# Data Science and Artificial Intelligence

# Machine Learning

**Classification**

**Lecture No. 2**


By- SIDDHARTH SABHARWAL SIR

# Recap of Previous Lecture

- **Topic** Ridge Reg
- **Topic** Lasso
- **Topic** Linear classification
- **Topic**
- **Topic**

# Topics to be Covered

Topic

Topic

Topic

Topic

Topic

Linear classification

$\downarrow$

Logistic Regression

Morning Routine
↓
Best-day

REVISION

# Linear Classification : Indicator Matrix

Concept ⇒

we create Y's ⇒ No of classes

2 Classes $Y_1$ $Y_2$

~~3 Classes $Y_1$ $Y_2$ $Y_3$~~

~~4 Classes $Y_1$ $Y_2$ $Y_3$ $Y_4$~~

Indicator matrix

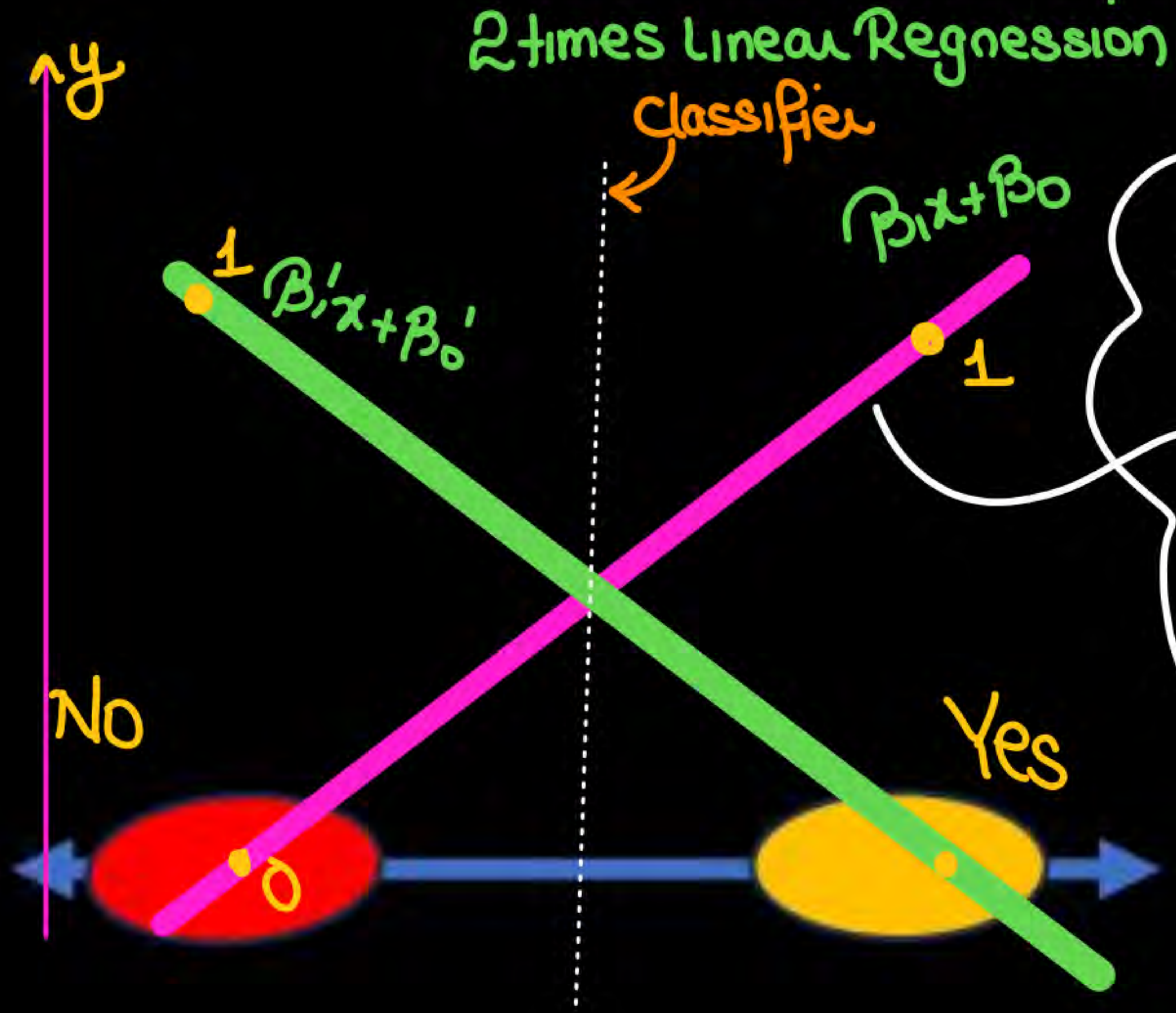
for a data point only one Y is '1'

## Linear Classification : Why LR cannot be used here ?

then we do linear regression for each Y and

in 2 class case

# Linear Classification




- Why we are not using a single line → LR
  - → we need a classifier that Bifurcate the points of two classes
  - → in LR line the value of Y has no limit.

