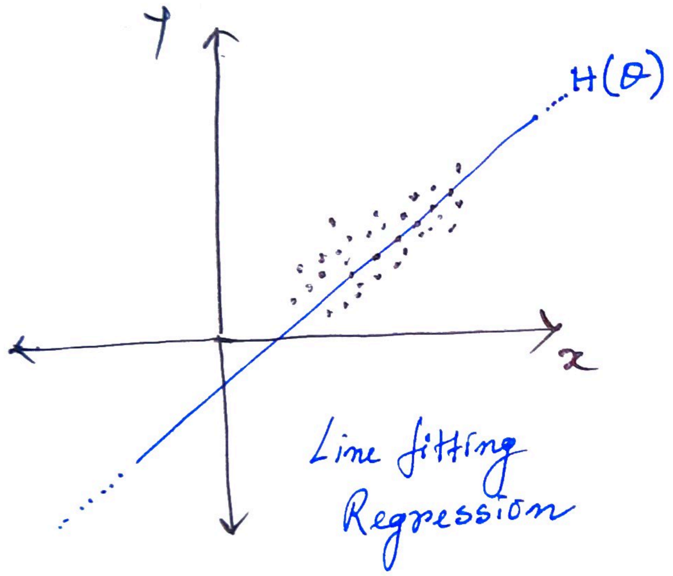
Regression Vs Classification

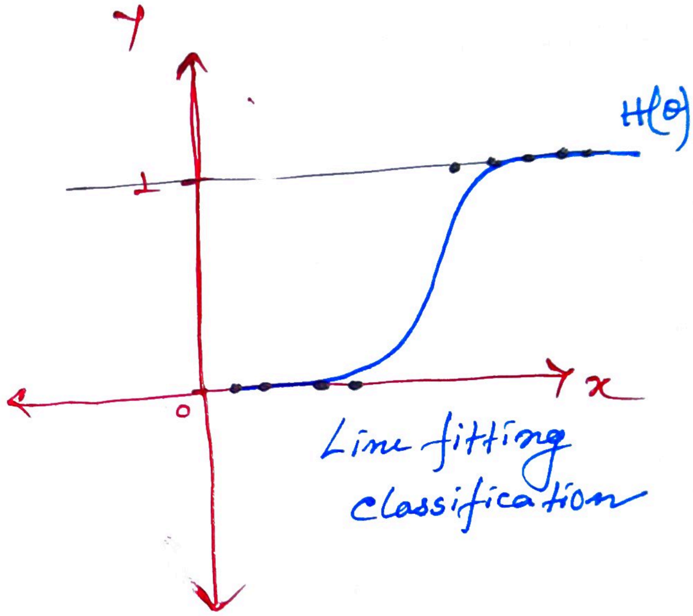
(X) Real Numbers → Regression Problem → (Y) Real Numbers

Example - Temperature, Price etc.

