Data Science and Artificial Intelligence

Machine Learning

Classification

Lecture No. 3

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

Topic Linear classification

Topic

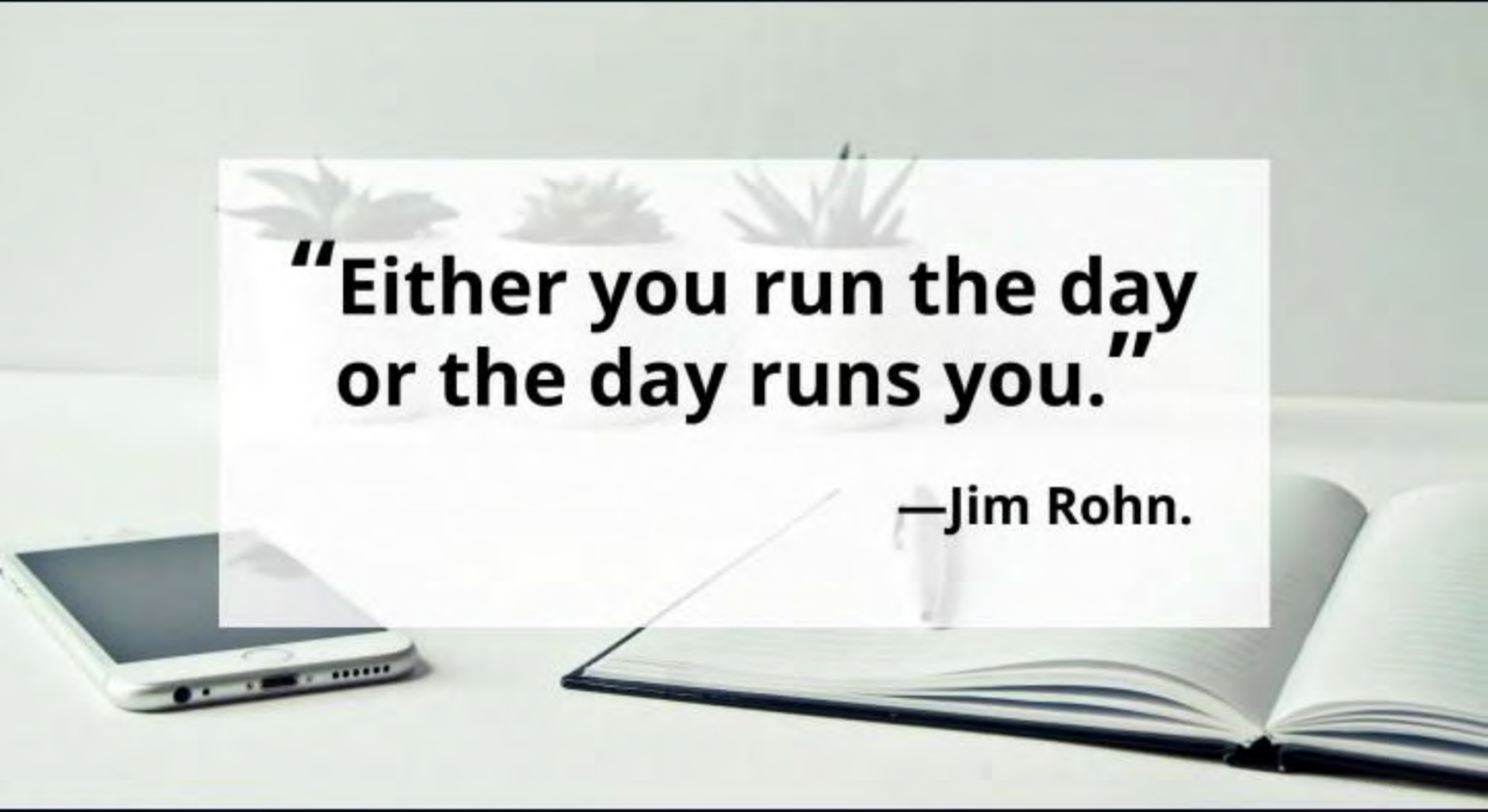
Topic

Topic

Topic

Topics to be Covered

- **Topic** Linear Classification
- **Topic** Problem
- **Topic** Logistic Regression
- **Topic**
- **Topic**


"Either you run the day
or the day runs you."
—Jim Rohn.

Linear Classification : The Loss function

loss fxn → $\max \sum_{i=1}^{N} (y_i x_i \beta)$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_D \end{bmatrix}$$

$$x_i = [1, x_i^1, x_i^2, ----]$$

Line shifted

REVISION

Linear Classification : The Problem

↳ (outlier')

So to max $\sum y_i x_i \beta$

- Line will get shifted

Linear classifier eq

$\Rightarrow \max \sum_{i=1}^{N} y_i x_i \beta = L$

$\hookrightarrow \sum_{i=1}^{N} y_i (\beta_0 + x_i^1 \beta_1 + x_i^2 \beta_2 + \ldots) = L$

$$\left.\begin{array}{l} \dfrac{\partial L}{\partial \beta_0} = 0 \\ \dfrac{\partial L}{\partial \beta_1} = 0 \end{array}\right\}$$

