(z^3) \end{aligned}$$

Tobig's 13기 이지용

Part 4. 역전파 (35 points)

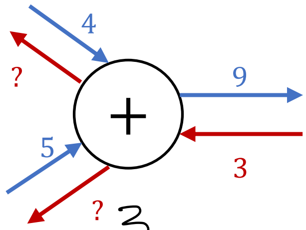
1. 다음 그림들의 물음표에 들어갈 숫자를 구하세요. (5 points)

$(5 \times \frac{6}{3} =)\ 10$



$15\ (= 5 \times \frac{6}{2})$

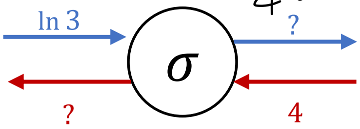
3



3

∵ +(Σ)는 그대로 분배해준다.

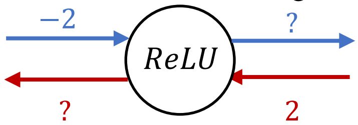
$\frac{3}{4}\left(= \frac{1}{1+e^{\ln 3}}\right)$



$\frac{3}{4}\left(= 4 \times \frac{dy}{dx}\Big|_{x=\ln 3}\right)$

$\frac{dy}{dx}\Big|_{x=\ln 3} = \frac{e^{-\ln 3}}{(1+e^{-\ln 3})^2} = \frac{1}{3} \div \left(\frac{4}{3}\right)^2 = \frac{3}{4^2}\ ,\ 4 \times \frac{3}{4^2} = \frac{3}{4}\ )$

0



$0\ (= 2 \times 0)$, $\frac{d\,\mathrm{ReLU}(x)}{dx}\Big|_{x=-2} = 0$

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

$\delta^3 = \frac{\partial J}{\partial z^3} = \begin{pmatrix} \delta_1^3 \\ \delta_2^3 \end{pmatrix}$, $\frac{\partial z^3}{\partial b^3} = I$ ($b^3 = h \times 1$ 벡터일때 $h \times h$ 크기의 단위행렬)

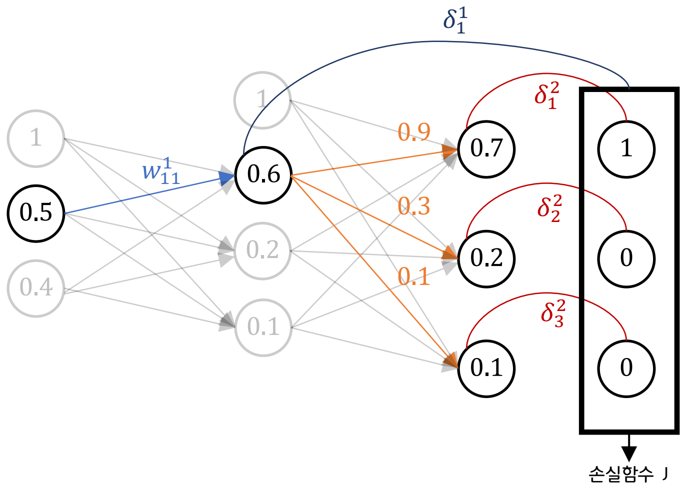
$\frac{\partial z^2}{\partial b^2}$ 와 $\frac{\partial z^1}{\partial b^1}$ 역시 단위행렬이므로

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

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

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

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)



$$\frac{\partial J}{\partial z^2} = \begin{pmatrix} p_1 - y_1 \\ p_2 - y_2 \\ p_3 - y_3 \end{pmatrix} = \begin{pmatrix} 0.7 - 1 \\ 0.2 - 0 \\ 0.1 - 0 \end{pmatrix} = \begin{pmatrix} -0.3 \\ 0.2 \\ 0.1 \end{pmatrix}, \quad \delta^2 = \frac{\partial J}{\partial z^2} = \begin{pmatrix} -0.3 \\ 0.2 \\ 0.1 \end{pmatrix}$$

$$\delta^1 = f'(z_1^1) \sum w_{k1}^2 \delta_k^2$$

$$f(z_1^1) = 0.6, \quad 0.6 = \frac{1}{1+e^{-z_1^1}}, \quad 1+e^{-z_1^1} = \frac{5}{3}, \quad e^{-z_1^1} = \frac{2}{3}, \quad z_1^1 = \ln\frac{3}{2}$$

$$f'(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad f'(z_1^1) = \frac{2/3}{(1+2/3)^2} = 0.24$$

$$\sum w_{k1}^2 \delta_k^2 = (0.9, 0.3, 0.1)\begin{pmatrix} -0.3 \\ 0.2 \\ 0.1 \end{pmatrix} = -0.2$$

$$\delta^1 = 0.24 \times (-0.2) = -0.048$$

$$w_{11}^1 \leftarrow w_{11}^1 + 0.05 \times 0.048 \times 0.5, \quad (\because x_1 = 0.5)$$

$$0.05 \times 0.048 \times 0.5 = 0.0012$$

고생하셨습니다~