For any new point → Rule

$\beta_1 x + \beta_0 > \beta_1' x + \beta_0'$ → Class 1

$\beta_1' x + \beta_0' > \beta_1 x + \beta_0$ → Class 0

## Linear Regression of an Indicator Matrix

Not imp

**So, now the analysis is as follows :**

So for more than 2 classes ⇒ K number of classes in label of data

↳ one hot coding $Y_1\ Y_2\ ---\ Y_K$ ⇒ indicator matrix

↳ Now linear regression for all Y's ⇒ we will get K linear eq ⇒ 1D case ⇒

$\beta_1 x + \beta_0 \rightarrow Y_1$

$\beta_1' x + \beta_0' \rightarrow Y_2$

$\beta_1'' x + \beta_0'' \rightarrow Y_3$

$\vdots$

## Linear Regression of an Indicator Matrix

Method: only for know.

### Lets extend the case for K classes

for any new point → decision Rule

Find $\beta_1 x + \beta_0$
$\beta_1' x + \beta_0'$
$\beta_1'' x + \beta_0''$
$\vdots$

So we find line eq that give max value → this give the class of new data point

* Here we are using Regression for Classification

## Linear Classification

### Linear Regression of an Indicator Matrix

One hot Coding

Indicator mat.

**Lets extend the case for K classes**

$Y \rightsquigarrow$ 3classes

| data | Y |
|---|---|
| 1 | Yes |
| 2 | Yes |
| 3 | No |
| 4 | No |
| 5 | Can't Say |
| 6 | Can't Say |
| 7 | No |
| 8 | Yes |

$\longrightarrow$

| data | $Y_1$ (Yes) | $Y_2$ (No) | $Y_3$ (Can't say) |
|---|---|---|---|
| 1 | 1 | 0 | 0 |
| 2 | 1 | 0 | 0 |
| 3 | 0 | 1 | 0 |
| 4 | 0 | 1 | 0 |
| 5 | 0 | 0 | 1 |
| 6 | 0 | 0 | 1 |
| 7 | 0 | 1 | 0 |
| 8 | 1 | 0 | 0 |

# Linear Classification

## Linear Regression of an Indicator Matrix

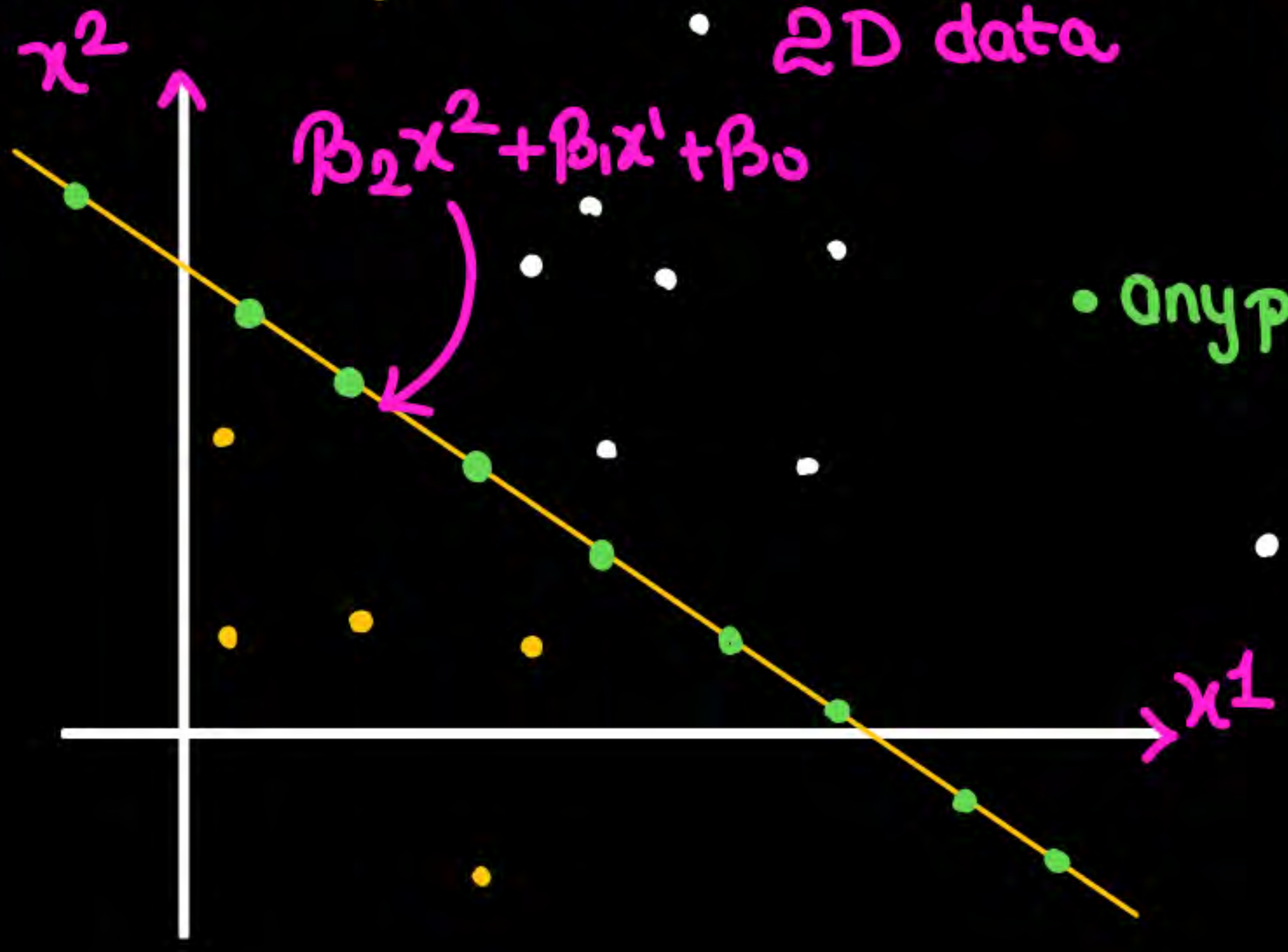
- So the previous algorithm
- But now lets see the linear Classification →
  - → in previous method for a 2 class case we need 2 linear Reg, and then by analysis we do classification

