Logistic Regression Cost Function

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log(1-h_\theta(x^{(i)}))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

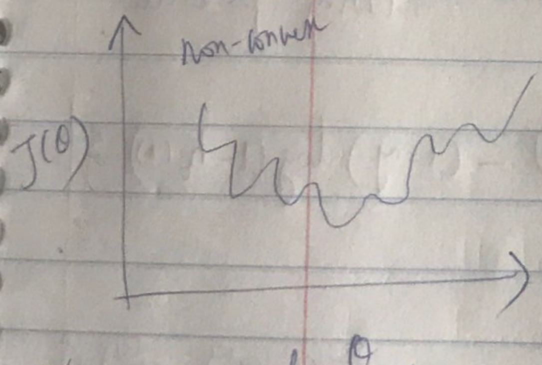
$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

Linear Regression $\quad J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \underbrace{\frac{1}{2}\left(h_\theta(x^i) - y^{(i)}\right)^2}_{cost(h_\theta(x^i),\, y^i)}$

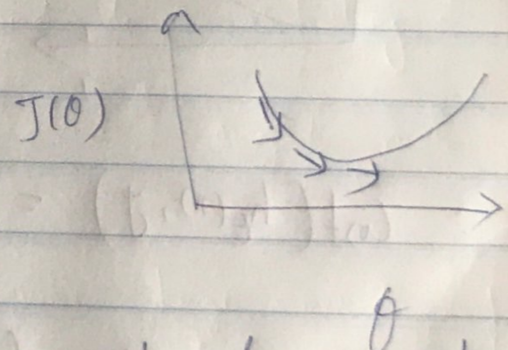
↓ Is convex function so we can minimize it of parameter's θ

for logistic

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



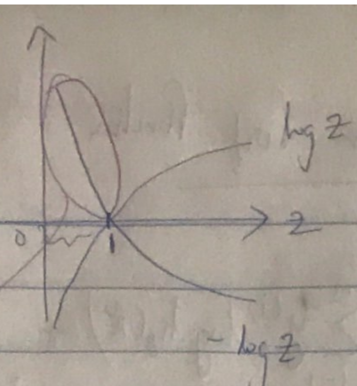
Non-convex due to very non-linear Sigmoid function.

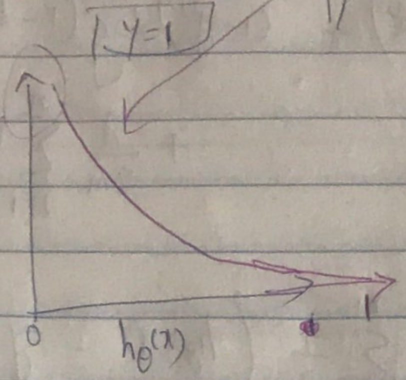


Here due to convexity gradient descent can converge to global optima.

Convert Non-convex to Convex

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

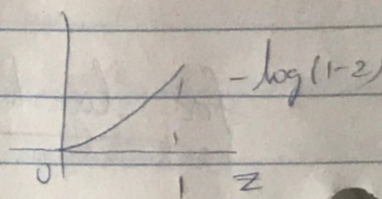


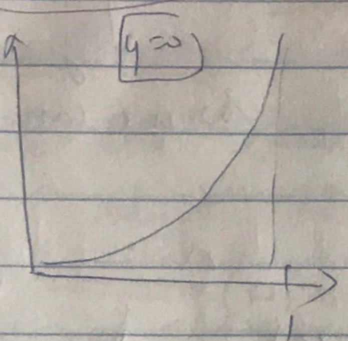


If $y = 1$ & $h_\theta(x) = 1 \Rightarrow$ cost = 0

If $y = 1$ & $h_\theta(x) = 0 \Rightarrow$ cost = α

wrong classification - heavily penalized.





If $y = 0$

$h_\theta(x) = 1$ & $y = 0 \Rightarrow$ cost = α

wrong prediction

$$\therefore \text{Cost}(h_\theta(x), y) = -y \log(h_\theta(x)) - (1-y)\log(1 - h_\theta(x))$$

Now we can apply Gradient descent.

Want $\min_\theta J(\theta)$

Repeat {

$$\theta_j := \theta_j - \alpha \underbrace{\frac{\partial}{\partial \theta_j} J(\theta)}_{\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}}$$

(Simultaneously update all θ)

}