- we will not get any Result
- Like we get in LR, RR

How algorithm Find β

Only for thinking

• not for exam

Let take 2D data

X^1	X^2	Y_i
a	b	1
c	d	1
e	f	-1
g	h	-1
i	j	1

$y_i(x_i\beta)$ shd be +ve

$$\beta_1 a + \beta_2 b + \beta_0 > 0$$

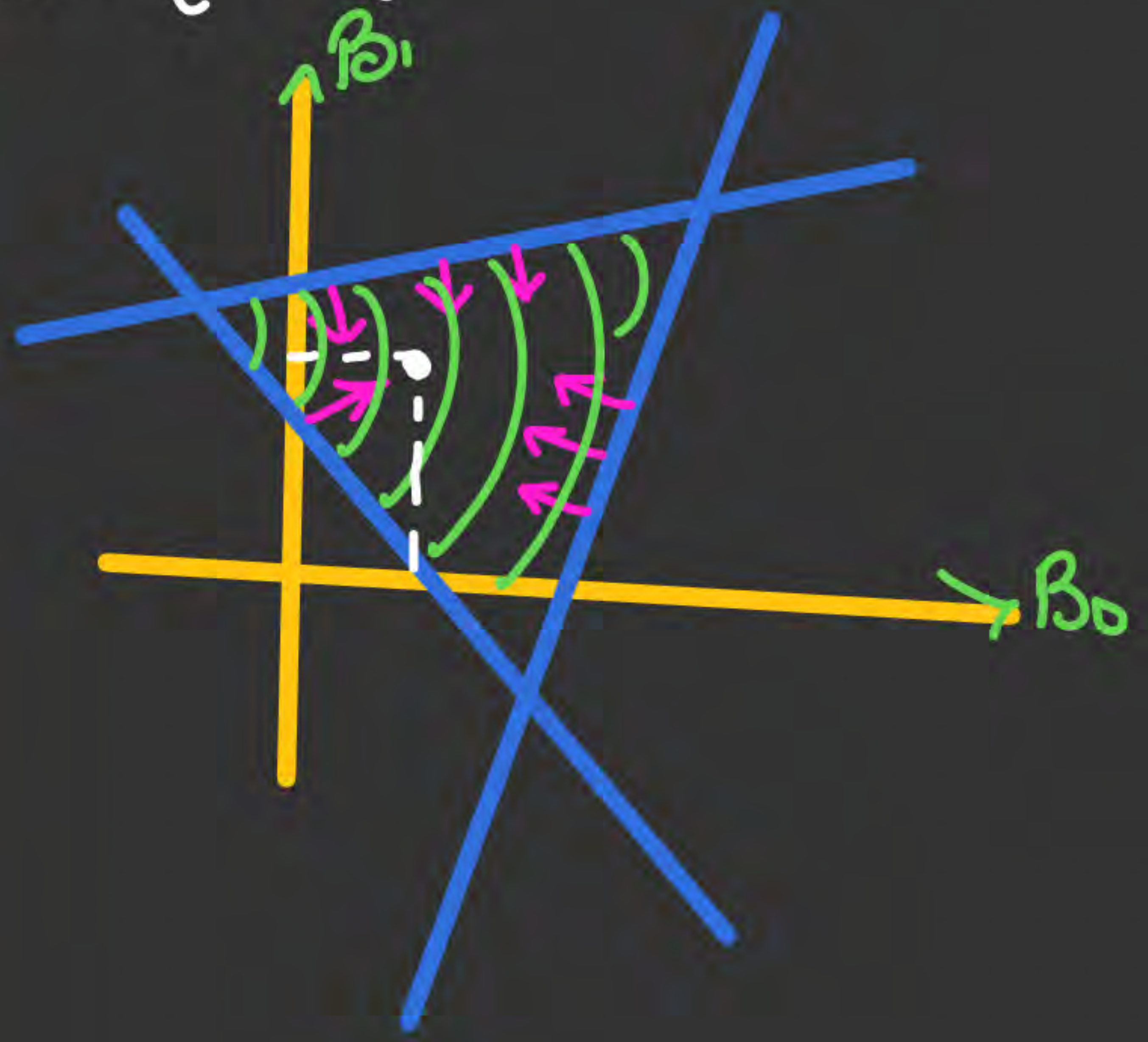
$$\beta_1 c + \beta_2 d + \beta_0 > 0$$

$$\beta_1 e + \beta_2 f + \beta_0 < 0$$

$$\beta_1 g + \beta_2 h + \beta_0 < 0$$

- The algo create all Planes and find region Common to all
- So from any point in the Common Region we can take $\beta_1, \beta_0, \beta_2$ & Classifier

Illustration



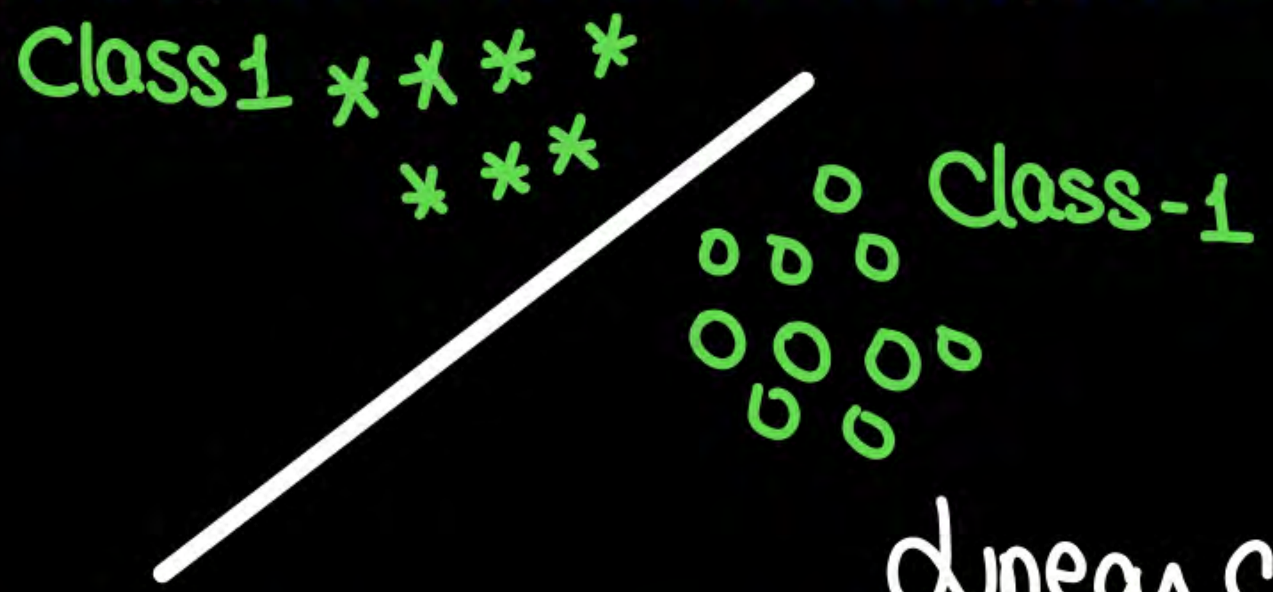
Linear Classification

You are given a trained ~~Logistic Regression~~ model with the following numerical weight vector and bias:

weight vector (w): [0.8, –1.2]

Bias (b): –0.5

You need to classify four points (A, B, C, D) using this model. The data points and their respective feature vectors are as follows:



Point A: (3, 5]

Point B: (–2, 4]

Point C: [1, –1]

Point D: [–4, –3]

linear c.

Which points will be classified as Class 1 (positive class) using this ~~Logistic Regression~~ model?

So class 1 $\Rightarrow$ positive class

$$(x_i \beta) > 0$$

$$\left(\beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2\right)$$

$$-.5 + .8 x_i^1 - 1.2 x_i^2$$

x^1	x^2		
3	5	-4.1	$\Rightarrow$ Class-1
-2	4	-11	$\Rightarrow$ Class-1
1	-1	$-.9$	$\Rightarrow$ Class-1
-4	-3	$-.1$	$\Rightarrow$ Class-1

So Concept

$$x_i \beta \rightarrow \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 \, \_ \, \_ \, \_$$

+ve → class 1

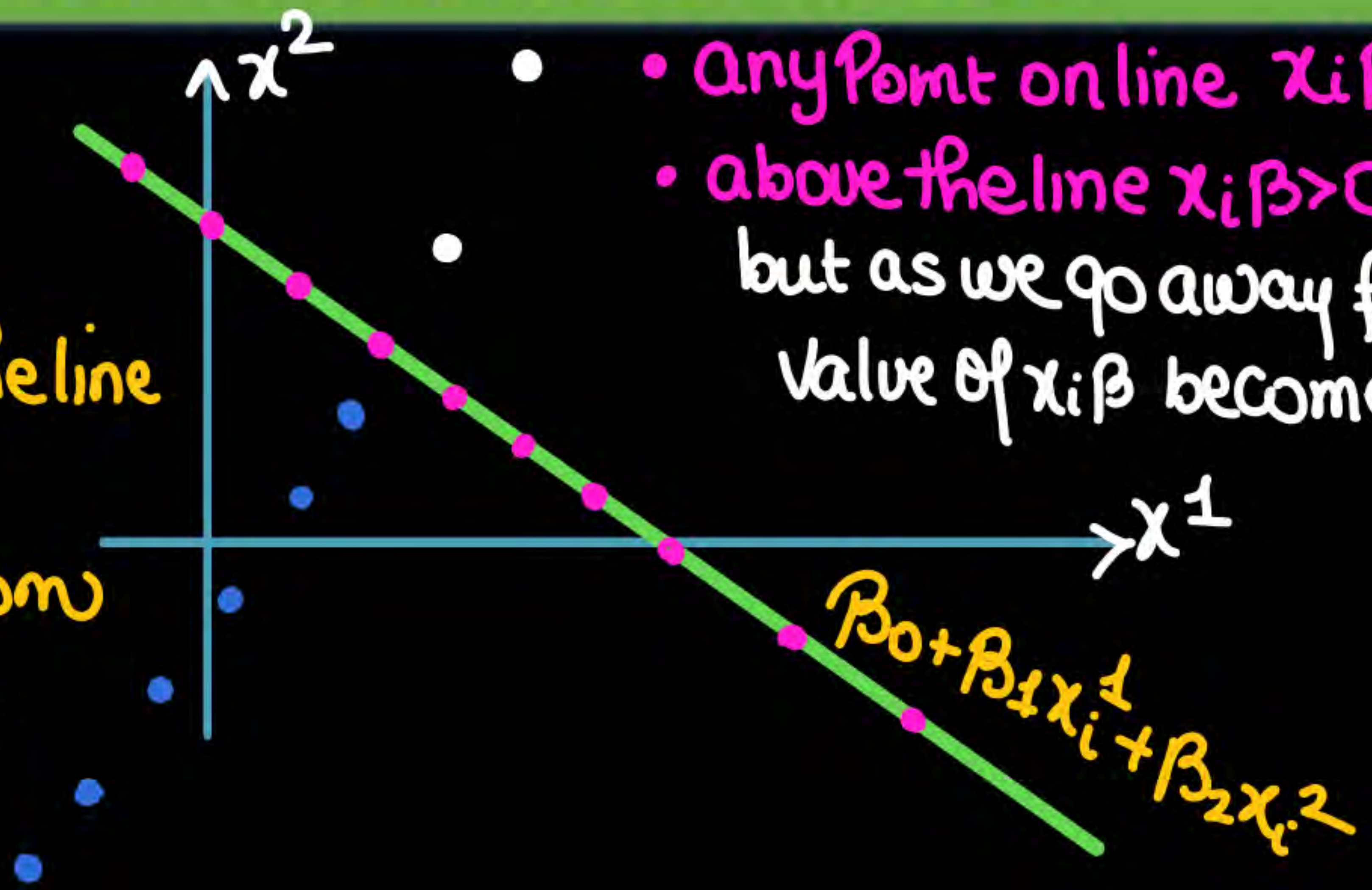
-ve → class -1

Linear Classification

Distance of a point from a line ...

