Logistic Regression : Supervised learning

It is a $\boxed{\text{classification}}$ technique, having discrete class labels
Objective: To classify a sample represented by features
In classification , the aim is to find the clas label given a set of input features $X \in \mathbb{R}^k$

Examples:
Given a set of emails: classify as SPAM / not SPAM

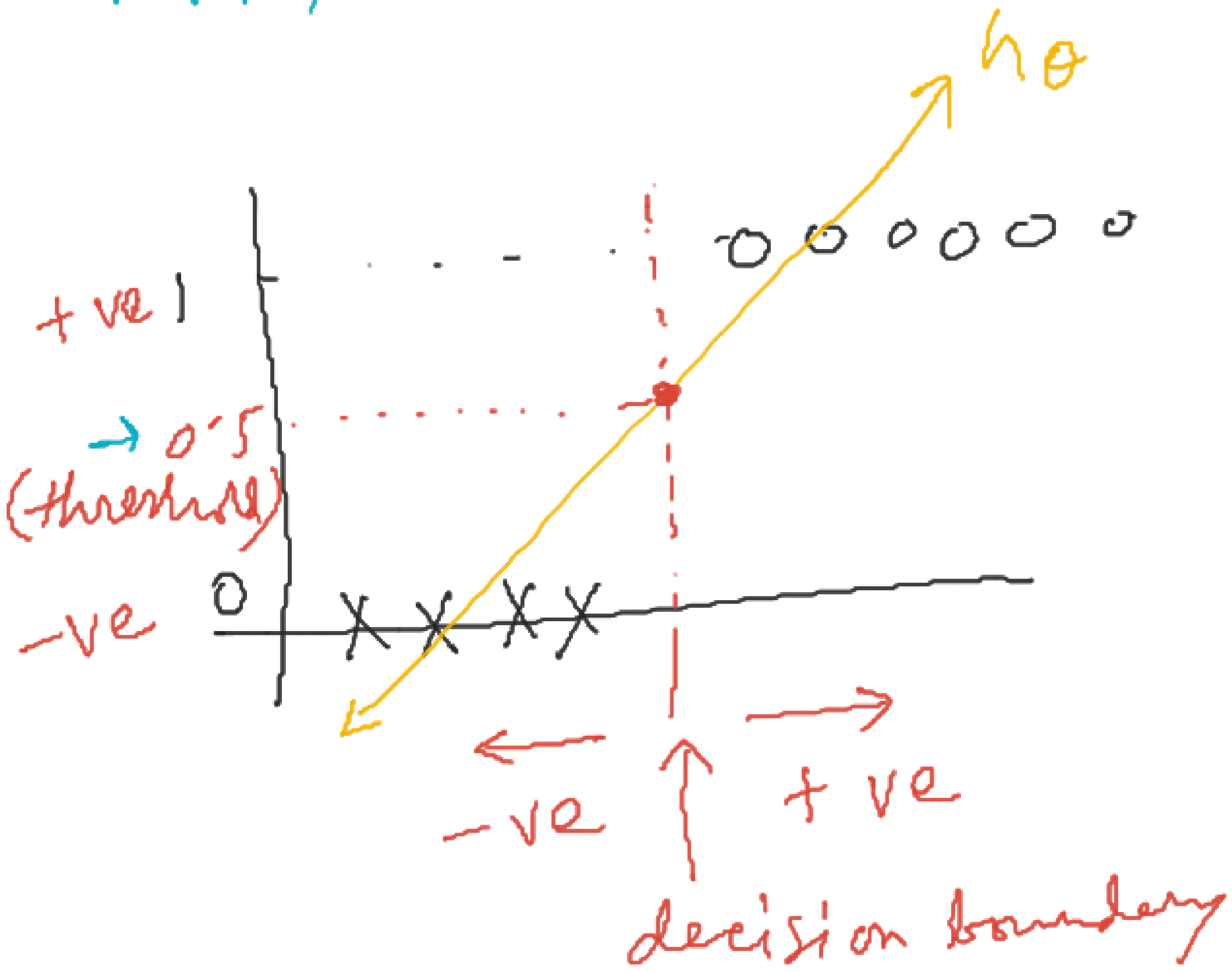
Given a set of images of animals, classify DOG/ CAT

