A travel agency wants an automated system to predict travel costs. The agency has the following data available with it.

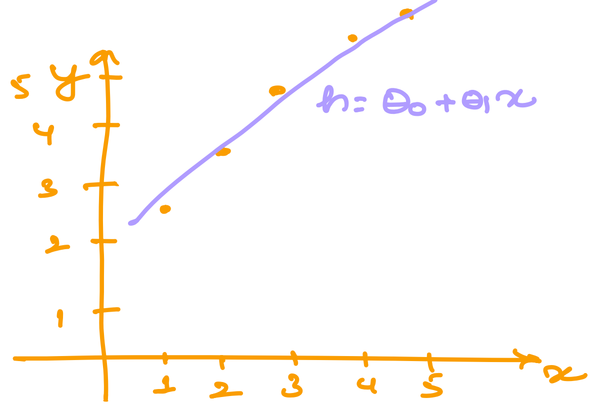


Table II

S. No.	Distance (in Km)	Travelling Cost (in Rupees)
1	1	2.75
2	2	3.5
3	3	4.25
4	4	5
5	5	5.75

[redacted]

Formulate the above problem as a linear model $h(x) = w_0 + w_1 x$ to predict the travelling cost for a given distance. The parameter w_0 is 2 (optimal). Apply gradient descent algorithm to find optimal parameter w_1. The learning rate for the first epoch is 0.073, and for the second epoch and later, the learning rate is 0.091. Let the initial value of w_1 is 0.5.

$$\theta = \theta - \eta \nabla_\theta J(\theta) \qquad \frac{\partial J(\theta)}{\partial(\theta)}$$

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

$\theta_0 = 2$ (optimal)

$\theta_1 = ?$

initial $\theta_1 = 0.5$ $\quad \eta = 0.073$

$$\theta_1 = \theta_1 - \frac{\eta}{m} 2 \sum_{i=1}^{m} \left[\hat{y}^{(i)} - y^{(i)}\right] x^{(i)}$$

$$\hat{y}^{(i)} = h_\theta(x^{(i)}) =$$

$$= \theta_0 + \theta_1 x_1 \qquad = \theta_0 x_0 + \theta_1 x_1$$

$$= \theta^T x$$

$$[\theta_0 \ \ \theta_1] \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \quad 1$$

$$[\theta_0 \ \ \theta_1] \begin{bmatrix} 1 \\ x \end{bmatrix}$$

$\theta_0 = 2$

$\theta_1 = 0.5$

$x^{(i)}$	$h_\theta(x^{(i)})$	y	$h_\theta(x^{(i)})-y$	$(h_\theta(x^{(i)})-y)x^{(i)}$
1	$2+0.5*1=2.5$	2.75	-0.25	-0.25
2	$2+0.5*2=3$	3.5	-0.5	-1
3	$2+0.5*3=3.5$	4.25	-0.75	-2.25
4	$2+0.5*4=4$	5	-1	-4
5	$2+0.5*5=4.5$	5.75	-1.25	-6.25
				-13.75

$\sum(h_\theta(x^{(i)})-y)x^{(i)}$

$$\theta_1 = 0.5 - \frac{0.073*2}{5}(-13.75)$$

$$= 0.5 + \frac{0.073*2*13.75}{5} = 0.9$$

2nd epoch

$$\theta_1 = 0.9 - \frac{0.091*2}{5}\left(\Sigma \_\_\_\_\_\_\right)$$

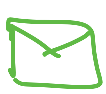
with θ_1 as 0.9

Logistic Regression

→ classification Algo.

eg:

→ wt, ht of dogs & cats:

→ spam or not spam  → NLP

→ $\boxed{\text{Anim}}$ → Cat, → Dog → CV

Training Data: $\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

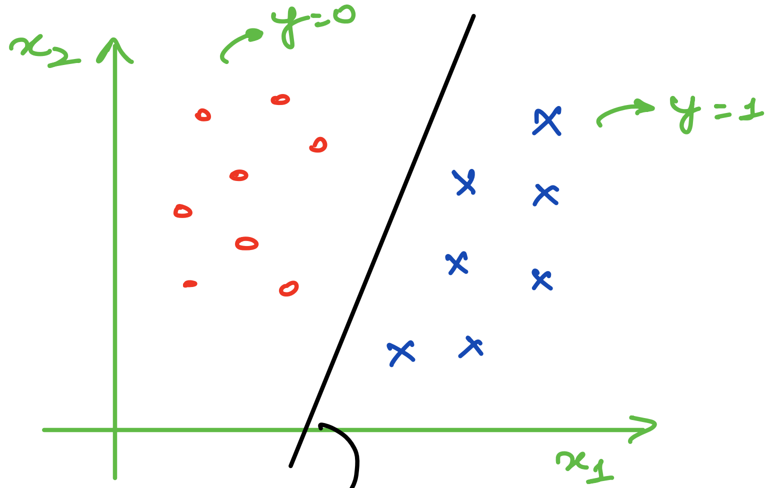
$x \in \mathbb{R}^n$

x has n features.

