<Logistic Loss Function>

- Squared error cost function은 convex function이기 때문에 내부에 새롭게 logistic loss function을 도입하여 해당 cost function을 logistic regression의 cost function으로 사용

~~$J_{\vec{w},b}($~~

cost

$$J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m} \underbrace{\frac{1}{2}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)^2}_{\Downarrow}$$

Loss $L(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)})$

logistic loss func

$$L(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)}) = \begin{cases} -\log(f_{\vec{w},b}(\vec{x}^{(i)})) \\ -\log(1-f_{\vec{w},b}(\vec{x}^{(i)})) \end{cases}$$

loss

$$L(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)})$$

$$= -y^{(i)}\log(f_{\vec{w},b}(\vec{x}^{(i)})) - (1-y^{(i)})\log(1-f_{\vec{w},b}(\vec{x}^{(i)}))$$

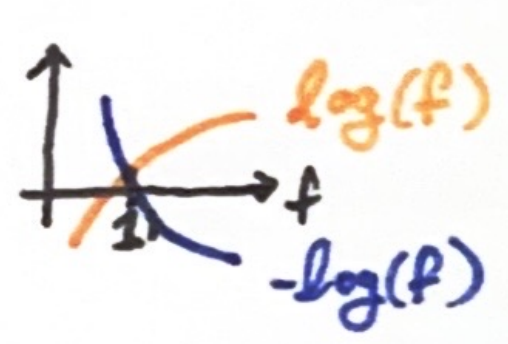
cost

$$J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\left[L(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)})\right]$$

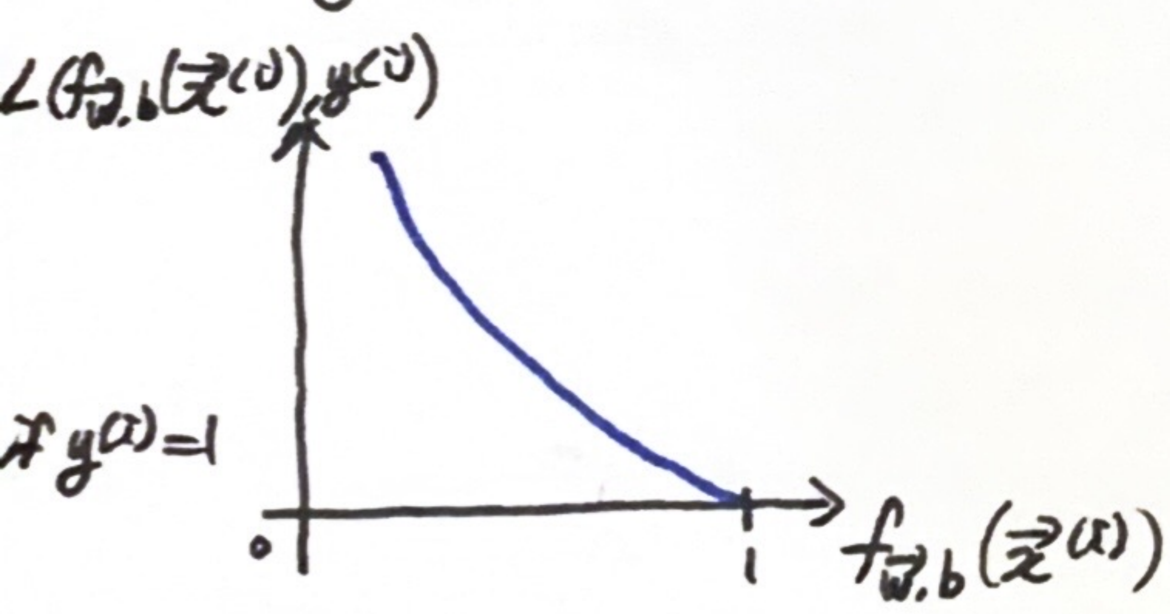
$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(f_{\vec{w},b}(\vec{x}^{(i)})) + (1-y^{(i)})\log(1-f_{\vec{w},b}(\vec{x}^{(i)}))\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(f_{\vec{w},b}(\vec{x}^{(i)})) + (1-y^{(i)})\log(1-f_{\vec{w},b}(\vec{x}^{(i)}))\right]$$