Given barometer's reading, predict the weather as rainy / not rainy

O/P

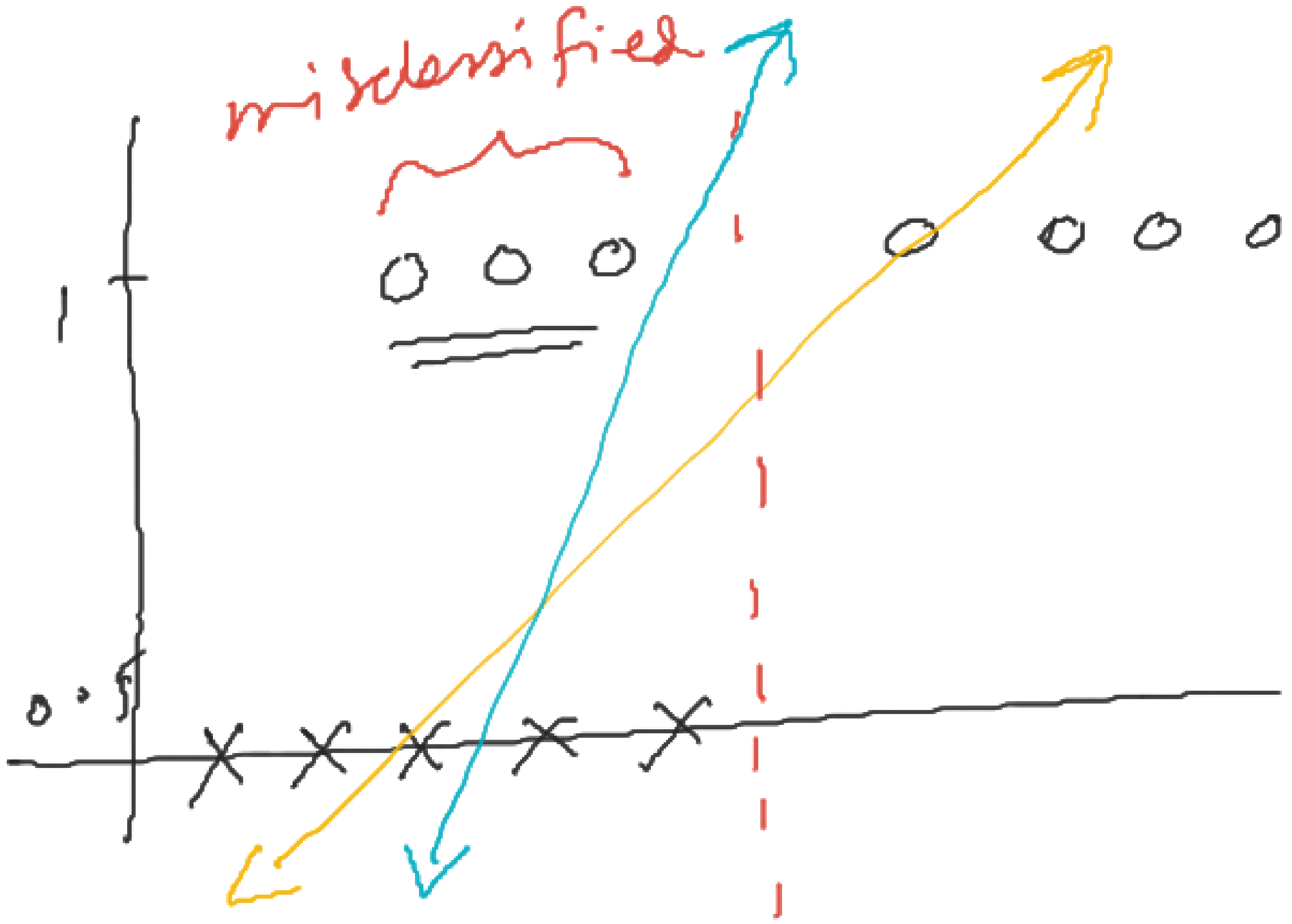
How to extend linear regression for classification?

Apply a threshold on regression o/p h_θ



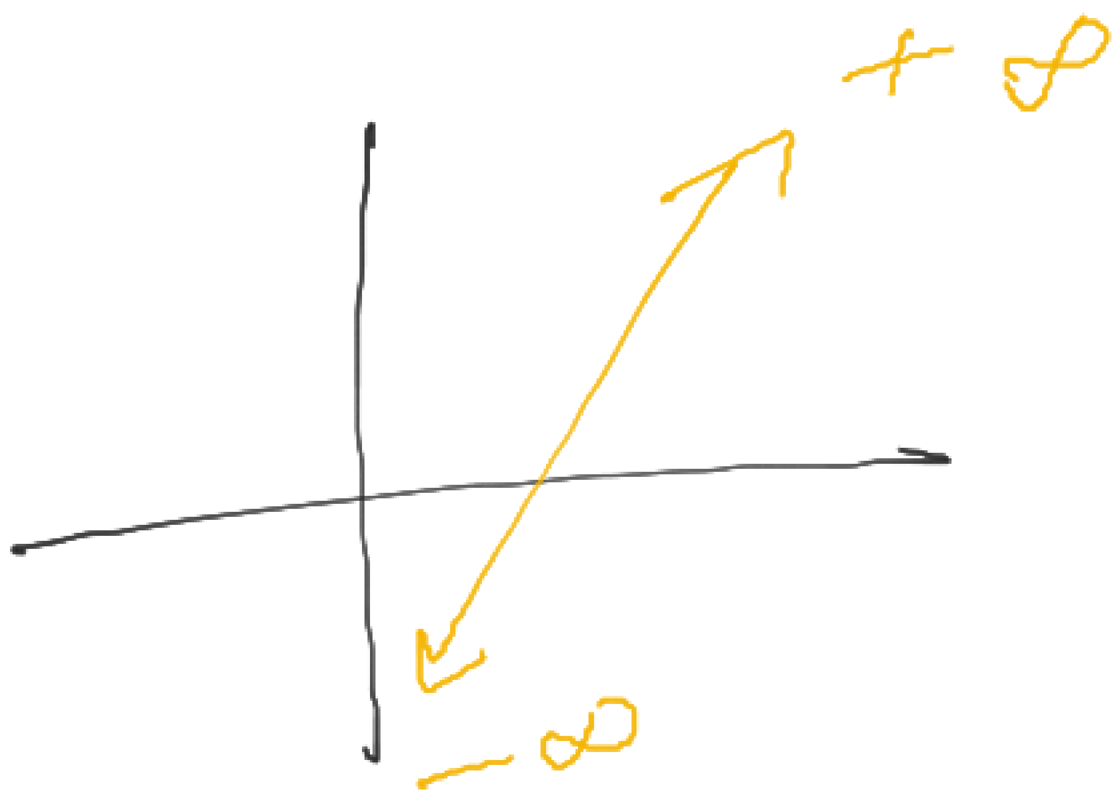
The span of h_θ is from $-\infty$ to $+\infty$