Lets consider a 2 class problem... We can have a single classifier for a 2 class problem...

# Linear Classification

## Linear Regression of an Indicator Matrix

But we can find the Classifier directly ⇒




**The loss function for a 2 class case...**

- any point on the line

  $\beta_2 x^2 + \beta_1 x^1 + \beta_0 = 0$

- any point above the line then $\beta_2 x^2 + \beta_1 x^1 + \beta_0 > 0$
- any " below " " then $\beta_2 x^2 + \beta_1 x^1 + \beta_0 < 0$

# Linear Classification

## Linear Regression of an Indicator Matrix

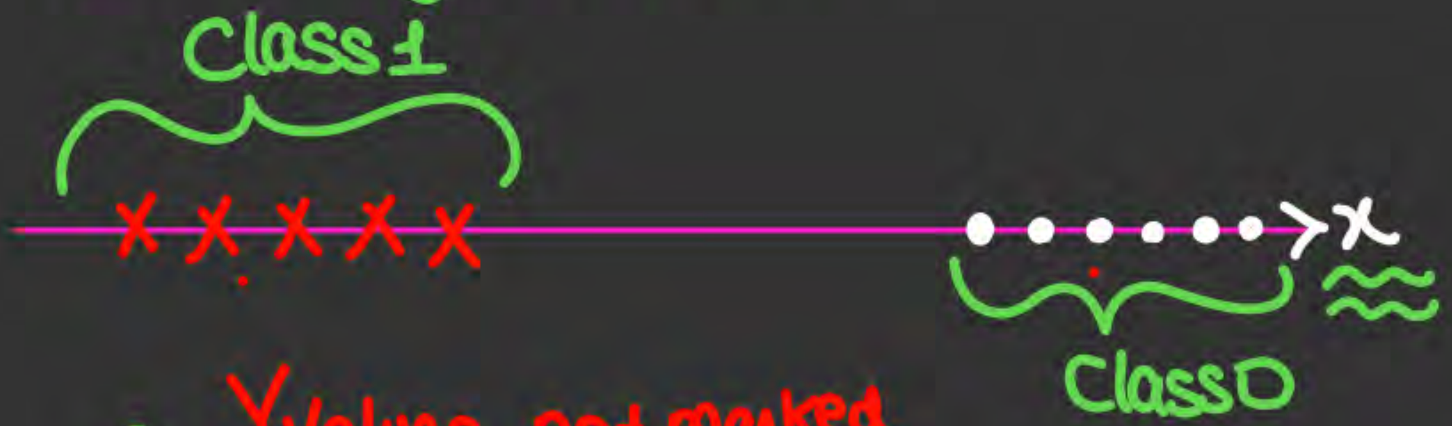
lets take example of 2D data points


| $x^1$ | $x^2$ | Y |
|---|---|---|
| | | Yes<br>No |

But this loss function has 2 problems 1. outlier and 2. value of predicted Y

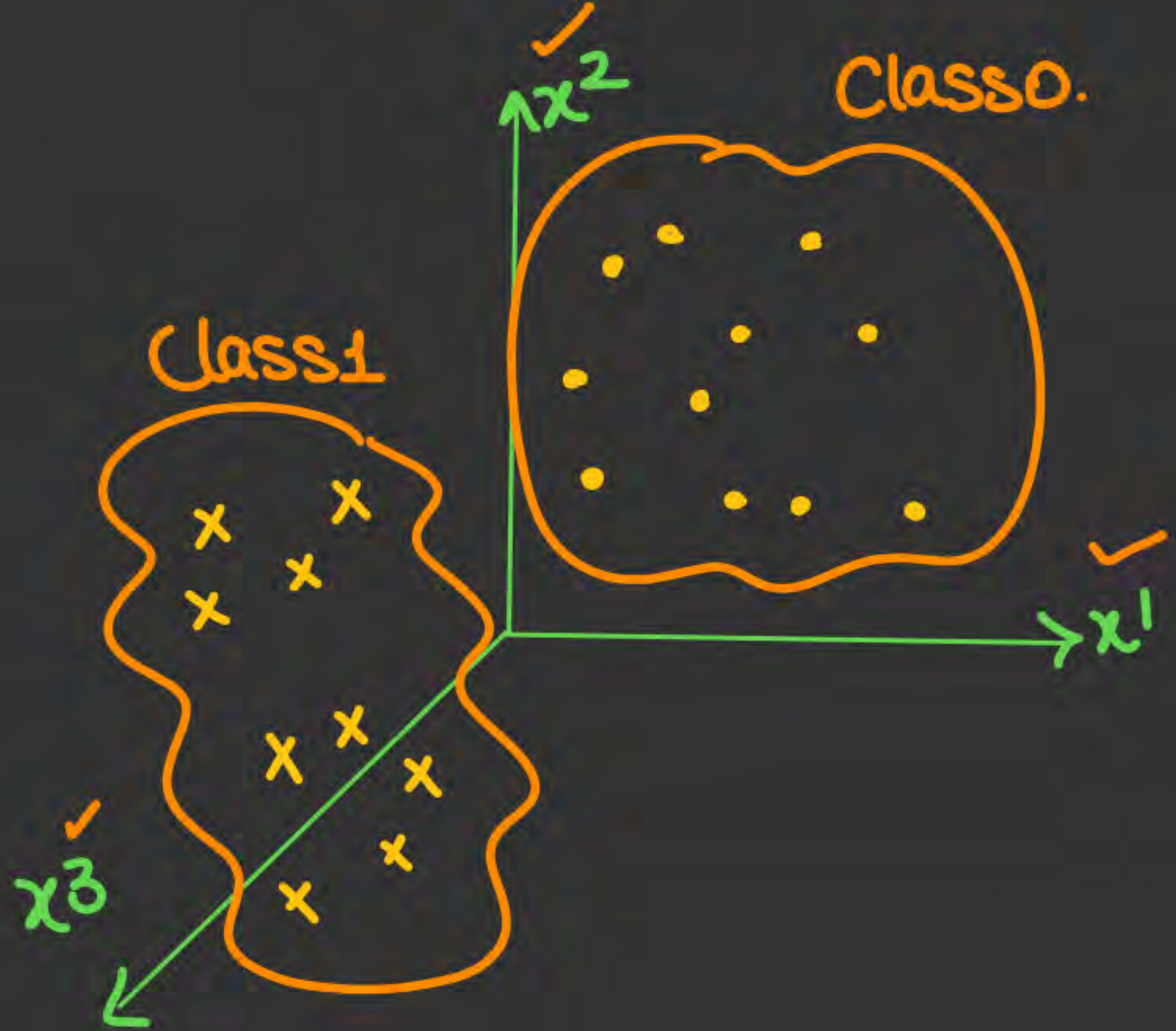
In case of classification we donot need to show y values to represent data because we can draw the points of different classes using different markings.

