3. (a) $\exp(x) > 0 \quad x \in \mathbb{R}$.

$$0 < \exp(f_y) < \exp(f_1) + \exp(f_2) + \cdots + \exp(f_k) \quad \text{where } f \in \mathbb{R}^k \text{ and } y \in \{1, \ldots, k\}$$

$$0 < \frac{\exp(f_y)}{\sum_{i=1}^{k} \exp(f_i)} < 1$$

$$-\infty < \log(x) < 0 \quad \text{where } 0 < x < 1$$ · [illegible]

$$\infty > -\log\left(\frac{\exp(f_y)}{\sum_{v=1}^{k} \exp(f_v)}\right) > 0$$

$$\therefore \; 0 < \ell^{CE}(f, y) < \infty$$

3 (b) $$\ell^{CE}(\lambda e_y, y) = -\log\left(\frac{e^{\lambda}}{e^0+e^0+\cdots+e^{\lambda}+\cdots+e^0}\right)$$

$$= -\log\left(\frac{e^{\lambda}}{e^{\lambda}+(k-1)e^0}\right)$$

$$= -\log\left(\frac{1}{1+(k-1)e^{-\lambda}}\right)$$

$$\therefore \lim_{\lambda\to\infty} \ell^{CE}(\lambda e_y, y) = \lim_{\lambda\to\infty} -\log\left(\frac{1}{1+(k-1)e^{-\lambda}}\right) = 0$$

4. I가 unique 하므로

$$f_I(x) > f_i(x)$$ i = I를 제외한 $1, \cdots, N$까지의 자연수.

$\therefore f_I(x) - f_i(x) > 0.$

f_I, f_i 모두 differentiable 하므로 continuous 임.

$N_{\varepsilon_i}(x)$ 안의 모든 원소가 $f_I(x) - f_i(x) > 0$을 만족하도록 하는 $\varepsilon_i > 0$를 각 i 마다 모두 잡을 수 있음.

$0 < \varepsilon < \min(\varepsilon_1, \cdots, \varepsilon_N)$ 인 ε 에 대해서

↑ I는 제외

$N_\varepsilon(x)$ 안의 모든 원소에 대해 $f_I(x) = f(x)$ 이고,

$f_I(x)$는 differentiable 이므로 $f(x)$도 differentiable.

$$\therefore f_I'(x) = f'(x)$$

5.(a) $z \in \mathbb{R}$, $z \geq 0$ 이라면

$$\sigma(z) = z$$

$$\sigma(\sigma(z)) = \sigma(z) = z$$

$$\therefore \sigma(z) = \sigma(\sigma(z)) \quad \text{where } z \geq 0$$

$z \in \mathbb{R}$, $z < 0$ 이라면

$$\sigma(z) = 0$$

$$\sigma(\sigma(z)) = \sigma(0) = 0$$

$$\therefore \sigma(z) = \sigma(\sigma(z)) \quad \text{where } z < 0$$

$$\therefore \sigma(z) = \sigma(\sigma(z)) \quad \forall z \in \mathbb{R}$$

5.(b) $\sigma'(z) = \dfrac{e^z}{1+e^z}$

$$|\sigma'(x) - \sigma'(y)| = \left| \frac{e^x}{1+e^x} - \frac{e^y}{1+e^y} \right|$$

M.V.T 사용하면, x와 y 사이의 다음을 만족하는 a 가 존재.

$$|\sigma''(a)| = \frac{\left| \frac{e^x}{1+e^x} - \frac{e^y}{1+e^y} \right|}{|x-y|}$$

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

$\alpha > 0$ 일 때 $g(\alpha) = \dfrac{\alpha}{(1+\alpha)^2}$ 의 최댓값은 $g'(\alpha) = \dfrac{1-\alpha}{(1+\alpha)^3}$ 이므로

$\alpha = 1$ 일 때 $\frac{1}{4}$ 이다.

$$\therefore \frac{\left| \frac{e^x}{1+e^x} - \frac{e^y}{1+e^y} \right|}{|x-y|} = |\sigma''(a)| \le \frac{1}{4}$$

$$\therefore |\sigma'(x) - \sigma'(y)| \le \frac{1}{4}|x-y|$$

$\therefore$ Lipschitz continuous derivatives

ReLU ([illegible])

$f(x) = \max(x, 0)$, $x = \varepsilon$, $y = -\varepsilon$, $\varepsilon > 0$ 이라고 가정

$$|f'(x) - f'(y)| = 1$$

$|f'(x) - f'(y)| = 1 \le M|\varepsilon - (-\varepsilon)| = 2M\varepsilon$. $\varepsilon < \frac{1}{2M}$ 일 때 모순.

$\therefore$ not Lipschitz.

5. (c) $\sigma(z) = \dfrac{1}{1+e^{-z}}$, $\ell(z) = \dfrac{1-e^{-2z}}{1+e^{-2z}}$

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

$y_1^\sigma = \sigma(A_1 x + b_1)$ 라하면

$C_1 x + d_1 = \frac{1}{2}(A_1 x + b_1)$ 이되도록 $C_1 = \frac{1}{2}A_1$, $d_1 = \frac{1}{2}b_1$으로 세팅.

$$y_1^\ell = \ell(C_1 x + d_1) = \ell\left(\tfrac{1}{2}(A_1 x + b_1)\right) = 2\sigma(A_1 x + b_1) - 1$$
$$= 2y_1^\sigma - 1$$

다음 step

$$y_2^\sigma = A_2 y_1^\sigma + b_2$$

$y_2^\ell = C_2 y_1^\ell + d_2 = 2C_2 y_1^\sigma - C_2 + d_2$ 이므로 $C_2 = \frac{A_2}{2}$, $d_2 = C_2 + b_2$로

세팅 하면 $y_2^\sigma = y_2^\ell$.

이후 모든 step도 $C_k = \frac{A_k}{2}$, $d_k = C_k + b_k$로 세팅하면 $y_k^\sigma = y_k^\ell$가 된다.

$(k \geq 2)$

6.

$$\nabla_a f_\theta(x) = \begin{bmatrix} u_1 \sigma'(a_1 x + b_1) x \\ \vdots \\ u_p \sigma'(a_p x + b_p) x \end{bmatrix}$$

$$\nabla_b f_\theta(x) = \begin{bmatrix} u_1 \sigma'(a_1 x + b_1) \\ \vdots \\ u_p \sigma'(a_p x + b_p) \end{bmatrix}$$

만약 모든 X_i 에 대해 $a_j^0 X_i + b_j^0 < 0$ 이라면 $\sigma'(a_j^0 X_i + b_j^0) = 0$

$$a_j^{k+1} = a_j^k - \alpha \frac{\partial \ell}{\partial f_\theta} \cdot u_j \sigma'(a_j^k X_i + b_j^k) X_i = a_j^k$$

$$b_j^{k+1} = b_j^k - \alpha \frac{\partial \ell}{\partial f_\theta} u_j \sigma'(a_j^k X_i + b_j^k) = b_j^k$$

7. σ에서 $\sigma'(a_j^0 x_i + b_j^0)$이 0이 아닌 α가 되면서

더이상 항상 0이되지 않는다. $\left(\frac{\partial J}{\partial f_\theta} = 0\right.$ or $a_j = 0$ 제거$\left.\right)$