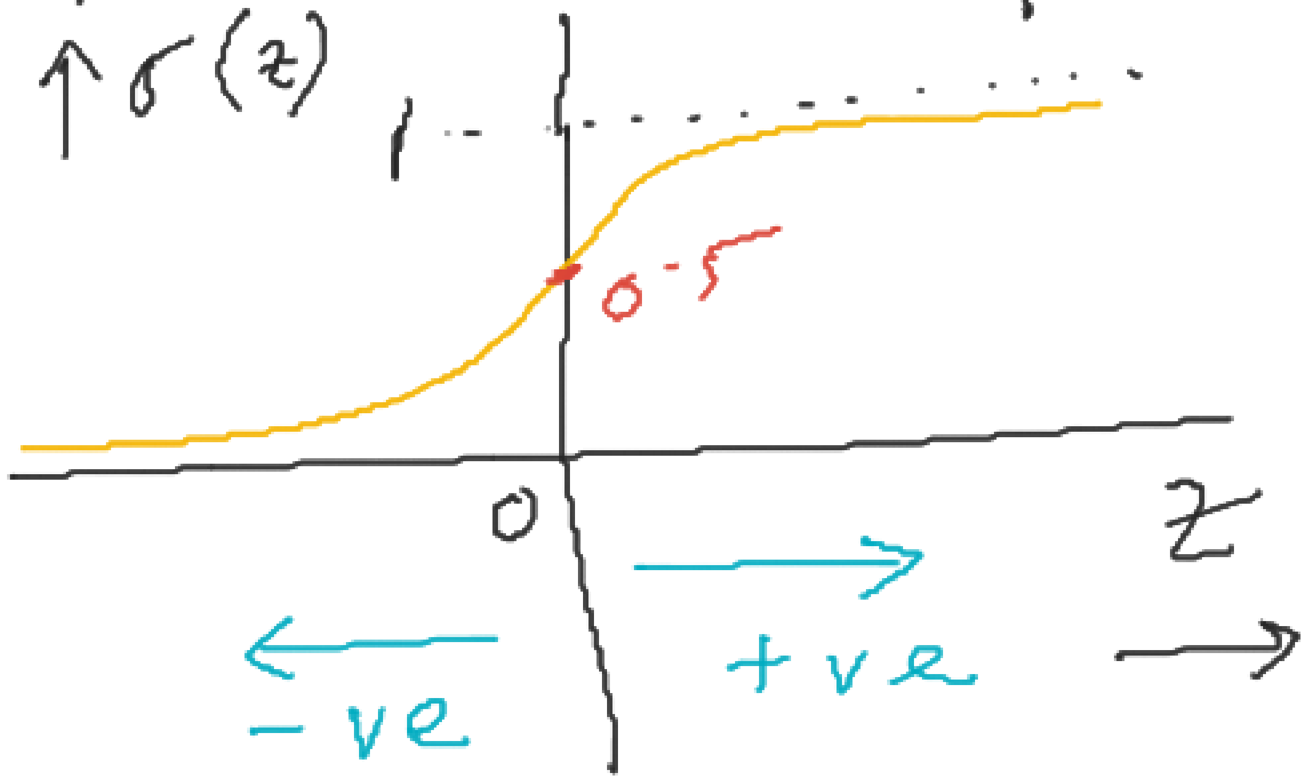
So, the value of h_θ can be much greater than 1 or less than 0



choosing threshold is challenging

Can we bound the hypothesis value?





Logistic f^n / Sigmoid function

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

$$\sigma(0) = \frac{1}{1+e^0} = \frac{1}{1+1} = 0.5 \leftarrow$$

$$0 \leq \sigma(z) \leq 1$$

$z \geq 0$

$\sigma(z) \geq 0.5$

$$z = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_k x_k \quad \Big| \; x_0 = 1$$