example of 1D data




- Y values not marked

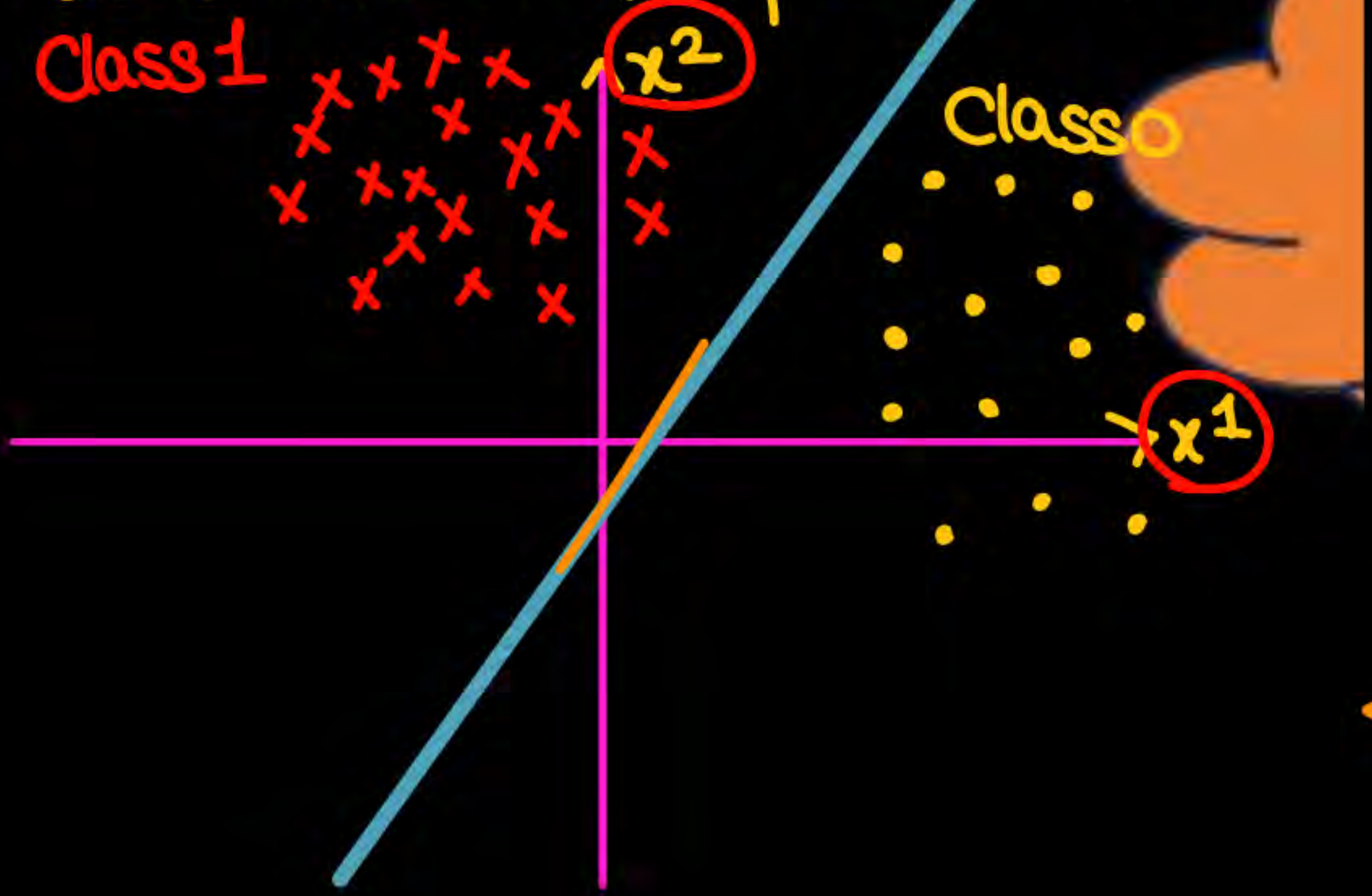
3D data



# Linear Classification

## Linear Regression of a

Lets take example of 2D data points




- So now we need to find the classifier
  - ↳ $B_2x^2 + B_1x^1 + B_0$

⇒ we can see that points of class 1 in the eq of classifier → (+ve)

⇒ also class 0 points in classifier eq ↳ (-ve)

