Data Science and Artificial Intelligence

Machine Learning

Classification

Lecture No. 1

By- SIDDHARTH SABHARWAL SIR

Recap of Previous Lecture

- Topic
- Topic
- Topic
- Topic
- Topic

Linear, Ridge, lasso

***.

WB + DPP + 80Q drive solved + 100Q more on drive + 50min lect. extra

Topics to be Covered

- Topic
- Topic
- Topic
- Topic
- Topic

linear classification

↓

Logistic regression

About the Faculty

- AIR 1 GATE 2021, 2023 (ECE).
- AIR 3 ESE 2015 ECE.
- M.Tech from IIT Delhi in VLSI.
- Published 2 papers in field of AI-ML.
- Paper 1 : Feature Selection through Minimization of the VC dimension.
- Paper 2 : Learning a hyperplane regressor through a tight bound on the VC dimension.

By- SIDDHARTH SABHARWAL SIR

GATE 2026 FODENGE

Join Our Telegram Channel!

Stay updated, get instant alerts & connect with the community

Join + Pin msg → drive link.

https://t.me/siddharthsirPW

By- SIDDHARTH SABHARWAL SIR


My Rule
* Moning
Productive
day
Productive
YOUR MORNING
SETS UP THE
SUCCESS
OF YOUR DAY
Fazil Azmaan

Linear Classification

Linear Classification :

What is classification ⇒ Categorical data ⇒ label is Categorical

So we can see that label has 2 class

Yes No

+1 -1

x^1	x^2	label
1	1	Yes
2	5	Yes
5	3	No
8	6	No

→ label is having Category

Linear Classification

Linear Classification :

What is classification ⇒ Categorical data ⇒ label is Categorical

So we can see that label has 2 class

Yes No

+1 -1

x^1	x^2	label
1	1	+1
2	5	+1
5	3	-1
8	6	-1

→ label is having Category

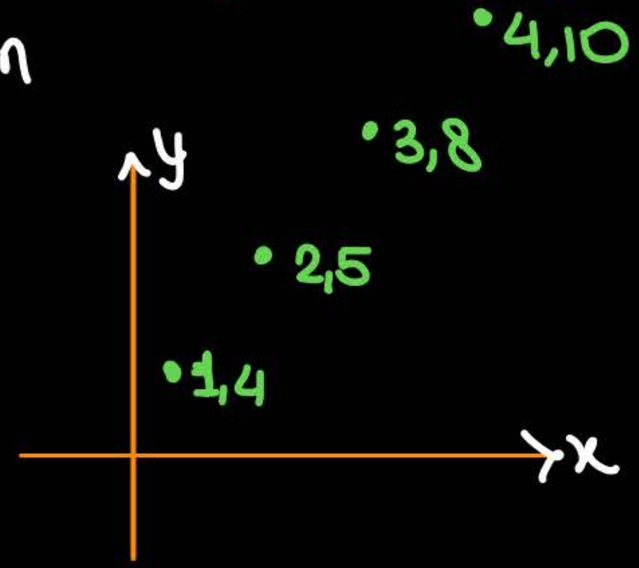
Linear Classification

Linear Classification :

In regression we mark point using x, y.