Geometry ⇛

- Similarly below the line $x_i\beta$ ⇛ -ve but as we go away from line $x_i\beta$ become more -ve.



- any Point on line $x_i\beta=0$
- above the line $x_i\beta>0$

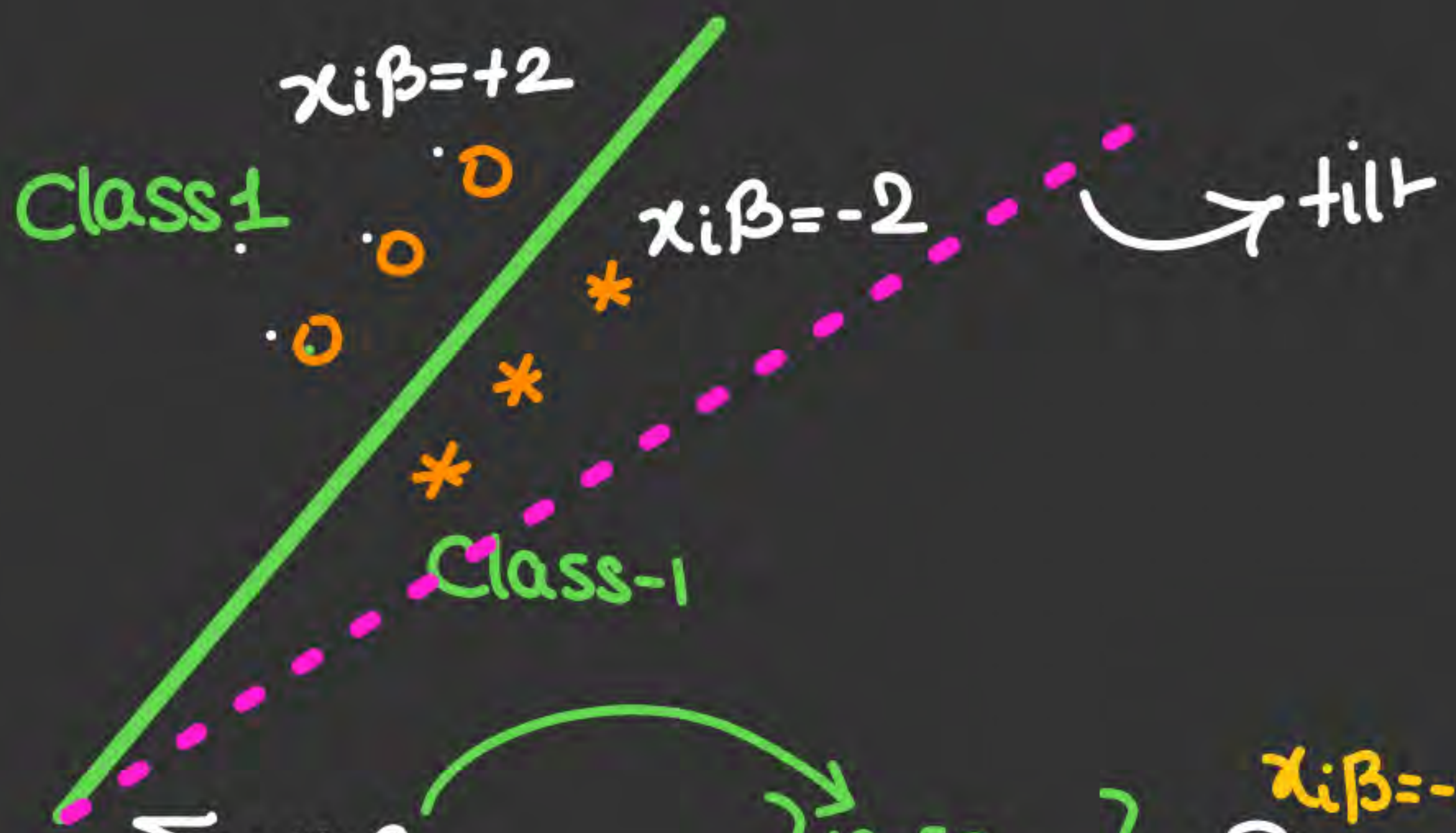
but as we go away from line value of $x_i\beta$ become highly +ve

- Outlier

effect the classifier becoz of huge value of $x_i\beta$

- Cost fxn $\Rightarrow \max \sum_{i=1}^{N} y_i x_i \beta$

- generally all points are close
- $x_i \beta$ is small
- But outlier have large $x_i \beta$, so classifier tilts to reduce $x_i \beta$ for outlier