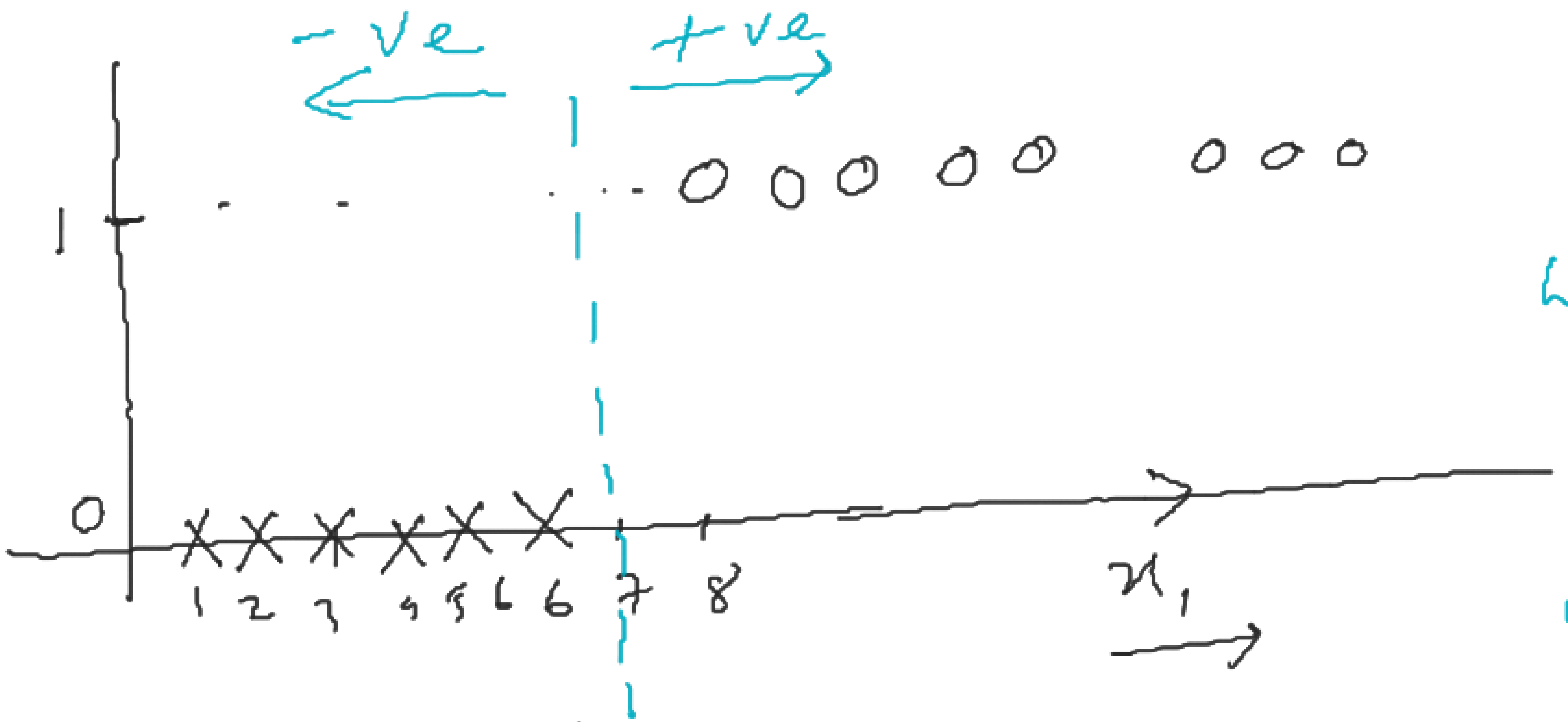
$$= \sum_{i=0}^{k} \theta_i x_i$$

The i/p of Sigmoid is $z = \sum_{i=0}^{k} \theta_i x_i$ for k dimensional features, $x_0 = 1$

$\sigma(z) = 0.5$ can be used as threshold

Sample $X(x_1, x_2, \ldots, x_k) \leftarrow \begin{cases} +\text{ve class}, & \text{if } \sigma(z) \geq 0.5 \leftarrow \\ -\text{ve class}, & \text{otherwise} \end{cases}$

or, $z \geq 0$



X ; -ve sample

o ; +ve "

$$z = \theta_0 + \theta_1 x_1$$

we need to predict θ_0 and θ_1

we assume $\theta_0 = -7$

$\theta_1 = 1$

$$z = -7 + x_1$$

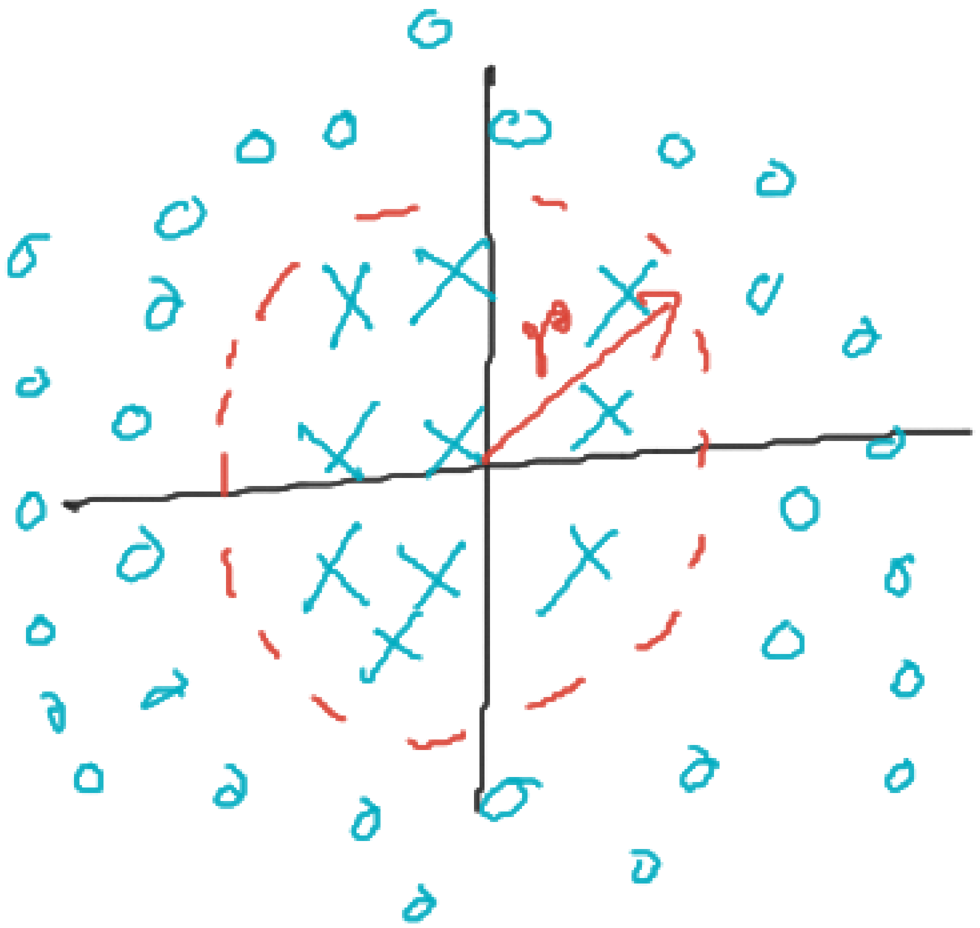
$$\sigma(z) \geq 0.5 \text{ or, } z \geq 0$$

$$-7 + x_1 \geq 0$$

$$x_1 \geq 7$$

Class boundary ; $x_1 = 7$

2



o : +ve class

x : -ve class

$$z = \theta_0 + \theta_1 x_1^2 + \theta_2 x_2^2 + \theta_3 x_1 x_2 + \theta_4 x_1 + \theta_5 x_2$$