(X) Real Numbers → Classification Problem → (Y) Discrete Set of Numbers

Example - True / False, Class1 / Class2

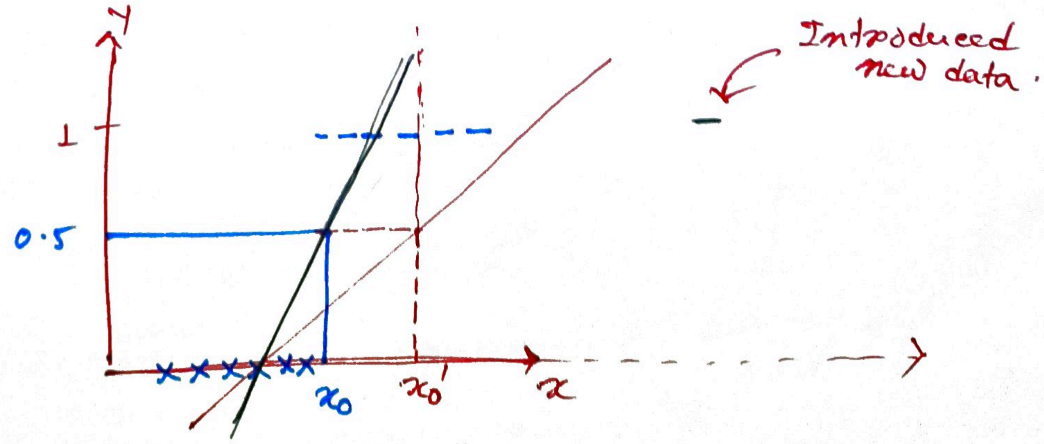




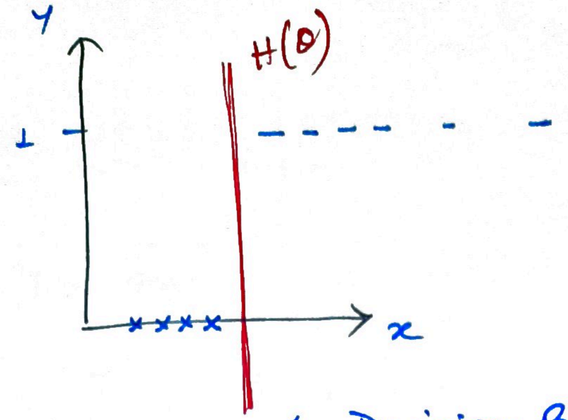
Unique Requirement ~~for~~ classification Problem:-

Req - 1 Output of the hypothesis should be restricted to the discrete set of values (like 0, 1)

Req - 2



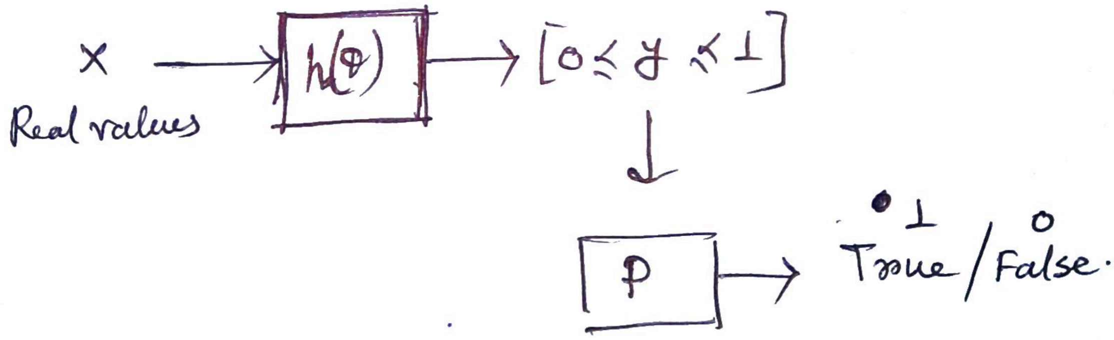
$x \geq x_0 \Rightarrow 1$
$x < x_0 \Rightarrow 0$



← Decision Boundary

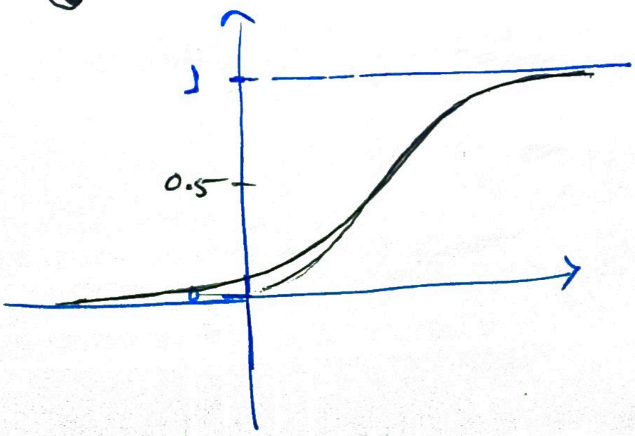
"So, unlike Regression line, that passes through the points minimizing the distance from the data points, for classification problem we will search for a Decision line/Plane to separate different classes of data."

So we need to develop a function like.



Sigmoid Function / Logistic function

$$g(z) = \frac{1}{1+e^{-z}}$$



For linear Regression

$y = mx$ or

$y = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots \theta_{n-1} x_{n-1} + \theta_n$.

$= \theta^T X$ [vectorized form]

Now passing output of L·R to Logistic fn. –

$$y = g(\theta^T X) = \frac{1}{1 + e^{-\theta^T X}}$$

$\therefore 0 \leq g(\theta^T X) \leq 1$

Now let us interpret $g(\theta^T X)$ is the estimated ~~probab~~ probability of $y=1$ for given x. parameterize by θ.

