LOGISTIC REGRESSION COST FUNCTION

$$Cost(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)), & \text{if } y=1 \\ -\log(1-h_\theta(x)), & \text{if } y=0 \end{cases}$$

$0 \leq h_\theta(x) < 1$

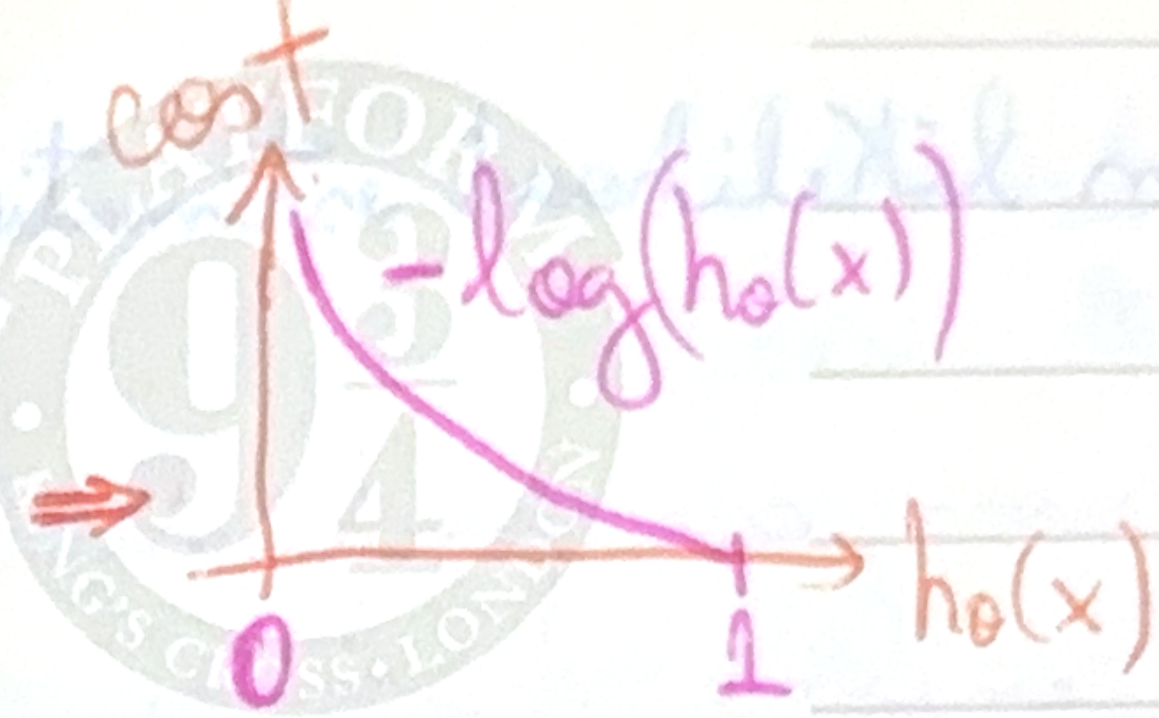
Se fosse usado a função das diferenças quadráticas, $J(\theta)$ seria não-convexa

Como $h_\theta(x) = \dfrac{1}{1+e^{-\theta^T x}}$

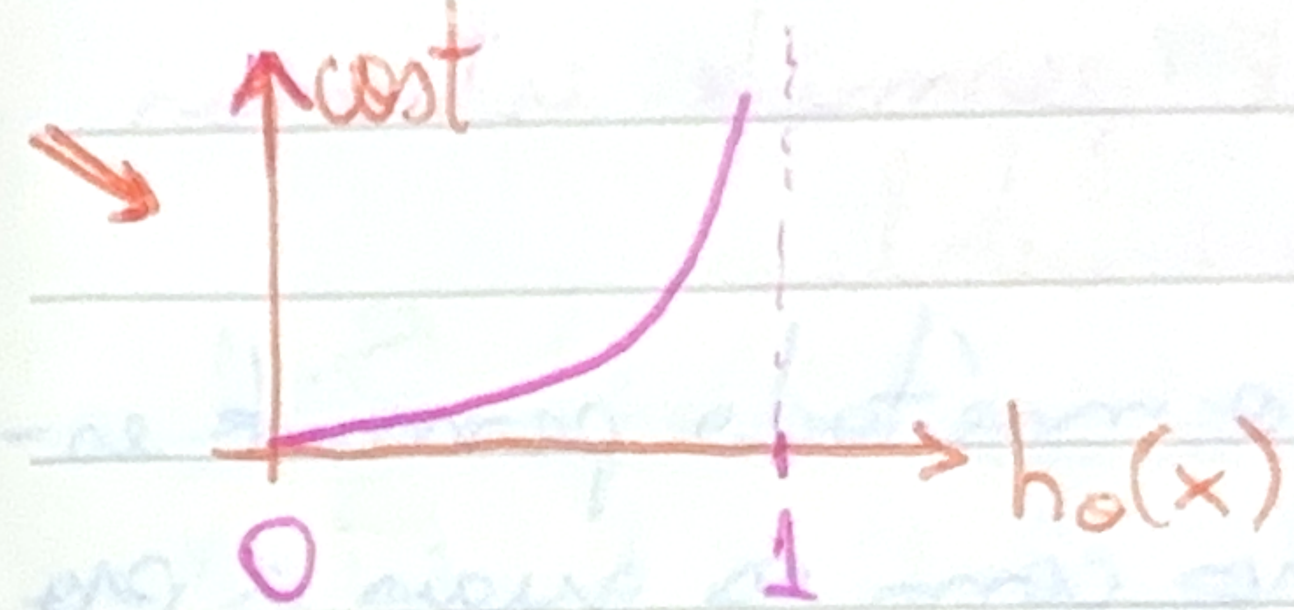
$$Cost(h_\theta(x), y) = \begin{cases} \log(1+e^{-\theta^T x}), & \text{if } y=1 \\ -\log\left(\dfrac{e^{-\theta^T x}}{1+e^{-\theta^T x}}\right), & \text{if } y=0 \end{cases}$$

$$Cost = \begin{cases} \log(1+e^{-\theta^T x}), & \text{if } y=1 \\ \log(1+e^{-\theta^T x}) - \theta^T x, & \text{if } y=0 \end{cases}$$

* **OBS:** ~~where~~ $\log a = \log_e a = \ln e$.



alto custo se $h_\theta(x) \to 0$ e $y = 1$



alto custo se $h_\theta(x) \to 1$ e $y = 0$

Minha versão simplificada:

$$Cost(h_\theta(x), y) = -\log\left(1 + (-1)^{1+y} \cdot h_\theta(x) - y\right)$$

não-diferenciável (?)

Versão simplificada do Prof. Andrew:

diferenciável

$$Cost(h_\theta(x), y) = -y \log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

$$\therefore J(\theta) = \frac{1}{m}\sum_{i=1}^{m} cost(h_\theta(x), y)$$

$$\Rightarrow \boxed{J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right]}$$

é uma função convexa

* OBS: com base no Princípio da Máxima Verossimilhança