⇒ Now Y label of Class 0 → -1

Y label of class 1 → 1

The best classifier will be where all the class 0 points are below the classifier and all the class 1 points are above the classifier ⇒ So to find the best classifier ⇒

we get classifier line such that class 1 oopar class 0 neeche

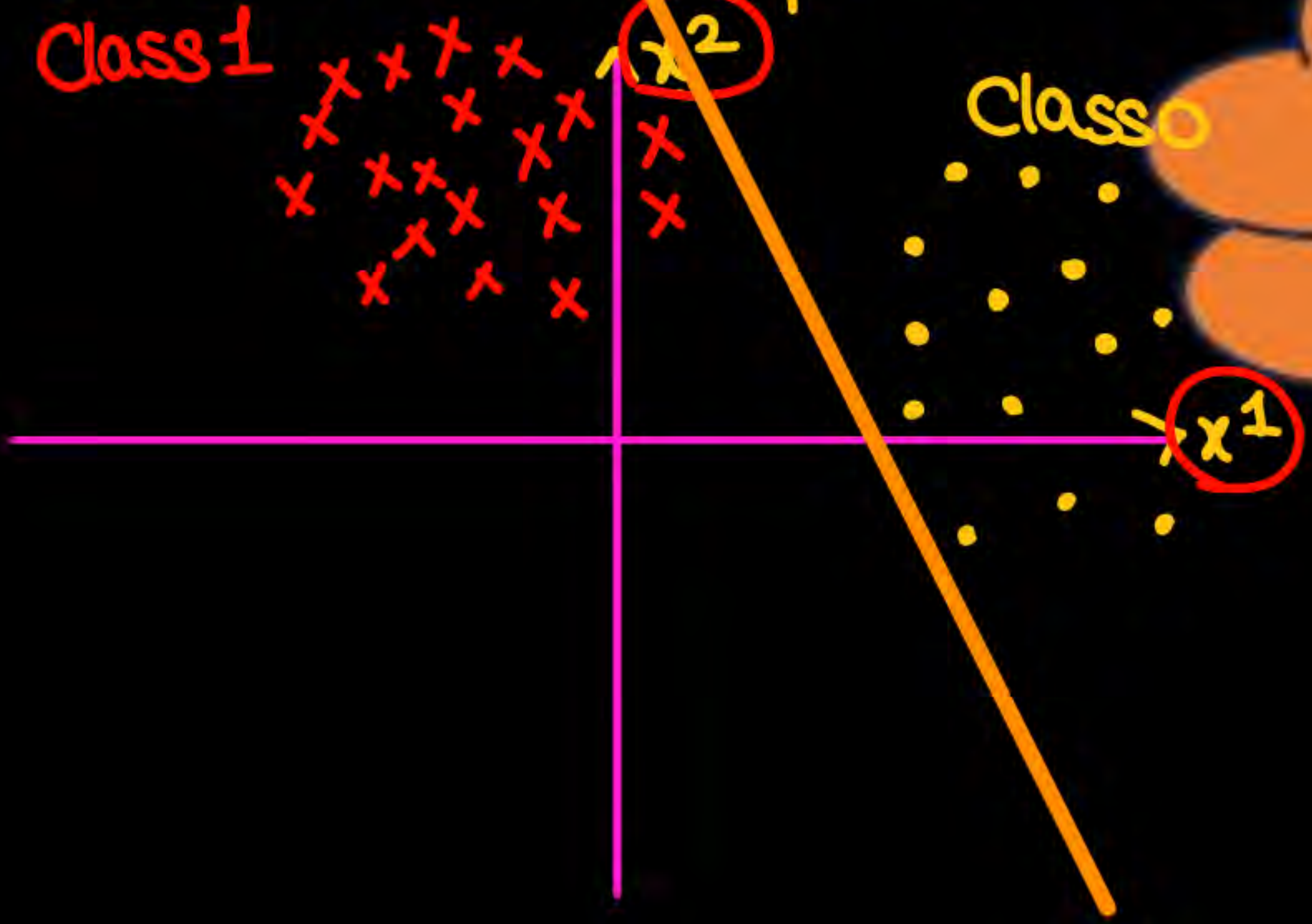
So for best classifier ⇒ $y_i (\beta_2 x^2 + \beta_1 x^1 + \beta_0)$