i.e. $P(y=1 \mid x, \theta) = h_\theta(x) = g(\theta^T x)$

we know, y is either 0 or 1

$P(y=1 \mid x, \theta) + P(y=0 \mid x, \theta) = 1$

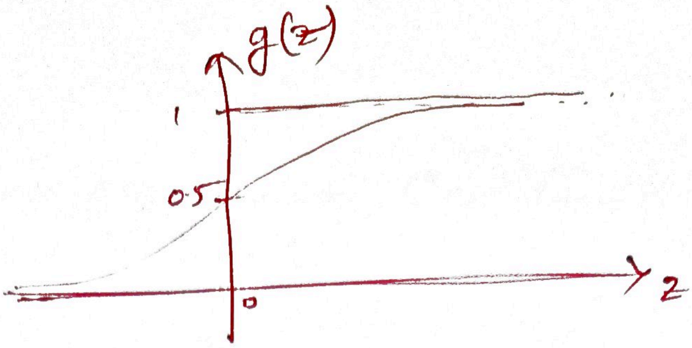
$P(y=0 \mid x, \theta) = 1 - P(y=1 \mid x, \theta)$

As we need to give the o/p as 0 or 1 (possible o/p value)

Let say if $P(y=1 \mid x, \theta) \geq 0.5 \Rightarrow Y = 1$

if $P(y=1 \mid x, \theta) < 0.5 \Rightarrow Y = 0$

$\therefore$ $h_\theta(x) \geq 0.5$ output $y=1$



$h_\theta(x) \geq 0.5$ we predict $y = 1$

$g(\theta^T x) \geq 0.5$

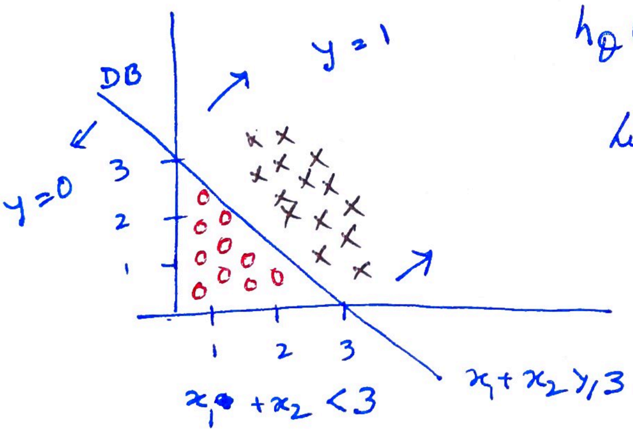
$g(z) \geq 0.5$

when $z \geq 0$

$\therefore \theta^T x \geq 0$.

Similarly

$\theta^T x < 0$ we predict $y = 0$.



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

Let us assume we learn the Parameters like -

$\theta_0 = -3$, $\theta_1 = 1$, $\theta_2 = 1$

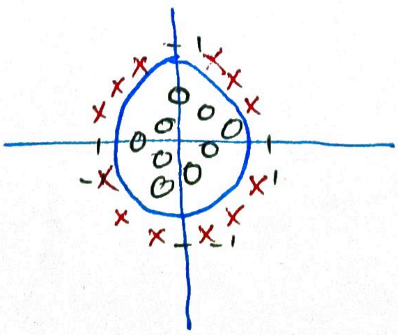
then, $\theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$

To predict $y = 1$ $h_\theta(x) \geq 0.5 \Rightarrow \theta^T x \geq 0$

$$-3 + x_1 + x_2 \geq 0$$

$$x_1 + x_2 \geq 3$$

<u>Non linear Decision Boundary</u>.



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

Predict $(y=1)$ if.

$$-1 + x_1^2 + x_2^2 \geq 0$$

$$x_1^2 + x_2^2 \geq 1$$

Equation of circle.

Cost Function

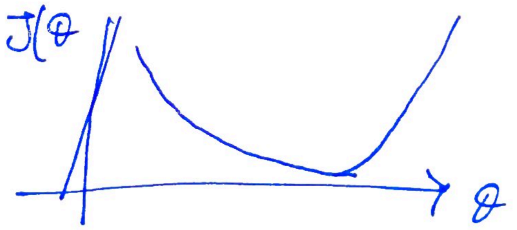
Linear Regression Cost function -

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} \left(h_\theta(x^i) - y^i \right)^2$$

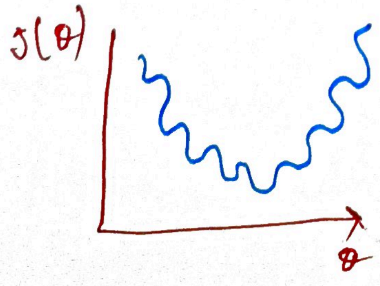
and we need to minimize $J(\theta)$ to find out the parameters ie. $\theta = \theta_0, \theta_1, \dots \theta_n$.