Task : to predict $\theta_0, \theta_1, \ldots, \theta_5$

θ_i's : learning parameters

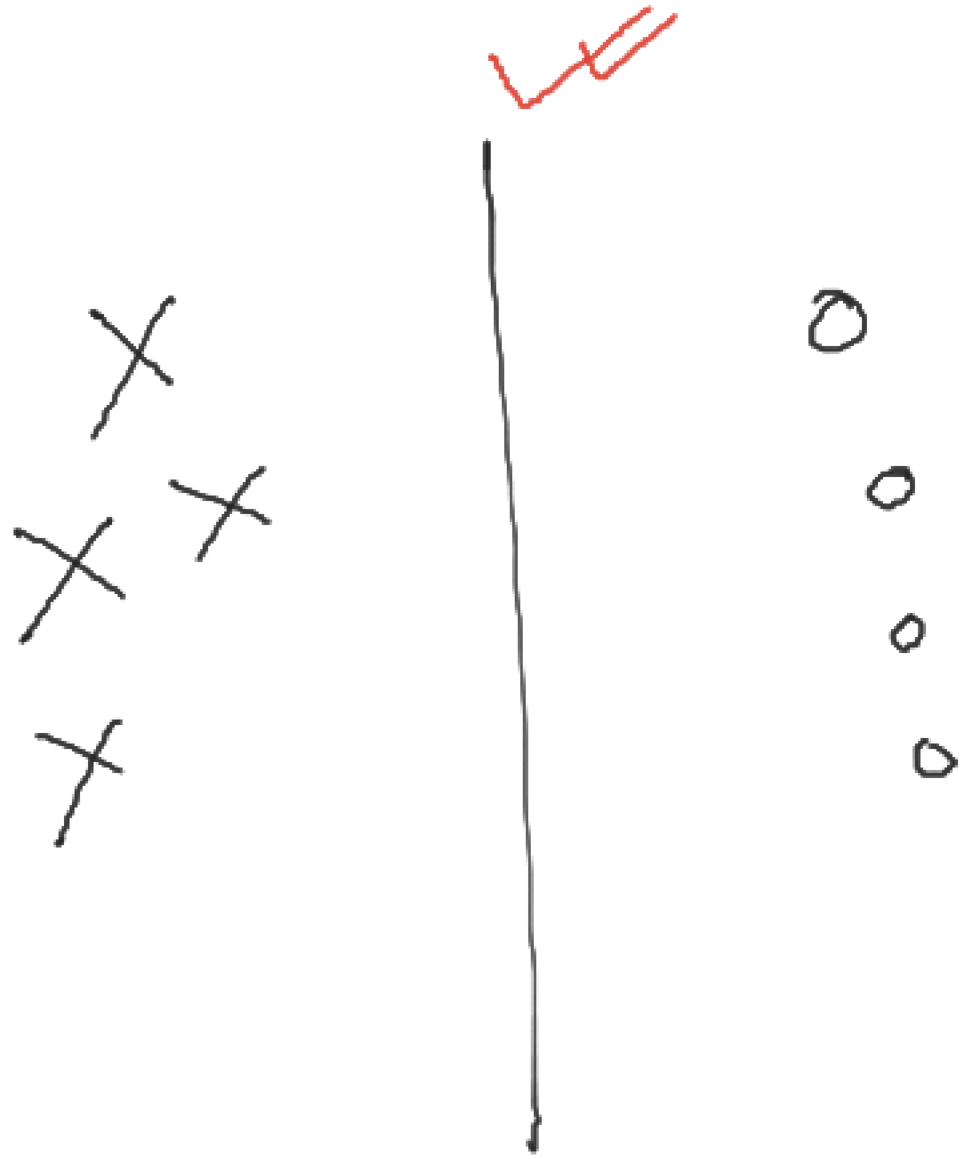
$\longrightarrow$ Assume $\theta_0 = -r^2$, $\theta_1 = 1$, $\theta_2 = 1$, $\theta_3 = \theta_4 = \theta_5 = 0$