$\sum y_i x_i \beta$

$\Rightarrow 2\times 1 + 2\times 1 + 2\times 1$

$-2\times -1 + -2\times -1 + -2\times -1$

$\Rightarrow 12$

$12 - 50$

$= -38$

$x_i\beta = -50$

Class 1

$\sum y_i x_i \beta$

$6\times1+6\times1+6\times1$

$+2\times-1+2\times-1+2\times-1$

$+1\times(-25)$

$12-25$

$\Rightarrow -13$

Class 1

Why $\max \sum_{i=1}^{N} y_i x_i \beta$

Classifier: class 1 → oopar → $x_i \beta > 0$

Class -1 → neeche → $x_i \beta < 0$

⇒ $y_i x_i \beta$ shd be +ve for good classifier

So Best classifier $\sum_{i=1}^{N} y_i x_i \beta$ max

Linear Classification

- How can distance decide the probability of a point being class 1/0

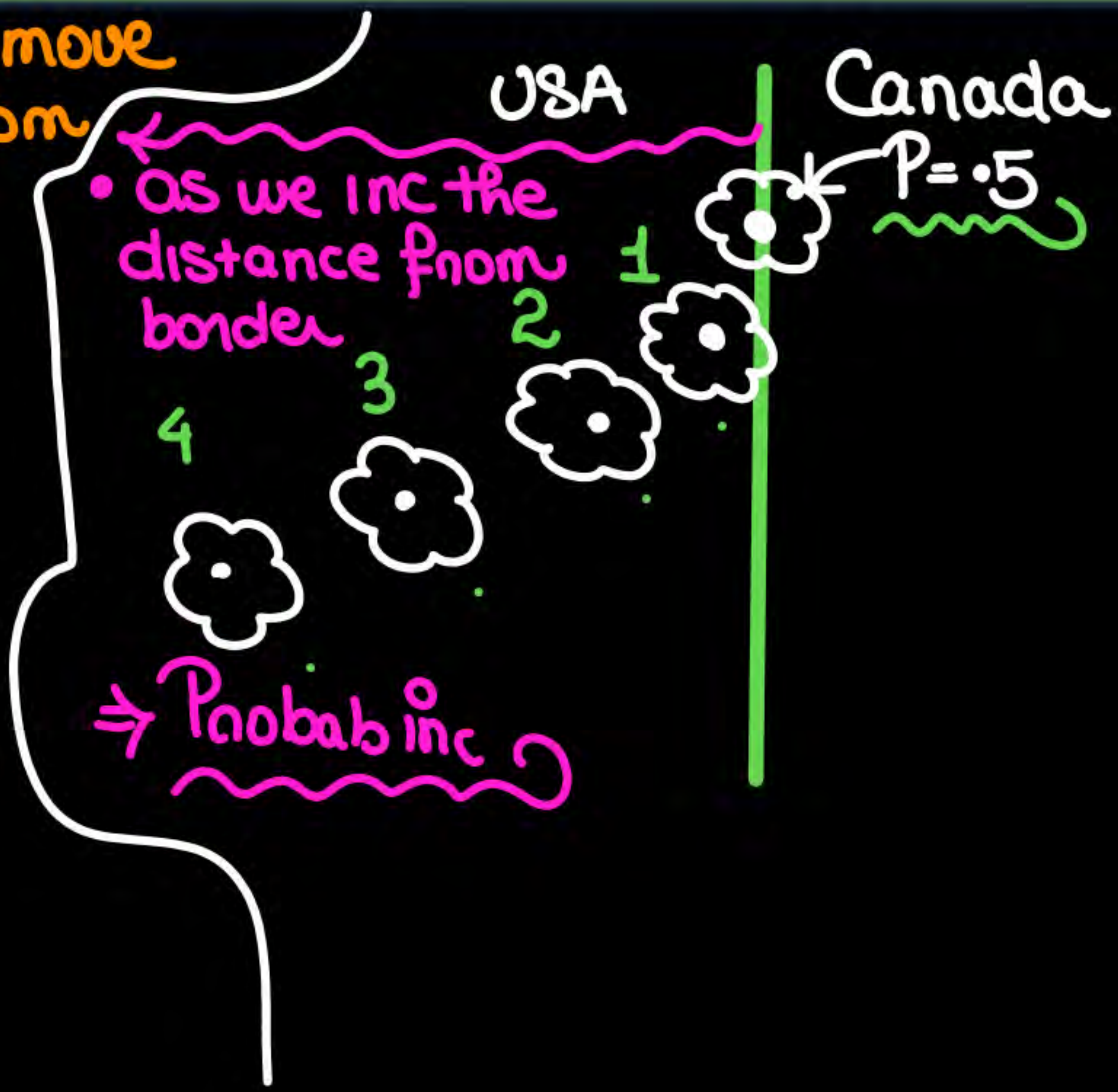
oopar class 1

neeche class 0

as we move away from line P_1 inc

@ boundary $P_1 = .5$, $P_0 = .5$

As we move below the line P_1 dec



Classifier

$x_i\beta$ more and more +ve
P_1 inc and $P_1 = 1$

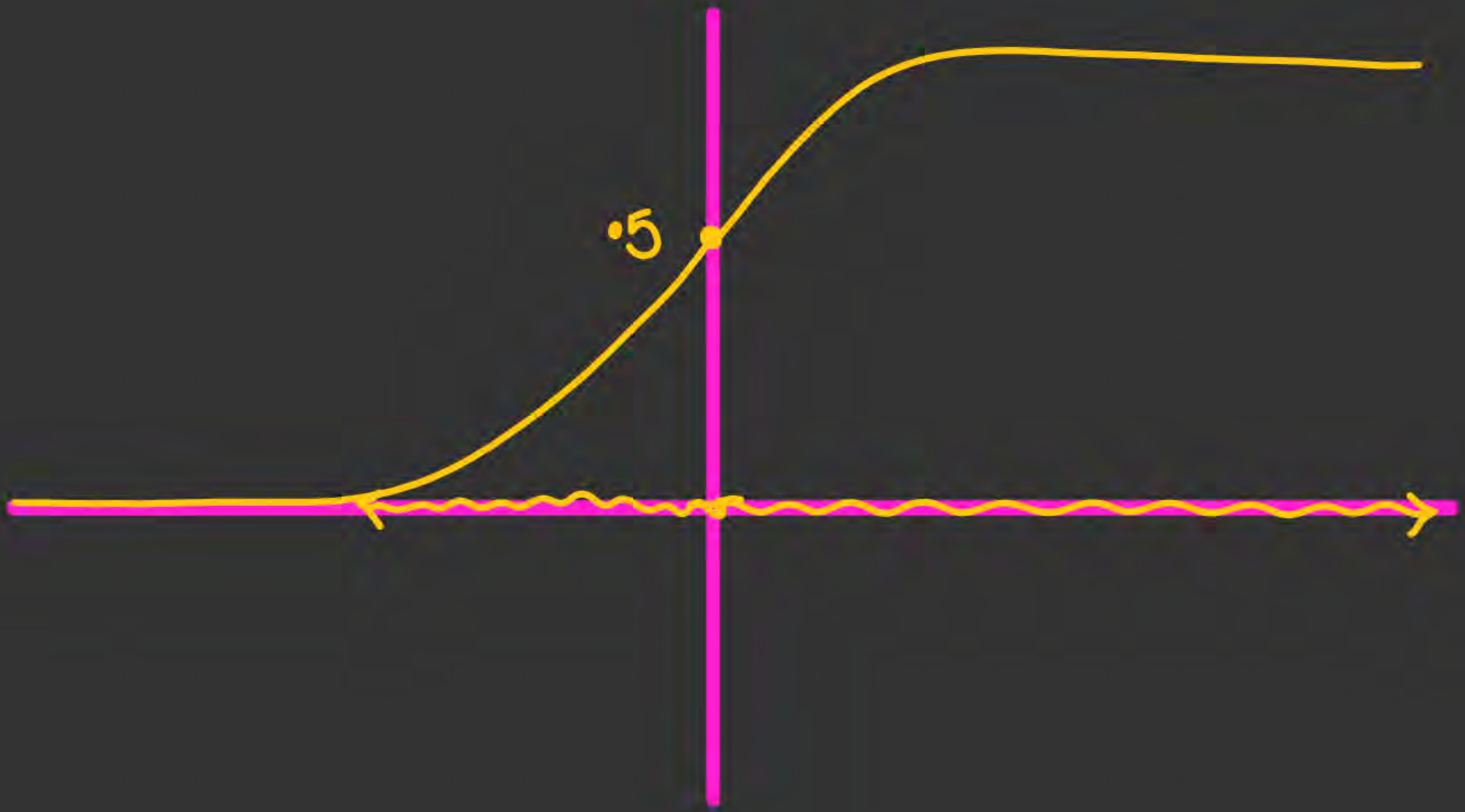
On the line $P_1 = .5$
$P_0 = .5$

- as $x_i\beta$ become -ve then P_1 dec

$x_i\beta = 0$