* log function



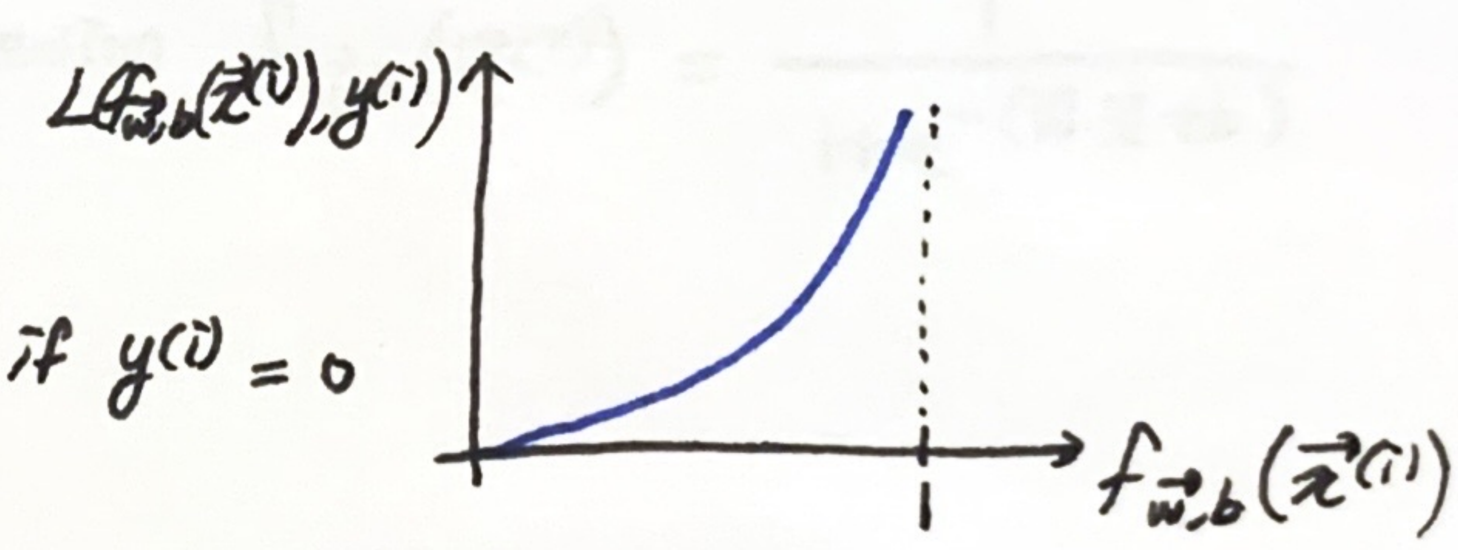
i) $-\log(f_{\vec{w},b}(\vec{x}^{(i)}))$



$y^{(i)}$ (=target) =1 일때

- $f_{\vec{w},b}(\vec{x}^{(i)})$ (=prediction) =1 $\rightarrow$ loss = 0
- $f_{\vec{w},b}(\vec{x}^{(i)})$ (=prediction) $\rightarrow 0 \Rightarrow$ loss $\rightarrow \infty$

ii) $-\log(1-f_{\vec{w},b}(\vec{x}^{(i)}))$



$y^{(i)}$ (=target) =0 일때

- $f_{\vec{w},b}(\vec{x}^{(i)})$ (=prediction) $\rightarrow 1 \Rightarrow$ loss $\rightarrow \infty$
- $f_{\vec{w},b}(\vec{x}^{(i)})$ (=prediction) = 0 $\Rightarrow$ loss $\Rightarrow 0$

∴ cost

$$J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m} \underbrace{L(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)})}_{\text{loss}}$$

$$\hookrightarrow = \begin{cases} -\log(f_{\vec{w},b}(\vec{x}^{(i)})) & \text{if } y^{(i)}=1 \\ -\log(1-f_{\vec{w},b}(\vec{x}^{(i)})) & \text{if } y^{(i)}=0 \end{cases}$$

$\Rightarrow$ overall cost function becomes convex

= can reach a global minimum by using gradient descent