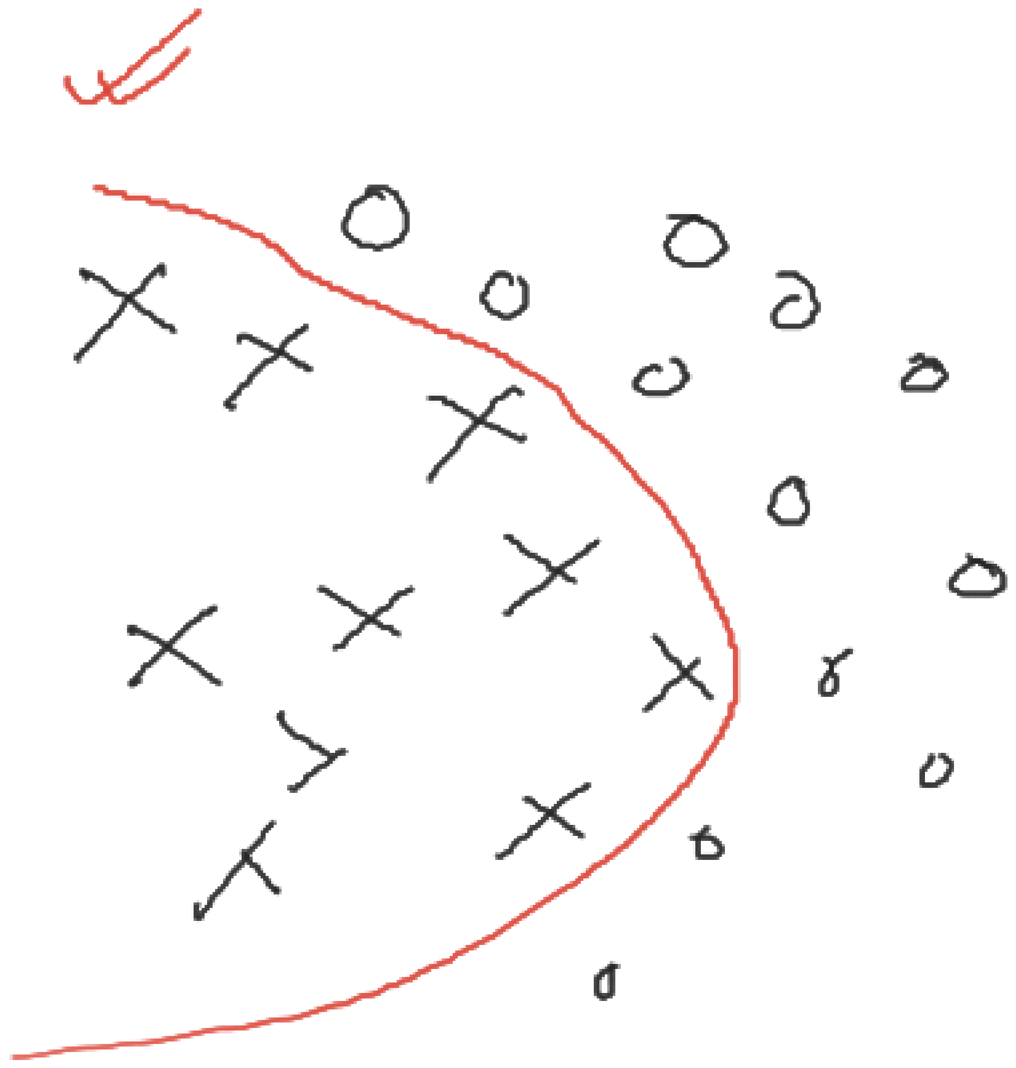
$$z = -r^2 + x_1^2 + x_2^2$$

$\sigma(z) \geq 0.5$ or, $z \geq 0$ or, $-r^2 + x_1^2 + x_2^2 \geq 0$

or, $x_1^2 + x_2^2 \geq r^2$

class boundary : $x_1^2 + x_2^2 = r^2$



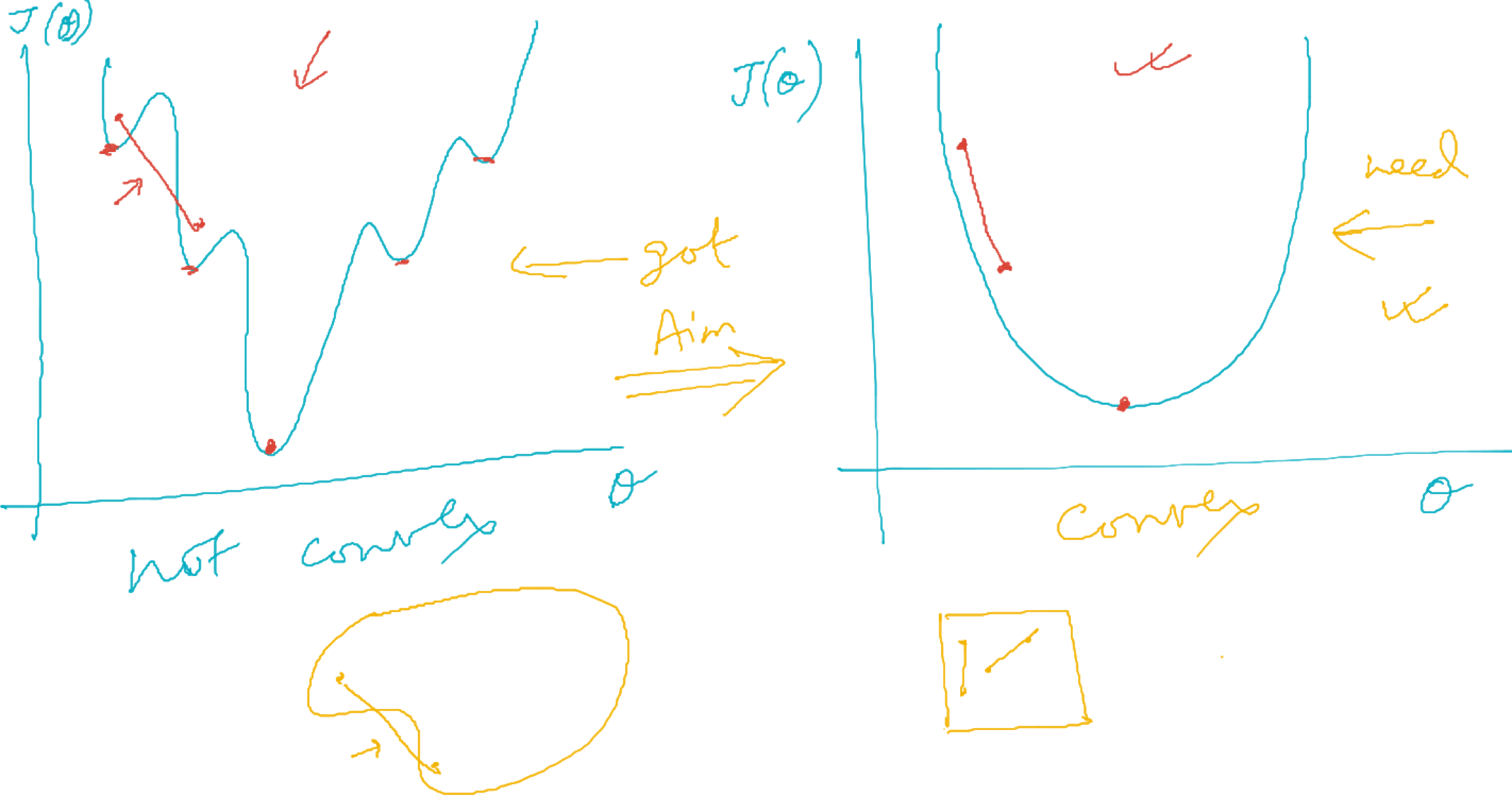


Cost function | $\underline{h_\theta^i} = h_\theta(x^i) = \dfrac{1}{1 + e^{-\sum_{j=0}^{k} \theta_j x_j^i}}$