$y \in \{0,1\}$

Classification task

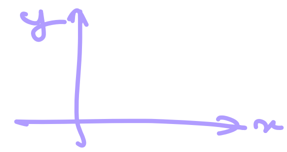
y should be a discrete value



$\underbrace{\theta_0+\theta_1 x_1+\theta_2 x_2}_{\theta^T x}=0$

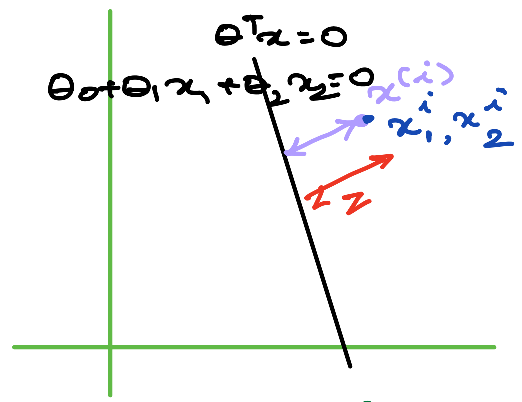
$y=mx+c$

$ax+by+c=0$



Linear Reg: $h_\theta(x)=\sum_{j=0}^{n}\theta_j x_j = \theta^T x$

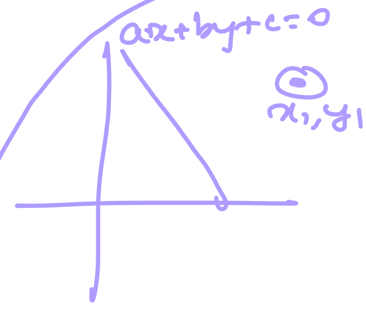
$=\theta_0+\theta_1 x_1+\theta_2 x_2$



$\theta^T x^{(i)} \to Z$

unnormalized distance b/w $x^{(i)}$ & line $\theta^T x=0$

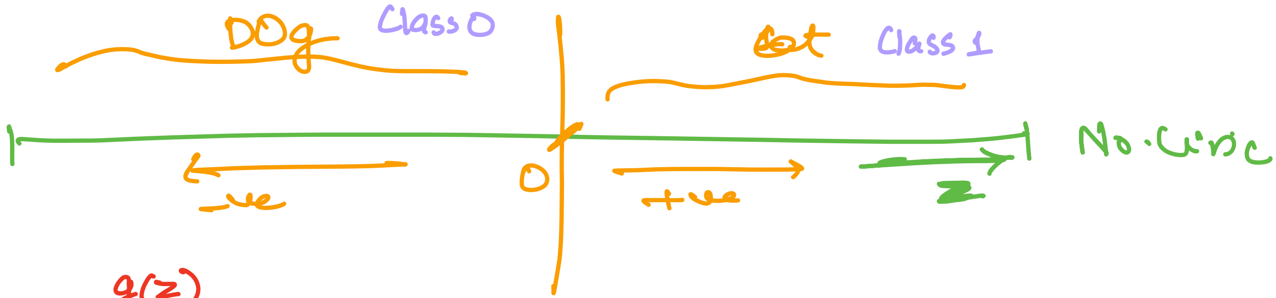
→ Real no.

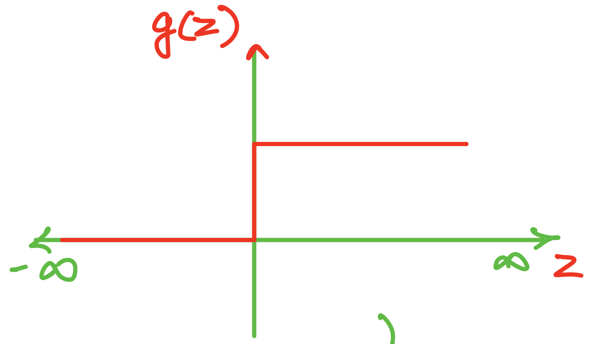


$\frac{ax_1+by_1+c}{\sqrt{a^2+b^2}}$ → normalized distance

$\theta_0+\theta_1 x_1^i+\theta_2 x_2^i$

$\underbrace{[\theta_0\ \ \theta_1\ \ \theta_2]}_{\theta^T} \cdot \underbrace{\begin{bmatrix}1\\ x_1^i\\ x_2^i\end{bmatrix}}_{x^i}$