→ Best classifier will be such that where this product is +ve for all points.

# Linear Classification

## Linear Regression of a

lets take example of 2D data points




Class 0 → $Y = +1$

Class 1 → $Y = -1$

$y_i(\beta_2 x^2 + \beta_1 x^1 + \beta_0) = +ve$ for all points

So class 0 → $\beta_2 x^2 + \beta_1 x^1 + \beta_0$ shd be $+ve$

When data has 2 classes

Now it is the freedom

* $\rightsquigarrow$ $Y=+1$

• $\rightsquigarrow$ $Y=-1$

Rule $\underbrace{y_i\left(\beta_2 x^2+\beta_1 x^1+\beta_0\right)}_{\text{+ve for all points}}$

So for Pink point $\beta_2 x^2+\beta_1 x^1+\beta_0>0$

" " green " $\beta_2 x^2+\beta_1 x^1+\beta_0<0$

When dat a has 2 classes

Now it is the freedom

$* \rightsquigarrow Y = -1$

$\bullet \rightsquigarrow Y = +1$

Rule $\underbrace{y_i(\beta_2 x^2 + \beta_1 x^1 + \beta_0)}_{\text{+ve for all points}}$

So for Pink point $\beta_2 x^2 + \beta_1 x^1 + \beta_0 < 0$

" " green " $\beta_2 x^2 + \beta_1 x^1 + \beta_0 > 0$

So linear classification ⇒

If datapoint

✓ $x_i = [1 \; x_i^1 \; x_i^2 \; \ldots \; x_i^D]$

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

For a 2 class problem

→ label the classes as ±1

→ The Best classifier

$y_i(\underbrace{\beta_0 + \beta_1 x_i^1 + \ldots \beta_D x_i^D}_{\text{shd be +ve}})$

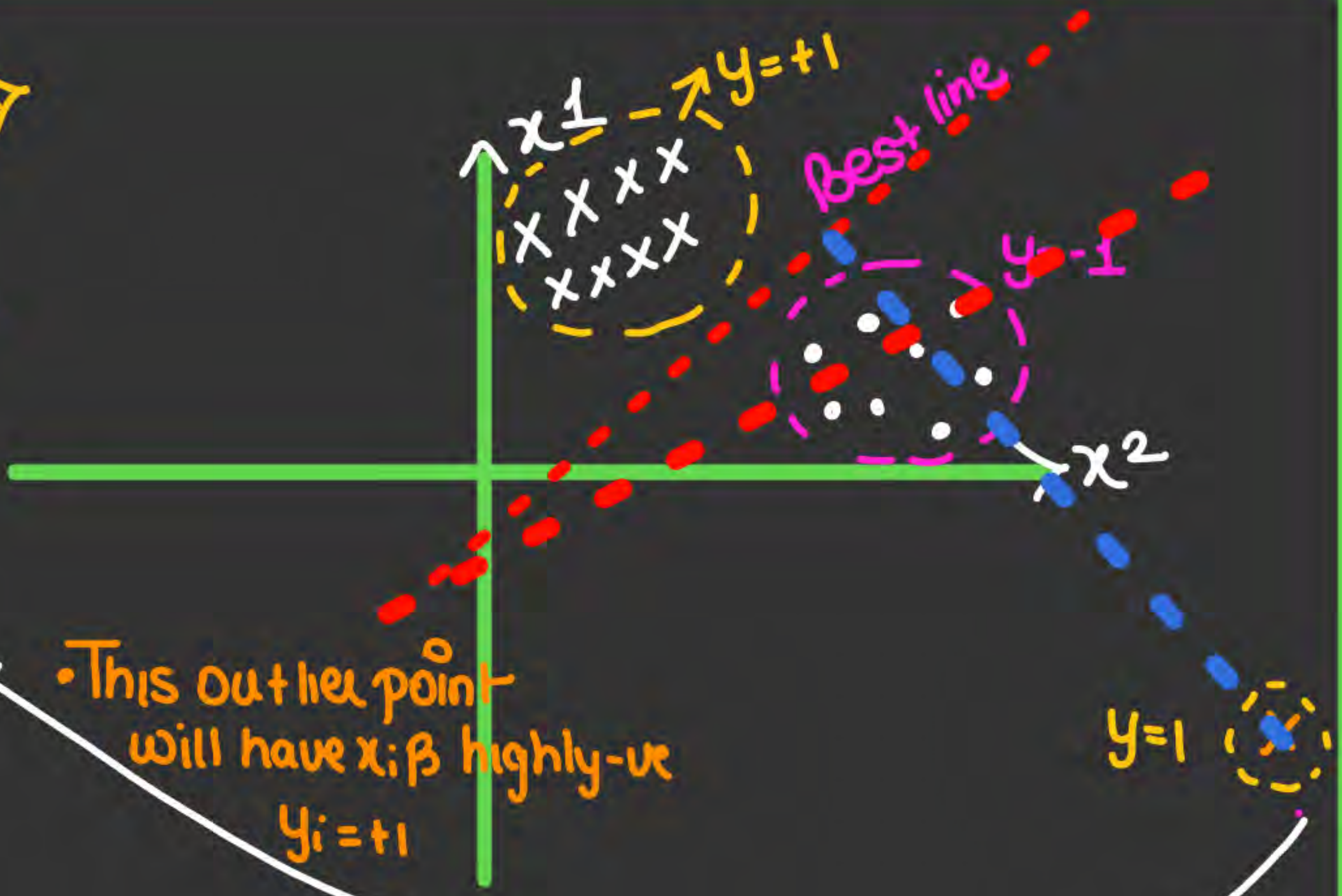
→ So Best classifier maximizes $\sum_{i=1}^{N} y_i (x_i \cdot \beta)$

