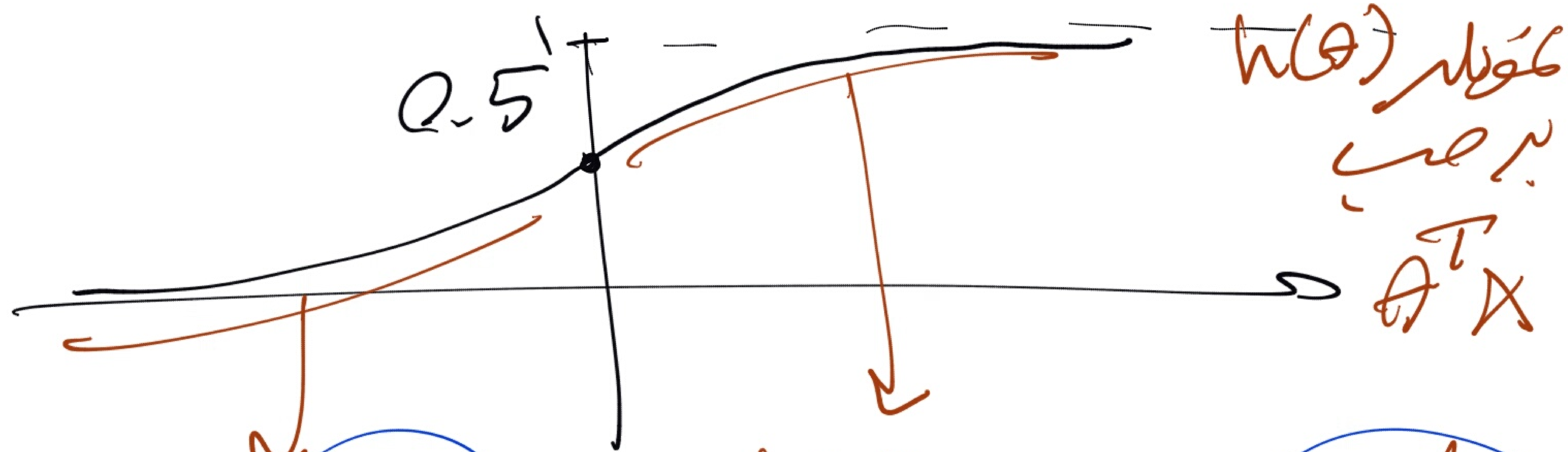
$$h(\theta) = \frac{1}{1+e^{-\theta^T x}}$$



$h(\theta) < 0.5 \Rightarrow y = 0$ $\qquad$ $h(\theta) \geq 0.5 \Rightarrow y = 1$

$P(y=1 \mid x;\theta) = h(\theta) \longrightarrow$ احتمال اینکه $y=1$ باشد

$P(y=0 \mid x;\theta) = 1 - h(\theta) \longrightarrow$ احتمال اینکه $y=0$ باشد

$$P(y \mid x;\theta) = (h_\theta(x))^y (1-h_\theta(x))^{1-y}$$

$$L(\theta) = \left(h_\theta(x^{(i)})\right)^{y^{(i)}} \left(1 - h_\theta(x^{(i)})\right)^{1-y^{(i)}}$$

$$\ell(\theta) = \log(L(\theta))$$

$$\ell(\theta) = \log\left(\left(h_\theta(x^{(i)})\right)^{y^{(i)}}\right) + \log\left(\left(1 - h_\theta(x^{(i)})\right)^{1-y^{(i)}}\right)$$

$$\ell(\theta) = y^{(i)} \log\left(h_\theta(x^{(i)})\right) + (1 - y^{(i)}) \log\left(1 - h_\theta(x^{(i)})\right)$$

حال برای آن که بتوانیم به جای maximize کردن آن، آن را minimize کنیم

کافی است آن را در یک منفی $\frac{1}{2}$ ضرب کنیم. وقتی آن را minimize

می‌کنیم این کمک را می‌کند که بعداً می‌توانیم $\frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$ را برای regularized کردن به آن اضافه کنیم

$$\frac{\partial}{\partial\theta_j}\ell(\theta) = \left(y\frac{1}{g(\theta^T x)} - (1-y)\frac{1}{1-g(\theta^T x)} \right)\frac{\partial}{\partial\theta_j} g(\theta^T x)$$

$$= \underline{\qquad\qquad\qquad} \times g(\theta^T x)(1-g(\theta^T x))\frac{\partial}{\partial\theta_j}\theta^T x$$

$$= \left(y(1-g(\theta^T x)) - (1-y)g(\theta^T x) \right) x_j$$

$$= \underbrace{(y - h_\theta(x))\,x_j} \Longrightarrow \frac{\partial J(\theta)}{\partial\theta_j} = \sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)}$$

برای اینکه gradient descent تبدیل بشود به gradient ascent ل را مشتق اجرا میکنیم