- Now we Relate $x_i\beta$ with Probab that x_i Belong to Class 1

$$\boxed{P_0 + P_1 = 1}$$

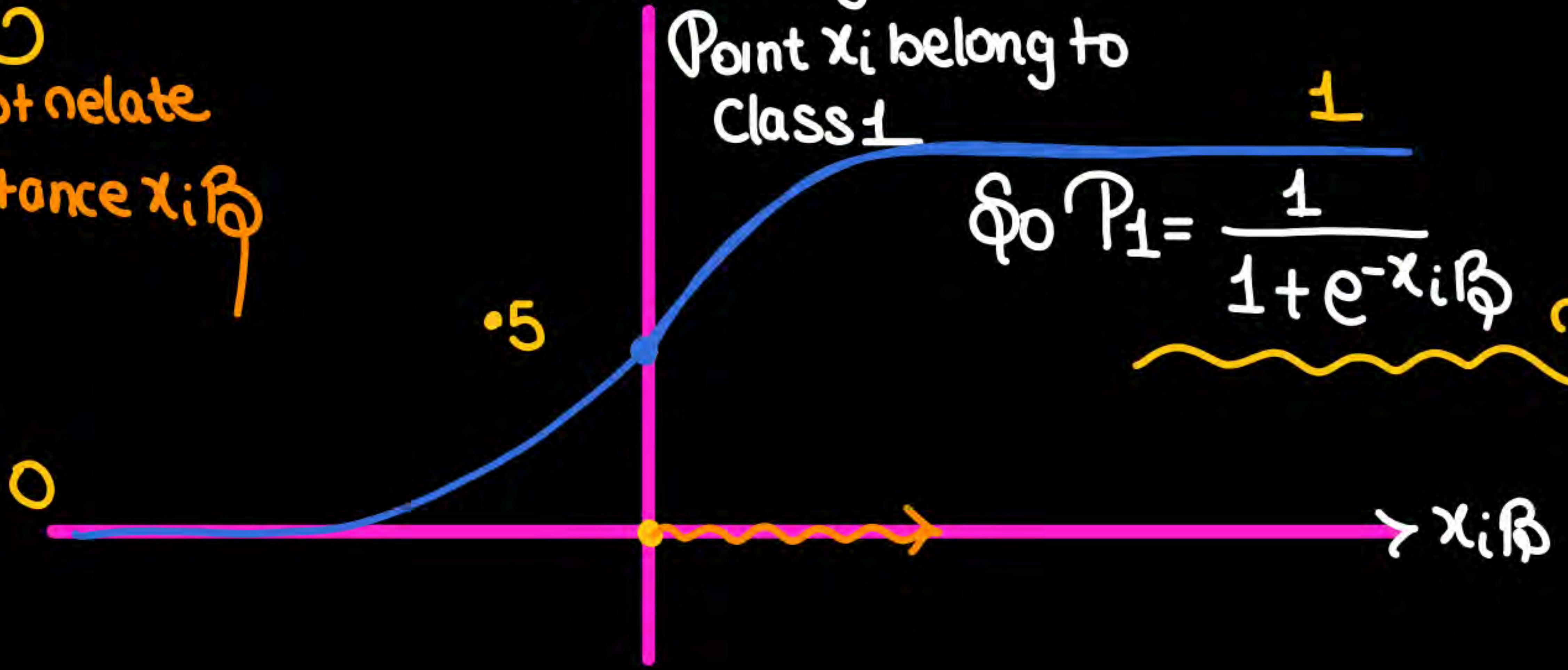

.5

Linear Classification

Probability

- So this plot relate P_1 to distance $x_i\beta$

- Probability that Point x_i belong to Class 1



So $P_1 = \dfrac{1}{1+e^{-x_i\beta}}$

So $P_1 \Rightarrow \dfrac{1}{1+e^{-x_i\beta}} \rightarrow$ Sigmoid function

$x_i\beta = \infty \Rightarrow P_1 = \dfrac{1}{1+e^{-\infty}} = \dfrac{1}{1} = 1$

$x_i\beta = 0 \Rightarrow P_1 = \dfrac{1}{1+e^{-0}} = \dfrac{1}{1+1} = \dfrac{1}{2}$

$x_i\beta = -\infty \Rightarrow P_1 = \dfrac{1}{1+e^{\infty}} = \dfrac{1}{\infty} = 0$

What will be

$P_0 \Rightarrow 1 - P_1$

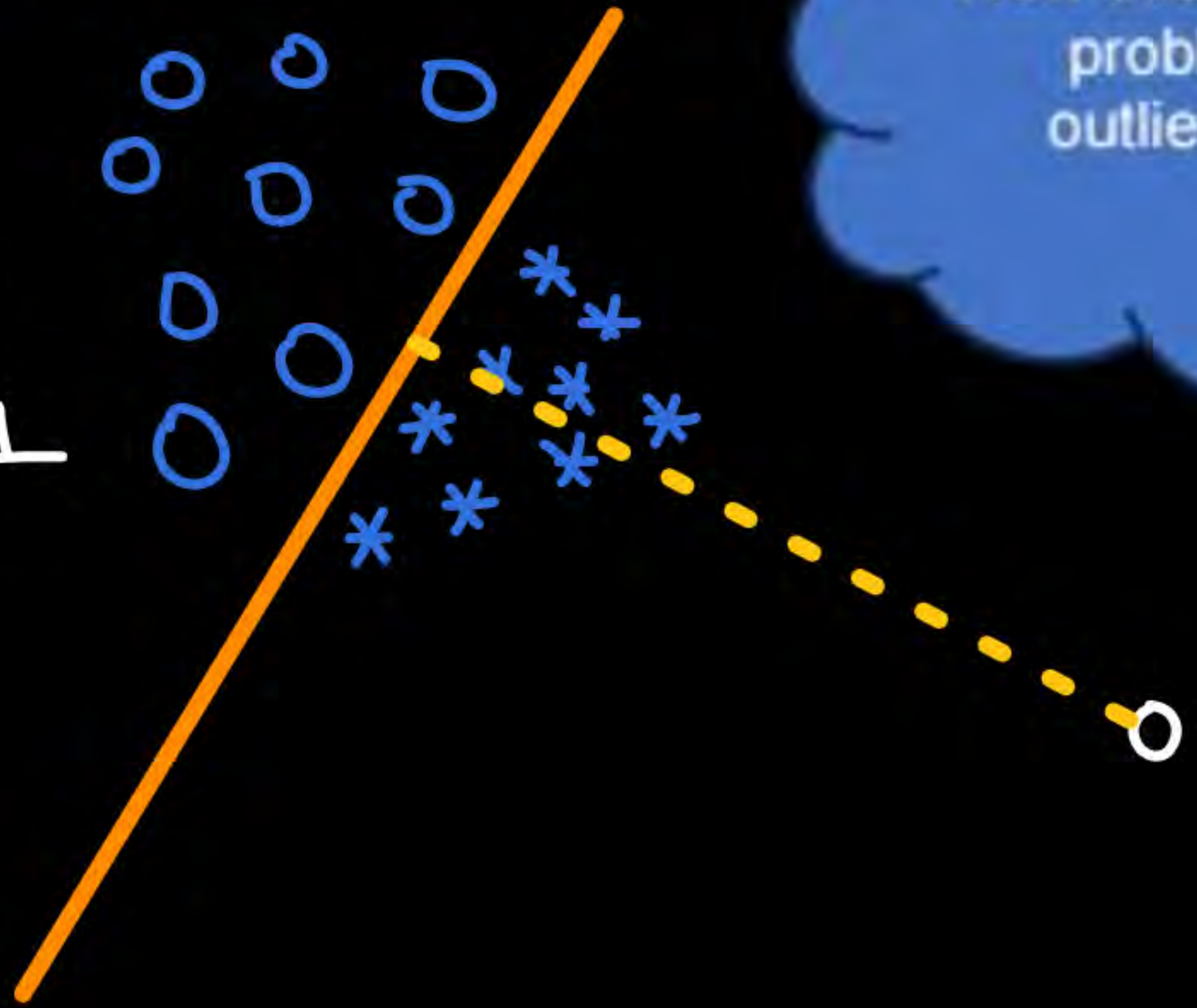
$\Rightarrow 1 - \dfrac{1}{1+e^{-x_i\beta}}$

$P_0 \Rightarrow \dfrac{e^{-x_i\beta}}{1+e^{-x_i\beta}}$

Linear Classification

So we Convert distance $x_i B$ into Probablity

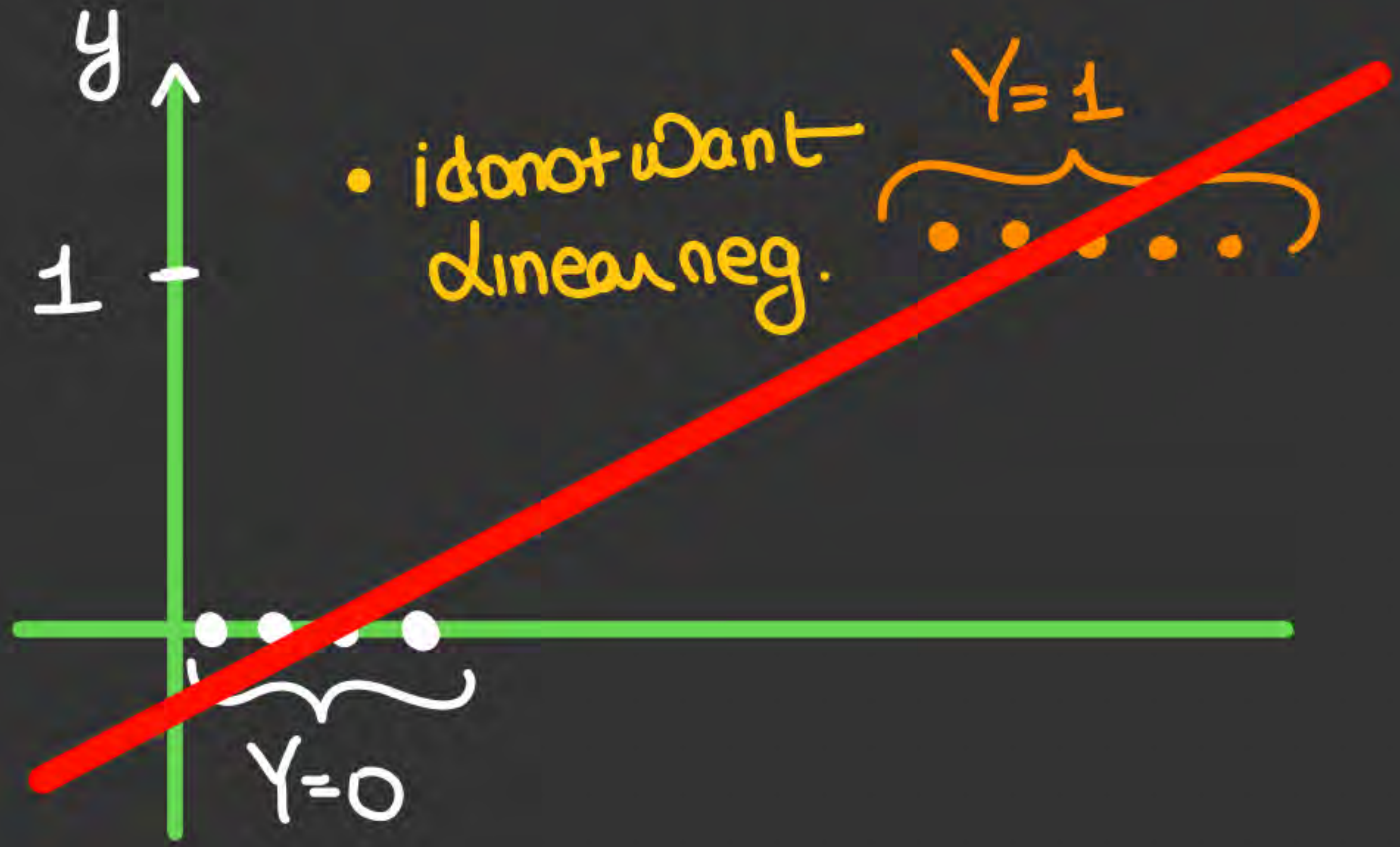
- So in this Case the outlier which have large $x_i B$, the P_1 will always be b/w 0 to 1 Only.