$j : 1 \to k$ (k-dim feature vector)
$i : 1 \to n$ (n no. of samples)

As per linear reg^n; Cost f^n $\boxed{J_\theta = \dfrac{1}{2n} \sum_{i=1}^{n} (h_\theta^i - y^i)^2}$

J_θ is no longer a convex func^n
— contains multiple local minima

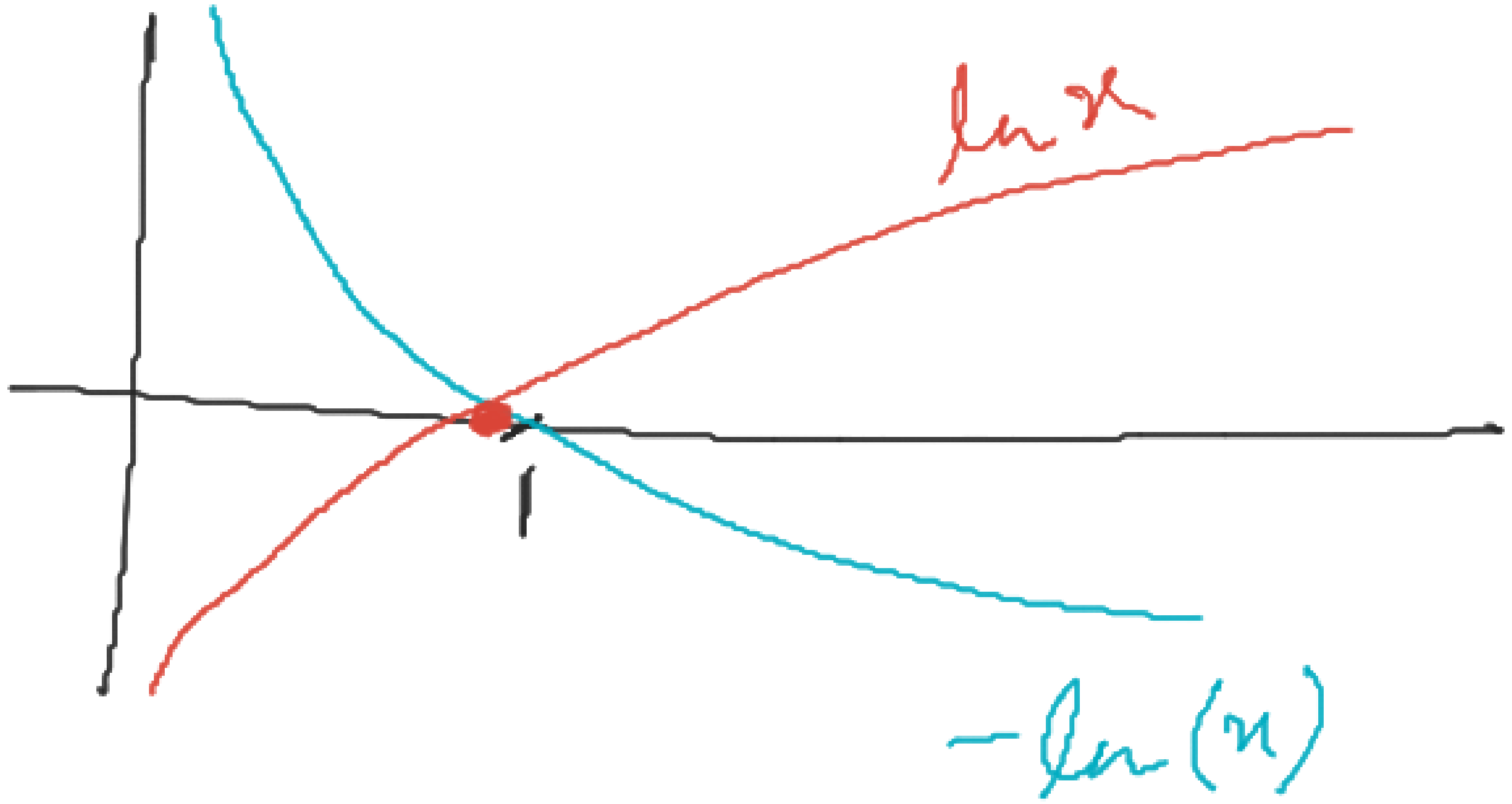


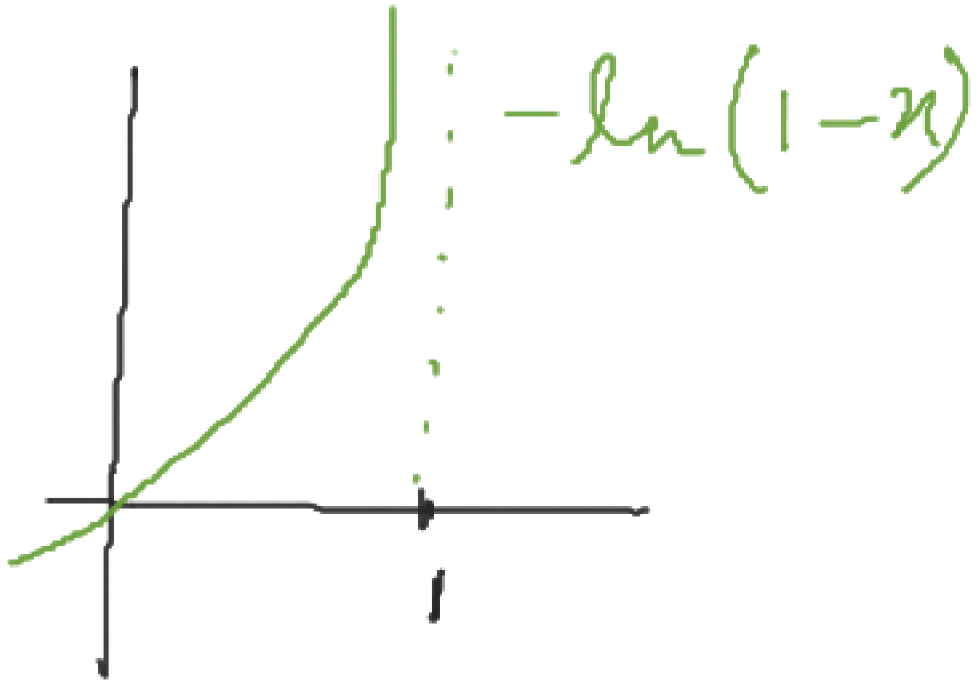
J(θ)
θ
Not convex
got
Aim
J(θ)
need
θ
Convex

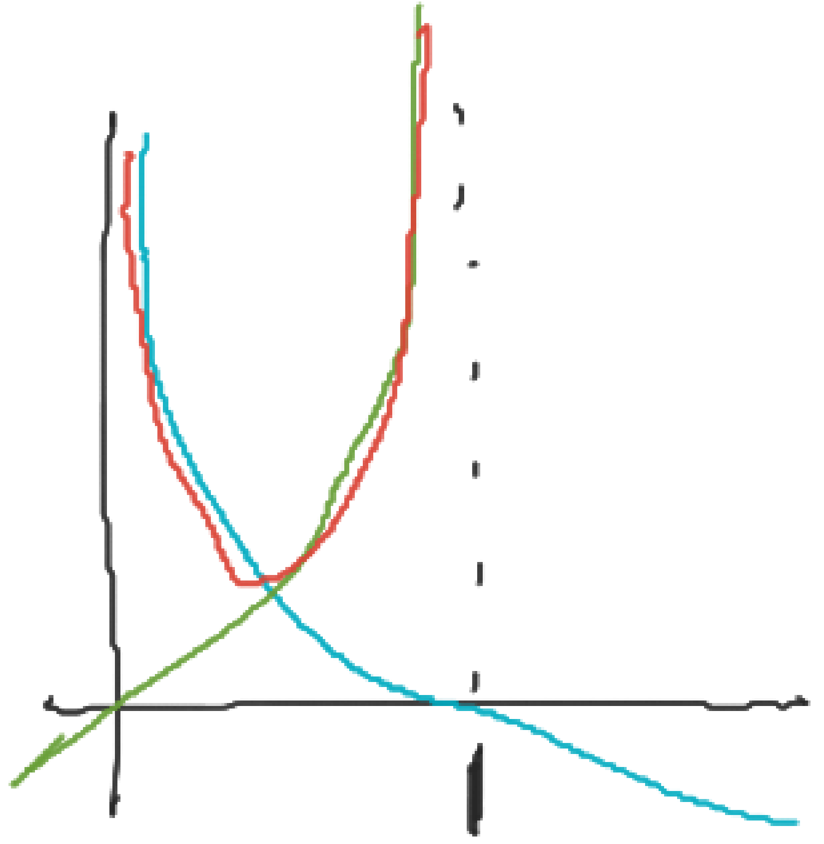
$$J_\theta = \frac{1}{2n} \sum_{i=1}^{n} \underline{\underline{(h_\theta^i - y_j^i)}}^2 = \frac{1}{n} \sum_{i=1}^{n} \underbrace{loss(h_\theta^i, y^i)}$$

function of actual y^i and predicted h_θ^i value

$$loss(h_\theta^i, y^i) = \begin{cases} \underline{-\log(h_\theta^i)} ; \text{if } \underline{y^i = 1} \\ \underline{-\log(1-h_\theta^i)} ; \text{if } \underline{y^i = 0} \end{cases}$$







Combining both :

$$loss(h_\theta^i, y^i) = y^i(-\log(h_\theta^i)) + (1-y^i)(-\log(1-h_\theta^i))$$

Binary Cross Entropy

If You can now perform GD, the updation formula of θ remains same

Task: Try to find the derivative of J_θ with respect to θ_j.

Logistic Regression $\longrightarrow$ Binary classification

Can we extend this for Multi-class classification?

Hint: use multiple times binary classification,

to perform multi-class classification

CAT | Dog → Cat   Dog ⇒ Bin. Cl. ←

input : images   o/p classify

CAT | Dog | Chicken

Cat   Dog & Chicken ⇒ Bin cl 1   0.3

Cat vs. others

one vs rest

not cat

Dog   Cat & chicken ⇒ Bin cl 2   0.2

Dog vs. others

not dog

chicken   Cat & Dog ⇒ Bin cl 3   0.5

chicken vs others

yes chicken

one vs. rest

prev. example, $m = 3$

Multiclass classification

- If you have multiple classes $C_1, C_2, \ldots, C_m$
- for each class C_i, $\forall i = 1, 2, \ldots, m$, classify C_i versus Rest
- classify a sample X in class

$$\max_{j=1,2,\ldots,m}\left(h_\theta^j(X)\right)$$