## Linear classification ⇒

↳ Outlier ⇒

But

$$\Rightarrow \alpha = \max \sum_{i=1}^{N} y_i (x_i\beta)$$




• This outlier point will have $x_i\beta$ highly -ve

$y_i = +1$

## Linear Classification

**Problem of outliers**

→ The linear classifier is effected by the outlier, because outlier have very high -ve value of $y_i(x_i\beta)$ thus the algorithm shift the whole classifier, to max $\sum y_i(x_i\beta)$

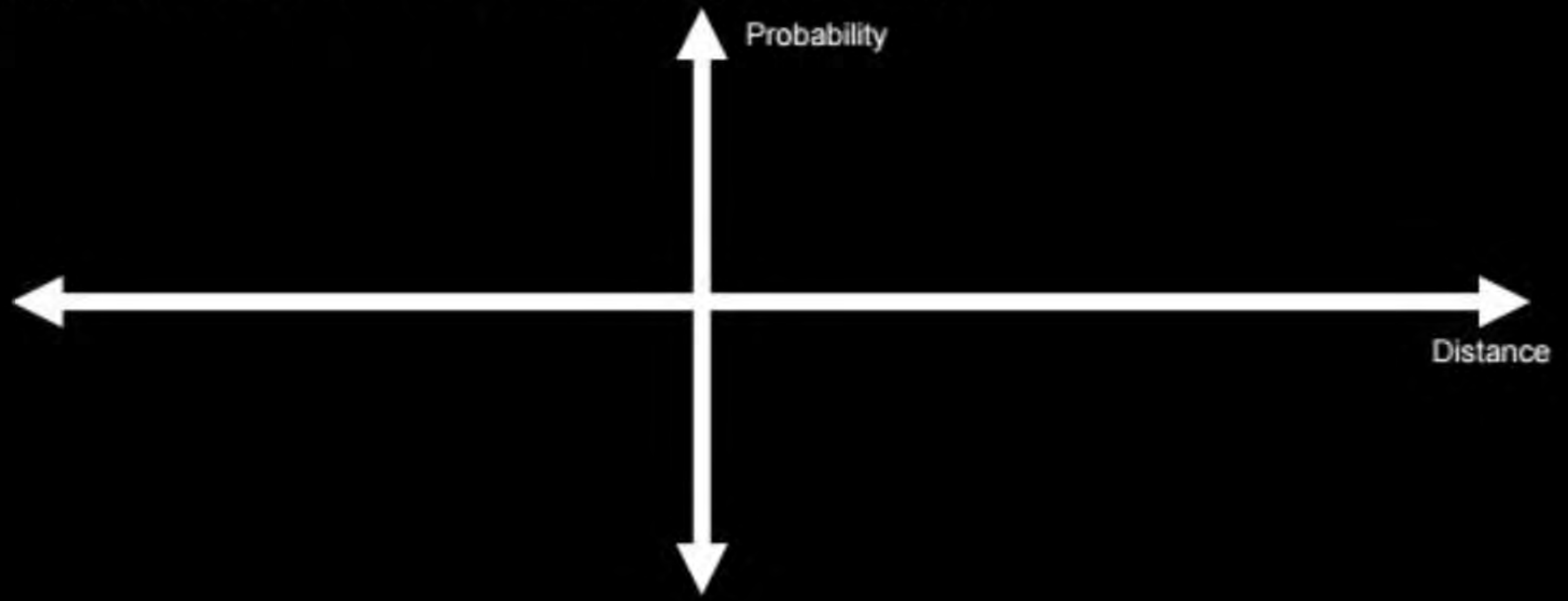
## • Linear Classification

To solve the problem of outlier we will not use the distance in the analysis rather we will use the probability.