How this solve the problem of outliers...??

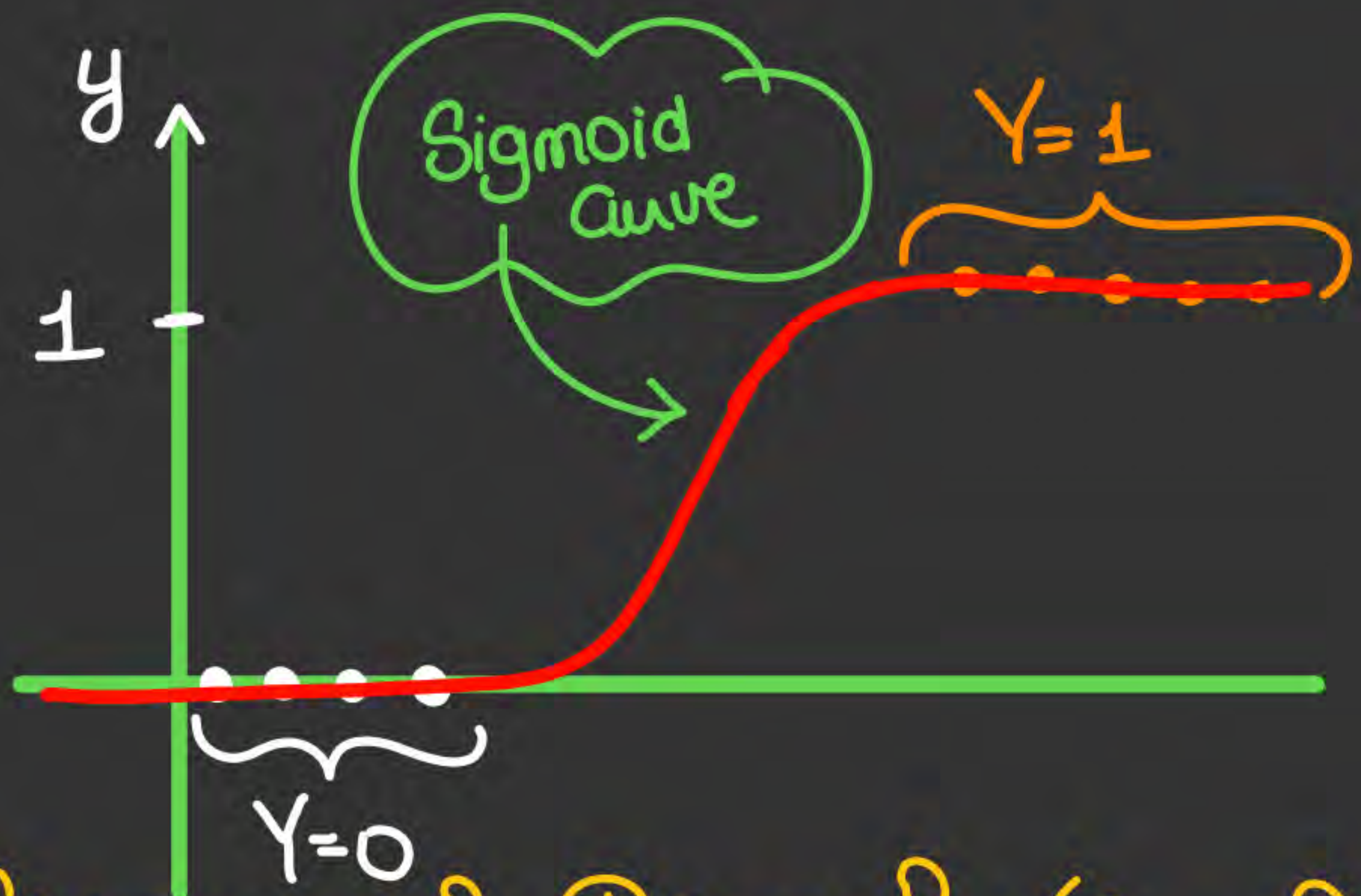
Feel the logistic Regression

data	Y
data1	1
data2	1
⋮	0
⋮	0
⋮	0
	1
data7	1



Feel the logistic Regression

data	Y
data1	1
data2	1
⋮	0
⋮	0
⋮	0
	1
data7	1

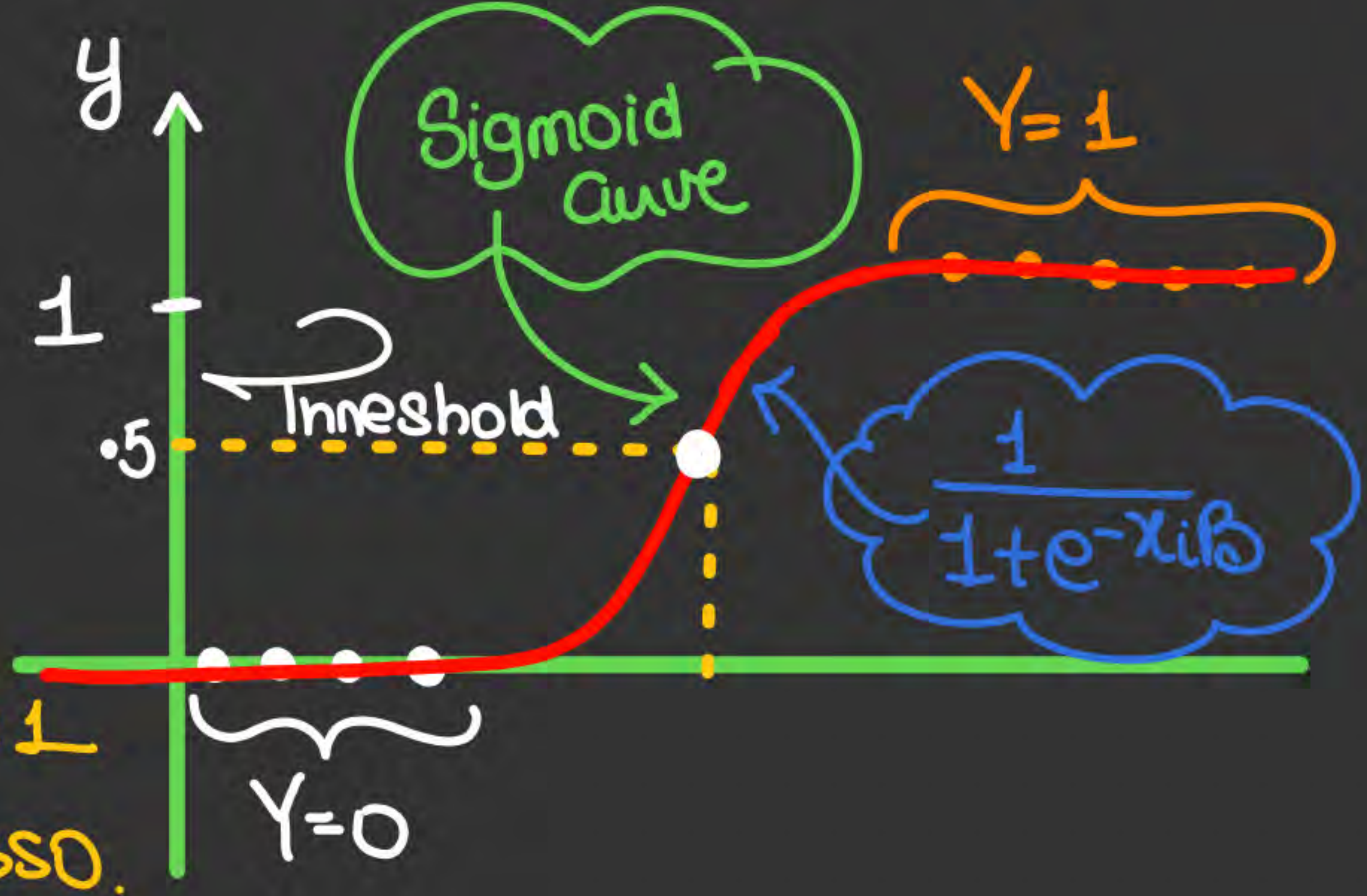


- So we are doing Regression (Curve fitting)
- 'S' type Curve fitted on data.

