c. Upgrading cost function to log likelihood function

- To get convex function in order to reach global minima, first we convert sigmoid function to probalistic function.
- Our cost function with sigmoid applied hypothesis function in probability form is,

$$P(y_i = 1 \mid x_i : \theta) = h_\theta(x_i)$$
$$P(y_i = 0 \mid x_i : \theta) = 1 - h_\theta(x_i)$$

→ combining this we get shorter form as,

$$\boxed{P(y_i \mid x_i : \theta) = [h_\theta(x_i)]^{y_i} [1 - h_\theta(x_i)]^{1-y_i}}$$

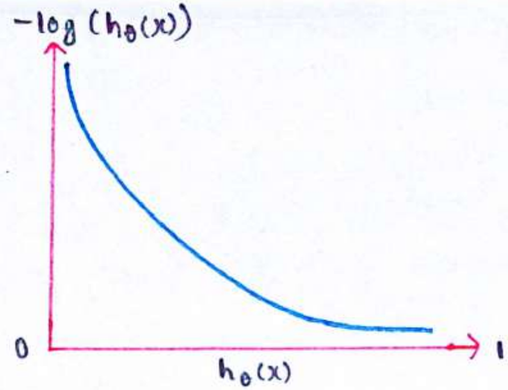
- Now we will further introduce log likelihood on $P(y_i \mid x_i : \theta)$ to get,

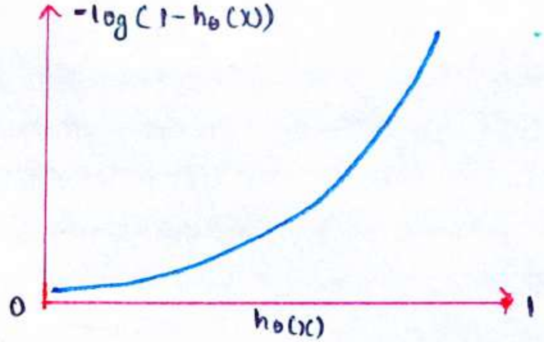
$$\Rightarrow \boxed{L(\theta) = -y \log(h_\theta(x_i)) - (1-y_i) \log(1-h_\theta(x_i))}$$

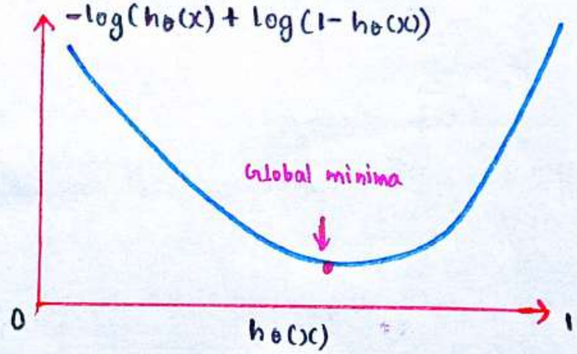
if we put $y=1$ or $y=0$ we can sumarize $L(\theta)$ as,

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

- If we plot both of these cases we will get something like this:







- After combining both we can observe we were able to get a convex function.
- Now substituting this log likelihood hypothesis function ($h_\theta(x)$) in cost function ($J(\theta_0, \theta_n)$) we get our final cost function.

$$\boxed{J(\theta_0, \theta_n) = \frac{1}{n} \left[y^i \log(h_\theta(x^i)) + (1-y^i) \log(1-h_\theta(x^i)) \right]^2}$$

- Now we will apply convergence algorithm on cost function to get the best fit line.

$$\boxed{\theta_n = \theta_n - \alpha \frac{\partial}{\partial \theta_n} (J(\theta_0, \theta_n))}$$

$$\boxed{\theta_0 = \theta_0 - \alpha \frac{\partial}{\partial \theta_0} (J(\theta_0, \theta_n))}$$