ML Assn 3 20160463 성해빈

1. (a) $L = \sqrt{w_1^2 + w_2^2} + \mu(-w_1 - 2w_2 + 5)$

$w_1 + 2w_2 \geq 5 \rightarrow -w_1 - 2w_2 + 5 \leq 0$

stationary condition : $\frac{\partial L}{\partial w_1} = 0$, $\frac{\partial L}{\partial w_2} = 0$

$$\frac{\partial L}{\partial w_1} = \frac{1}{2} \frac{1}{\sqrt{w_1^2 + w_2^2}} \cdot 2w_1 - \mu = \frac{w_1}{\sqrt{w_1^2 + w_2^2}} - \mu = 0$$

$$\frac{\partial L}{\partial w_2} = \frac{1}{2} \frac{1}{\sqrt{w_1^2 + w_2^2}} \cdot 2w_2 - 2\mu = \frac{w_2}{\sqrt{w_1^2 + w_2^2}} - 2\mu = 0$$

$\Big) \mu = \frac{w_1}{\sqrt{w_1^2 + w_2^2}} = \frac{w_2}{2\sqrt{w_1^2 + w_2^2}}$

$\therefore 2w_1 = w_2$ to guarantee existence of μ

feasibility : $-w_1 - 2w_2 + 5 \leq 0 \rightarrow -5w_1 + 5 \leq 0 \rightarrow w_1 \geq 1$, $w_2 \geq 2$

$$\mu = \frac{\overset{\geq 1}{w_1}}{\underset{> 0}{\sqrt{w_1^2 + w_2^2}}} > 0$$

complementary : $\underset{> 0}{\mu} \underset{0}{\underbrace{(-w_1 - 2w_2 + 5)}} = 0 \quad \therefore -w_1 - 2w_2 + 5 = 0$

$\downarrow$

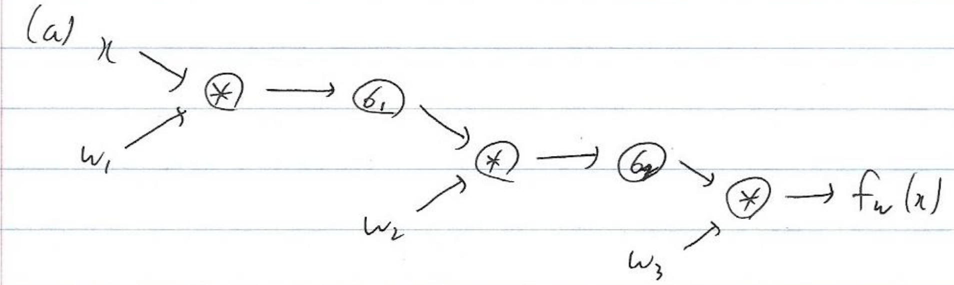
$-5w_1 + 5 = 0$

$w_1^* = 1$, $w_2^* = 2$ from KKT

(b) $\max \frac{1}{\sqrt{w_1^2 + w_2^2}}$ is equivalent to $\min \sqrt{w_1^2 + w_2^2}$, $\therefore (w_1^*, w_2^*) = (1, 2)$

2.

2. (a)



(b) $\frac{\partial \sigma_1}{\partial u} = \left(\frac{1}{1+e^{-u}}\right)' = -\frac{1}{(1+e^{-u})^2} \cdot e^{-u} \cdot -1 = \frac{e^{-u}}{(1+e^{-u})^2}$ (i)

$\frac{\partial \sigma_1}{\partial u} = \frac{e^{-u}}{(1+e^{-u})^2} = \frac{1}{1+e^{-u}} \times \left(1 - \frac{1}{1+e^{-u}}\right) = \sigma_1(u)(1-\sigma_1(u))$ (ii)

(c) forward pass: calculating $x \to f_w(x)$, going left to right in (a)

backward pass: calculating gradients $\frac{\partial f_w}{\partial w_3}, \frac{\partial f_w}{\partial w_2}, \frac{\partial f_w}{\partial w_1}$ in order going right to left

(d) $\frac{\partial f_w}{\partial w_3} = \boxed{\sigma_2(w_2\sigma_1(w_1 x))}$ we should retain $\sigma_2(w_2\sigma_1(w_1 x))$ from forward pass

(e) $\frac{\partial f_w}{\partial w_2} = w_3 \boxed{\sigma_2'(w_2\sigma_1(w_1 x))} \cdot \sigma_1(w_1 x) = w_3 \boxed{\sigma_2(w_2\sigma_1(w_1 x))}\left(1 - \boxed{\sigma_2(w_2\sigma_1(w_1 x))}\right) \cdot \boxed{\sigma_1(w_1 x)}$

we should retain $\sigma_2(w_2\sigma_1(w_1 x))$ and $\sigma_1(w_1 x)$

(f) $\frac{\partial f_w}{\partial w_1} = w_3 \boxed{\sigma_2'(w_2\sigma_1(w_1 x))} \cdot w_2 \sigma_1'(w_1 x) \cdot x$

$= w_3 \boxed{\sigma_2(w_2\sigma_1(w_1 x))}\left(1 - \boxed{\sigma_2(w_2\sigma_1(w_1 x))}\right) \cdot w_2 \boxed{\sigma_1(w_1 x)}\left(1 - \boxed{\sigma_1(w_1 x)}\right) \boxed{x}$

we should retain $\sigma_2(w_2\sigma_1(w_1 x))$ and $\sigma_1(w_1 x)$ and x

order $\frac{\partial f_w}{\partial w_3} \to \frac{\partial f_w}{\partial w_2} \to \frac{\partial f_w}{\partial w_1}$ is efficient since we can use info of $\frac{\partial f_w}{\partial w_3}$ to compute $\frac{\partial f_w}{\partial w_2}$

" $\frac{\partial f_w}{\partial w_2}$ to compute $\frac{\partial f_w}{\partial w_1}$

this order is right to left in (a), opposite from forward pass: left to right