Basic Rule to decide the Class of a new point

So for any new point find Y ⇒ if Y > .5 ⇝ Class 1

Y < .5 ⇝ Class 0.



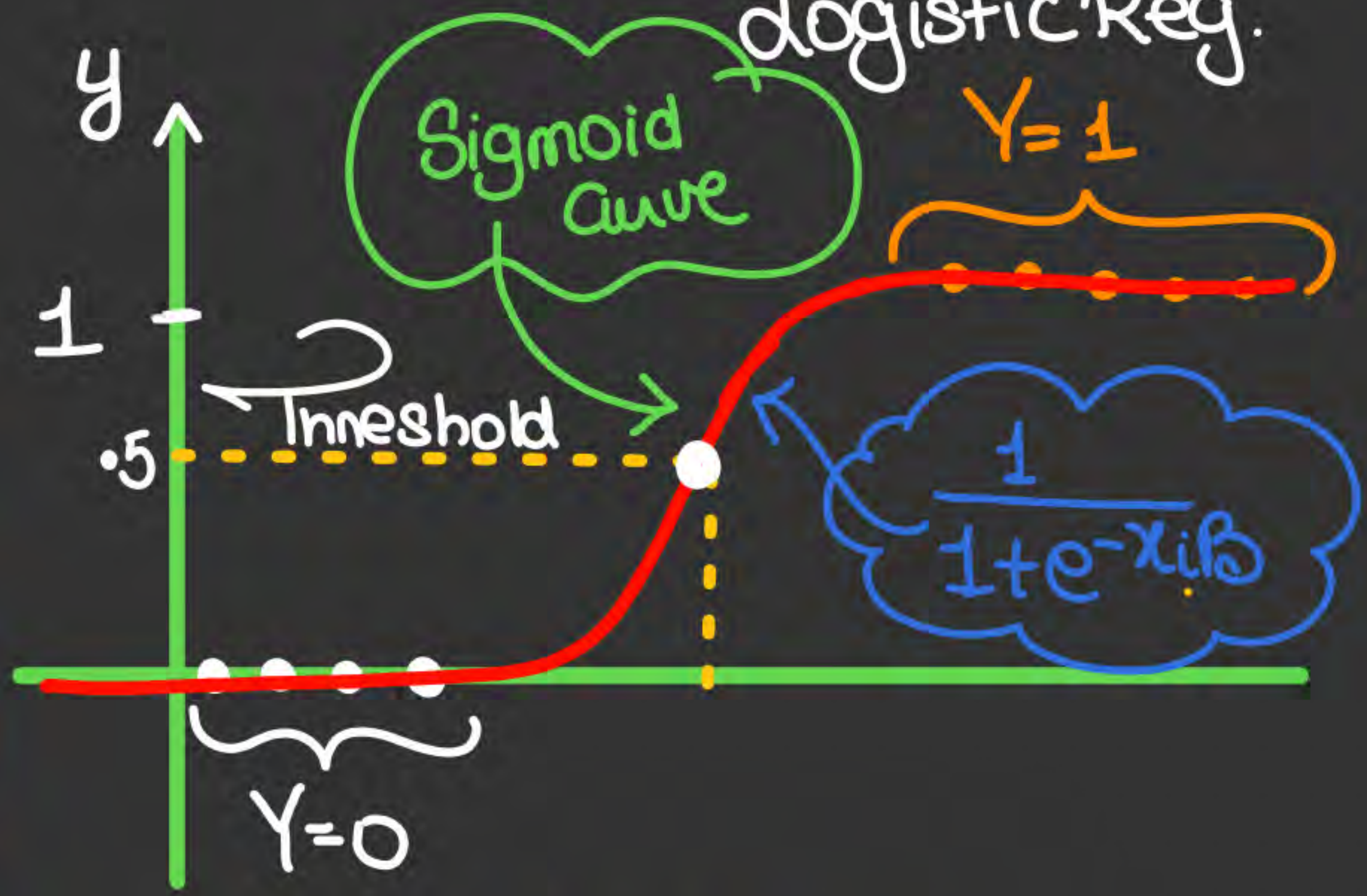
Linear classifier

$x_i\beta > 0 \rightarrow$ class1

$x_i\beta < 0 \rightarrow$ class0

Cost fxn $\max \sum_{i=1}^{N} y_i x_i\beta$

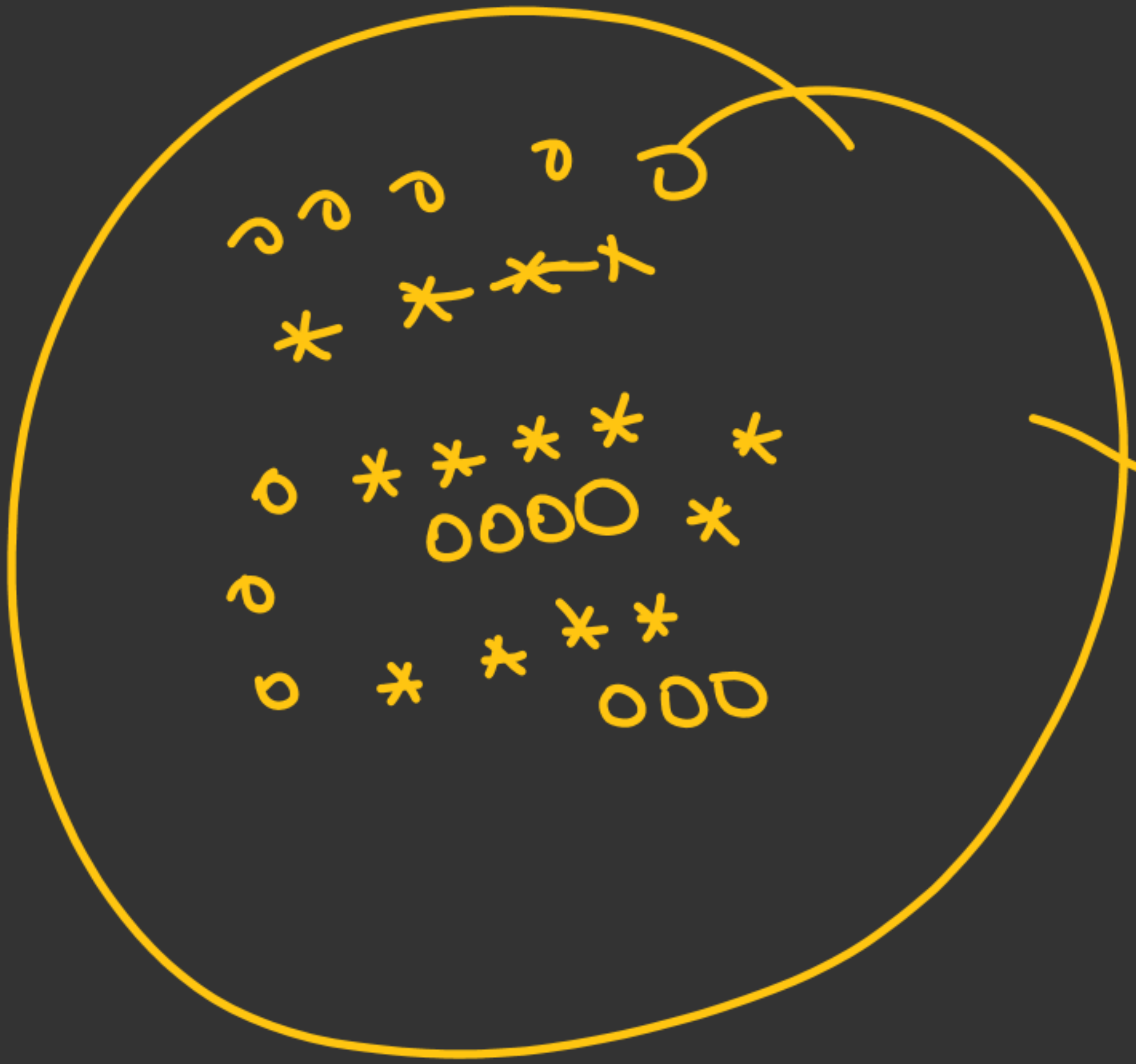
Logistic Reg.