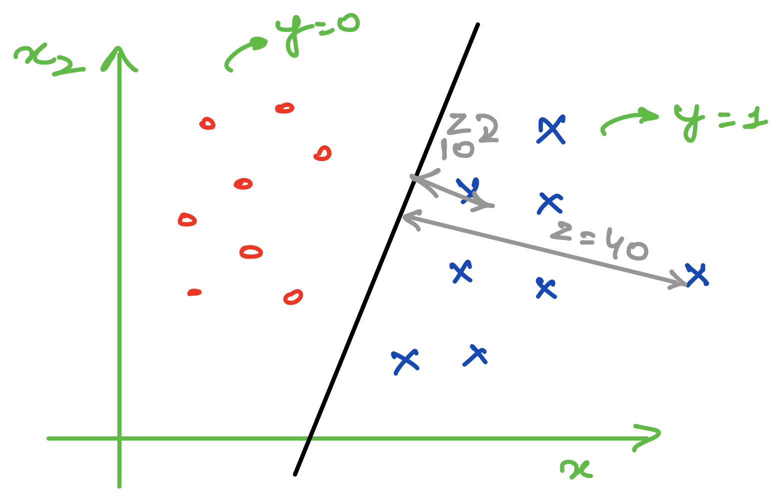




$g(z) = 1 \quad \text{if } z \geq 0$

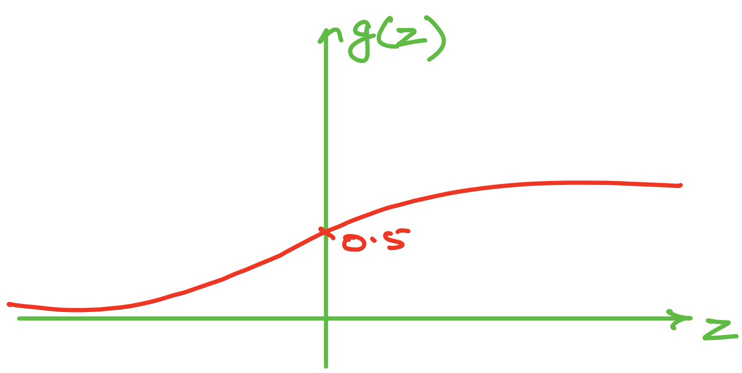
$g(z) = 0 \quad \text{if } z < 0$

PROBLEM ?



Not able to distinguish b/w points which are close to line & which are far.

Sigmoid function:



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

$z = \infty \qquad g(z) = \frac{1}{1+\frac{1}{e^{z}}} \; 0 \quad = 1$

$z = 0 \qquad g(z) = \frac{1}{1+1} = 0.5$

↳ Point is on the line then probability is 0.5

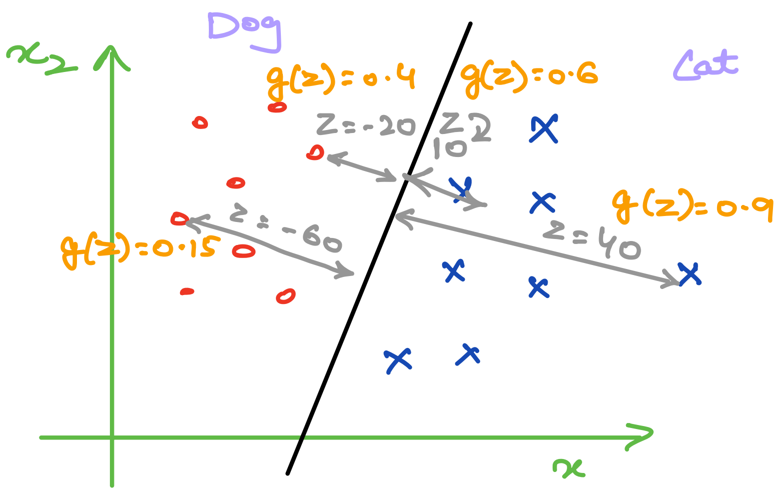
$$z = -\infty \qquad g(z) = \frac{1}{1+\frac{1}{e^{z}}\infty} = 0$$

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

$$= g(z) = \frac{1}{1+e^{-z}} \qquad \text{where } z = \theta^T x$$

Value b/w 0 & 1

Probability / confidence with which you say the point belong to class 1.



g(z) = 0.4

40% sure point ∈ Cat

60% sure point ∈ Dog.

y follows Bernoulli Distribution