→ Remember we used numerical method called Gradient Descent to find out the minima value of $J(\theta) / \theta_0, \theta_1, .. \theta_n$.

→ ~~To~~ Remember Gradient Descent algorithm would converge if $J(\theta)$ vs θ is a convex curve. like.

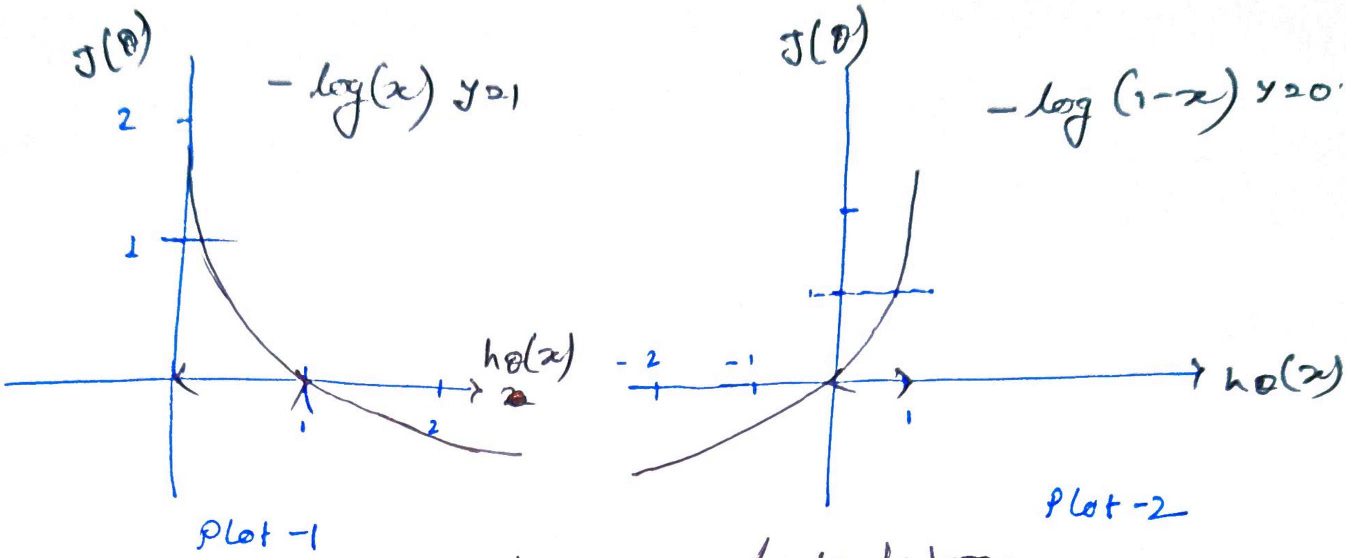


" If we try to use linear regression Cost function for logistic regression it can be shown that the resultant cost function won't be convex any more. So we need to find out a new cost function for logistic Regression.



Logistic Regression Cost Function

$$Cost(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y=1 \rightarrow (1) \\ -\log(1-h_\theta(x)) & \text{if } y=0 \rightarrow (2) \end{cases}$$



desmos.com/calculator.

We are only interested for the segment $0 \leq x \leq 1$

From Plot-1

for a data x_i if we predict $y_i = 0$ ie ~~h(x)~~ $h_\theta(x) = 0$ we can see the cost approaches to '∞' and if $y_i = 1$ (ie. correct prediction) cost $\rightarrow 0$

From Plot-2

for a data x_i if we predict $y_i = 1$ ie $h_\theta(x) = 1$ we can see the cost approaches to ∞ and for $y_i = 0$ (ie. correct prediction) cost $\rightarrow 0$.

Combining Eq-1 & E-2.

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

This cost f^n is convex in nature

So the Cost f^n for logistic regression –

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

$$= -\frac{1}{m} \sum_{i=1}^{m} y^i \log(h_\theta(x^i)) + (1-y^i) \log(1-h_\theta(x^i))$$

So we need to →

minimize $(J(\theta))$ to get the θs

To minimize the $J(\theta)$ we would use G.D.

$$\theta_{j+1} = \theta_j - \alpha \frac{\delta}{\delta\theta_j} J(\theta)$$

$$\frac{\delta}{\delta\theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i) x_j^i$$

$$\therefore \theta_{j+1} = \theta_j - \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i) x_j^i$$