# Linear Classification

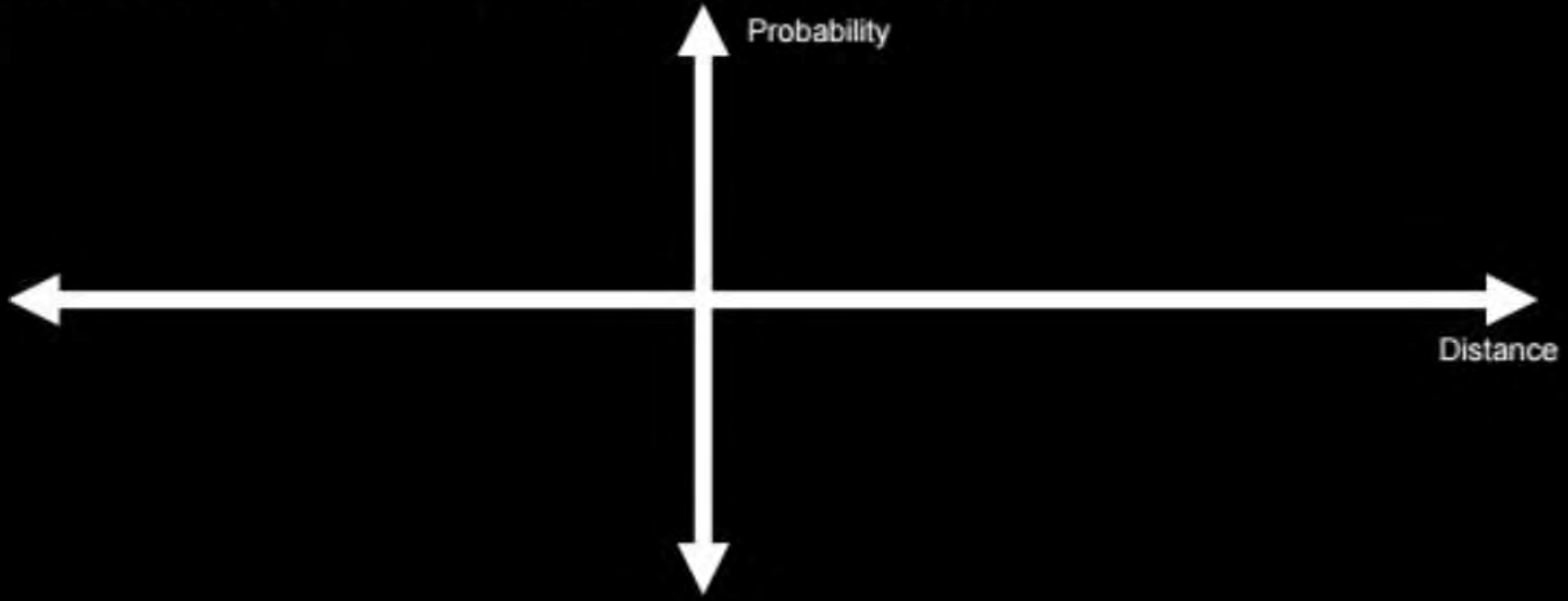
## Linear Classification

To solve the problem of outlier we will not use the distance in the analysis rather we will use the probability.



## • Linear Classification

To solve the problem of outlier we will not use the distance in the analysis rather we will use the probability.



## • Linear Classification

### • Linear Classification

Why linear regression was not good in case of classification problem – because Y was either 0/1 but the line that we learn was giving very large values also.

Hence in logistic regression we are doing regression but we are using sigmoid function here for perfect 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.

## • Logistic Regression

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

## • Logistic Regression

The concept of threshold

## • Logistic Regression

**Comparison of the linear classification and logistic Regression**

In linear classification we find a line and say value <>0
but here we say value <> some threshold

This is called sigmoid...

Also called Sigmoid Function

This is called sigmoid...

Also called Sigmoid Function

This is called sigmoid...

Also called
Sigmoid
Function

# Linear Classification

## Logistic Regression

- This can be used when the data is linearly seperable...

## Logistic Regression

- Logistic regression cannot solve XOR problem...

## What is Logit ??

# Linear Classification

## Logistic Regression

2 class case

The loss function...

Logistic Regression

2 class case

We find the parameters by rule ...

# Linear Classification

## Logistic Regression

2 class case

We find the parameters by rule ...

## Linear Classification

Logistic Regression

2 class case

We find the parameters by rule ...

## Linear Classification

**Logistic Regression**

2 class case

We find the parameters by rule ...

# Linear Classification

## Logistic Regression

2 class case

We find the parameters by rule ...

## Logistic Regression

Now calculation
Probability is easy...

## Logistic Regression

Simple decision rule in 2 class case

THANK - YOU