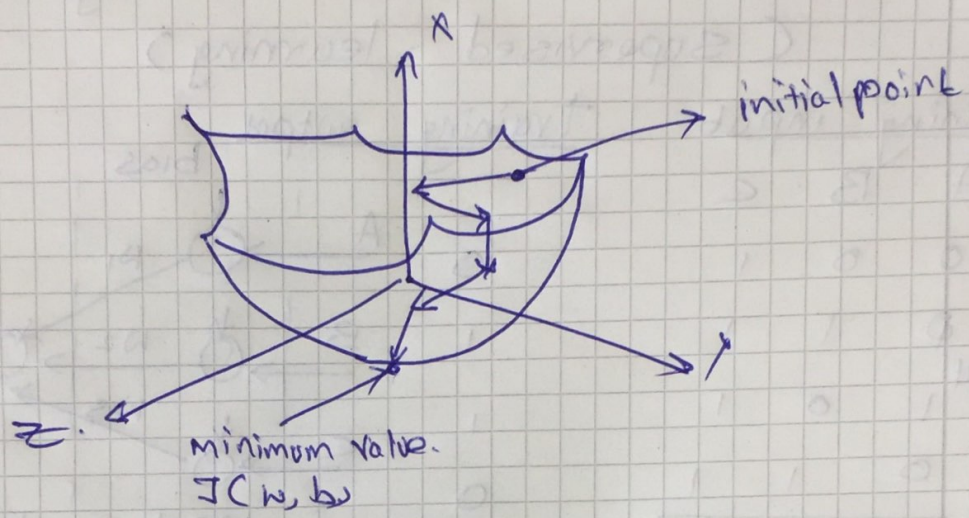
error function.

Gradient decent - We need to minimize the error to its optimal point, gradient decent starts the algo at any random point and it calculates steps towards the down direction i.e min value.



Sigmoid function $= \frac{1}{1+e^{-x}}$

derivative of Sigmoid $= \frac{d}{dx}(1+e^{-x})^{-1}$

$= (-1)(1+e^{-x})^{-2}$

$= (-1)(1+e^{-x})^{-2}(-1)e^{-x}$

$= (1+e^{-x})^{-2} \cdot e^{-x}$

$= \frac{e^{-x}+1-1}{(1+e^{-x})^{+2}}$

$= \frac{e^{-x}+1}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2}$

$= \frac{1}{1+e^{x}}\left(1 - \frac{1}{1+e^{-x}}\right)$

$\boxed{= y(1-y)}$