- $x_i\beta > 0$, $Y > .5 \Rightarrow$ class1

  $x_i\beta < 0$, $Y < .5 \Rightarrow$ class0

  Cost fxn ----



not in Syllabus

Logistic Regression

Logistic Regression

Let us have a data with some classes 1 and 0, these are the Y values of the input.
In logistic Regression we actually try to fit a S curve on the data.

done

The Sigmoid Function…

Logistic Regression

Logistic Regression

Now we have the concept of the threshold, how to find the best coefficients ?

- generally threshold = .5
- But it can be changed.

$$Y = \frac{1}{1+e^{-x_i\beta}} \begin{cases} > .5 \Rightarrow \text{class 1} \\ < .5 \Rightarrow \text{class 0} \end{cases}$$

we have to find β's

• Logistic Regression

The concept of threshold

done

What is logit or log of odds

- Odds $\Rightarrow \dfrac{\text{Probab of Success}}{\text{Probab of failure}}$

- In our Case Odds $= \dfrac{P_1}{P_0}$

$$= \frac{P}{1-P}$$

- Probab of win $\Rightarrow .8$
- Probab of loosing $\Rightarrow .2$

what is odds against wining $\Rightarrow \underbrace{P_{loosing}}/P_{win}$

$\Rightarrow \underbrace{.2/.8}$

Odds of Winning $\Rightarrow P_{win}/P_{loose}$

Logistic Regression

What is Logit ?

$$\text{log}_e \text{odds}$$

$$\left\{\log_e \frac{P}{1-P}\right\} \Rightarrow \left\{\log_e \frac{1/1+e^{-x_i\beta}}{1-\frac{1}{1+e^{-x_i\beta}}}\right\}$$

$$\Rightarrow \log_e \frac{1}{1+e^{-x_i\beta}-1}$$

$$\Rightarrow \log_e \frac{1}{e^{-x_i\beta}} \Rightarrow \log_e e^{+x_i\beta} \Rightarrow x_i\beta$$

$$\text{Odds of wining} = P_{win}/P_{loose}$$

$$\text{Odds of loosing} = \frac{P_{loose}}{P_{win}}$$

• Logistic Regression

In logistic Reg

$$\log_e odds \Rightarrow x_i\beta$$
$$\Rightarrow (\beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 \ldots)$$

Importance of Logit ??

• Logistic Regression

• Logistic Regression

The pattern of Questions on this ...

Linear Classification

40/70 ⇒ Success | 30/40 ⇒ fail.

Logistic Regression

7) Consider the data collected from 410 customers in a restaurant. It is observed that 40 of the 70 customers tipped the server who was wearing a black shirt and 130 of the 340 customers tipped the server who was wearing a different color. Compute the logit or log-odds of tipping a server wearing a black shirt.

- 0.2877
- 0.1249
- -0.7677
- -1.7677

410 Customers ⇒ 40 out of 70 Customers tip server with black shirt
130 " " 340 " " " " white shirt

loge odds of tipping the server wearing black shirt.

$$\Downarrow$$

$$\log_e\left[\frac{40/70}{30/70}\right] \Rightarrow \log_e(4/3) \Rightarrow .287.$$

Q 1D data,

if $x = 1$ $\log_e$ odd $\Rightarrow .45$

if $x = 5$ $\log_e$ odd $\Rightarrow .75$

find β_1, β_0 for logistic Reg $\Rightarrow \dfrac{1}{1+e^{-x_i\beta}}$

$\hookrightarrow (\beta_0 + \beta_1 x_i^1)$

V. Imp $\rightarrow$ in logistic Reg.

$\log \text{odd} = \beta_1 x^1 + \beta_0$

$.45 = \beta_1 \times 1 + \beta_0$

$.75 = 5\beta_1 + \beta_0$

$\beta_1 = 3/40, \beta_0 = 3/8.$

Linear Classification

Logistic Regression

8) In continuation with question 7, let $x = 1$ if the server is wearing black shirt and $x = 0$ for servers wearing other colored shirts. We know that there are 270 observations with $x = 1$ and 340 observations with $x = 0$. The response variable is also an indicator variable given by $y = 1$ if the customer left a tip and $y = 0$ if the customer did not leave a tip. Use this data to fit a logistic regression model to compute the log-odds of leaving a tip depending on the color of the server's shirt..

H.W

- $-0.4797 + 0.1249x$
- $0.2877 + 0.1249x$
- $0.1249 + 0.4317x$
- $-0.4797 + 0.7674x$

Logistic Regression

What type of dependent variable is suitable for logistic regression?

A) Continuous variable
B) Categorical variable with multiple categories
C) Binary or dichotomous variable
D) Ordinal variable

Logistic Regression

In logistic regression, what is the role of the logistic function (sigmoid function)?

A) It transforms the independent variables.
B) It models the relationship between the dependent and independent variables.
C) It converts the log-odds into probabilities.
D) It calculates the likelihood of the data.

Logistic Regression

Which term represents the natural logarithm of the odds of an event occurring in logistic regression?

A) Odds ratio
B) Probability
C) Log-odds or logit
D) Coefficient

Logistic Regression

What is the likelihood function used for in logistic regression?

A) To estimate the coefficients of the model.
B) To calculate the odds ratio.
C) To find the best threshold for classification.
D) To assess the fit of the model by maximizing the likelihood of the observed outcomes.

Logistic Regression

1. What kind of algorithm is logistic regression?

a) Cost function minimization

b) Ranking

c) Regression

d) Classification

Logistic Regression

6. Probability of an event occurring is 0.9. What is odds ratio?

a) 0.9:1

b) 9:1

c) 1:9

d) 1:0.9

#Q. The following table gives the binary labels $(y^{(i)})$ for four points $\left(x_1^{(i)}, x_2^{(i)}\right)$ where $i = 1, 2, 3, 4$. Among the given options, which set of parameter values β_0, β_1, β_2 of a standard logistic regression model $p(x_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 x+\beta_2 x)}}$ results in the highest likelihood for this data?

(a) $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = 2.0$

(b) $\beta_0 = -0.5, \beta_1 = -1.0, \beta_2 = 2.0$

(c) $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = -2.0$

(d) $\beta_0 = -0.5, \beta_1 = 1.0, \beta_2 = 2.0$

x_1	x_2	y
0.4	–0.2	1
0.6	–0.5	1
–0.3	0.8	0
–0.7	0.5	0

Logistic Regression

- The Loss function

Linear Classification

Logistic Regression

- **The Loss function**

How can we use log into this function

Linear Classification

Logistic Regression

- Extending the case for more than 2 classes… (not imp)

2 mins Summary

Topic

Topic

Topic

Topic

Topic

THANK - YOU