1D data Regression

x	y
1	4
2	5
3	8
4	10

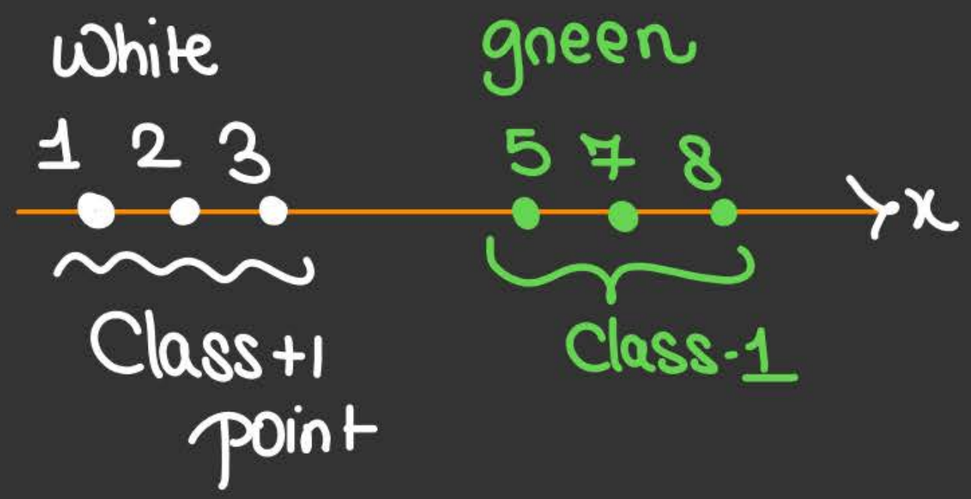


? To plot data we need y axis

1D, Categorical data

x	y
1	+1
2	+1
3	+1
5	-1
7	-1
8	-1

we donot need y axis to show data

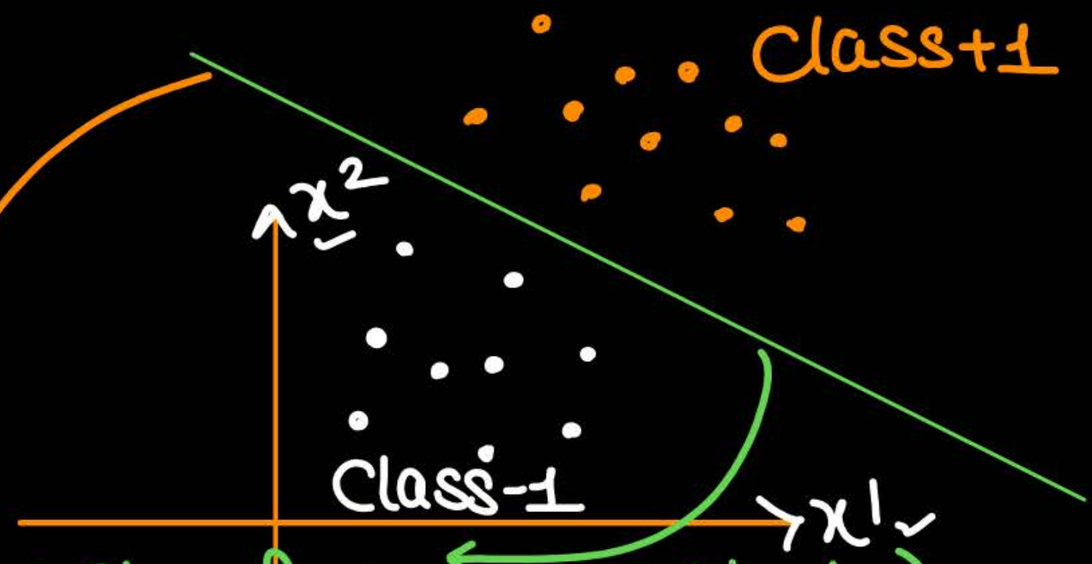


Linear Classification

Simple Line in a 2 D plane

2D data

x^1	x^2	y
1	2	-1
2	3	-1
4	2	-1
5	6	+1
7	8	+1

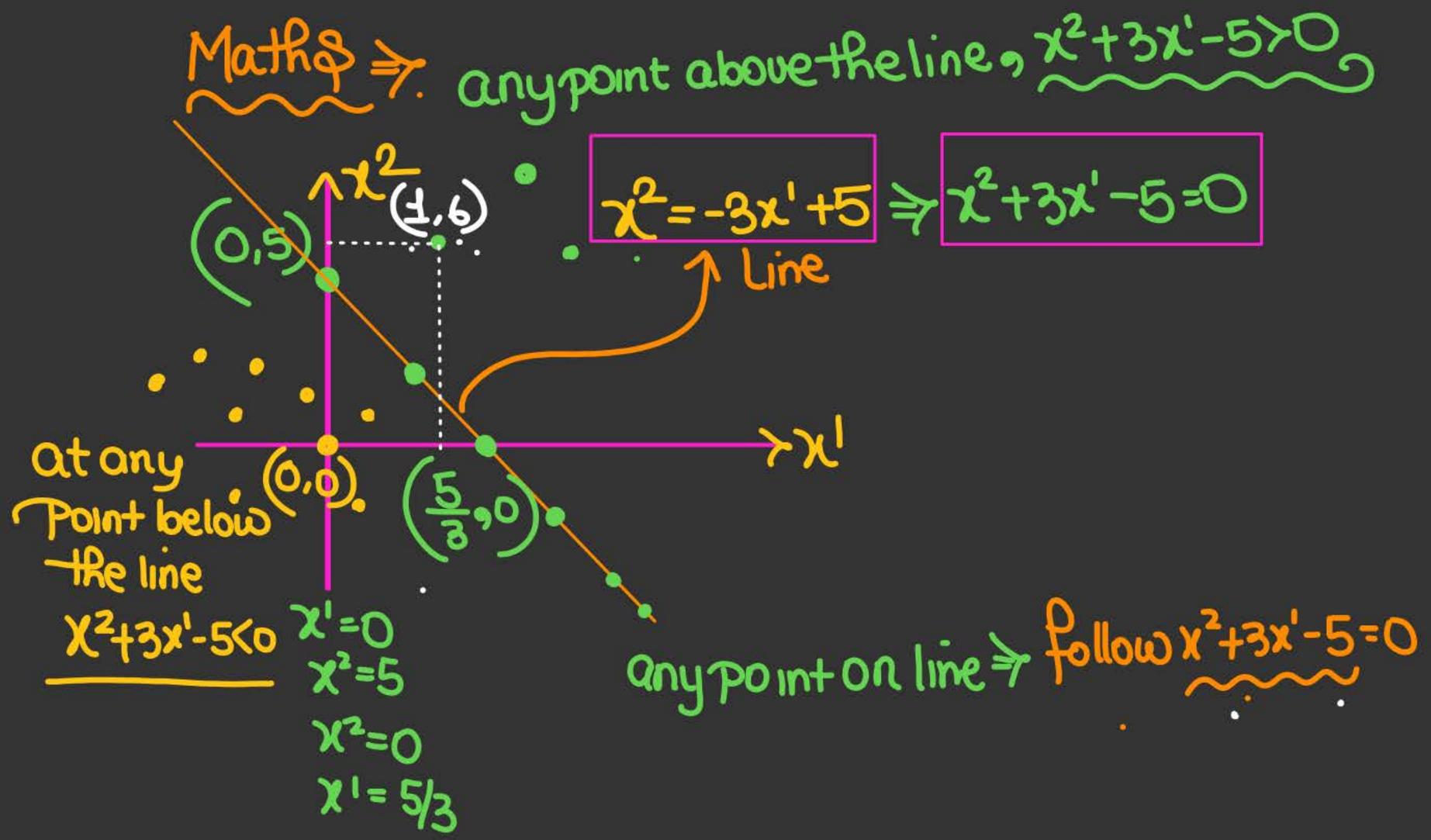


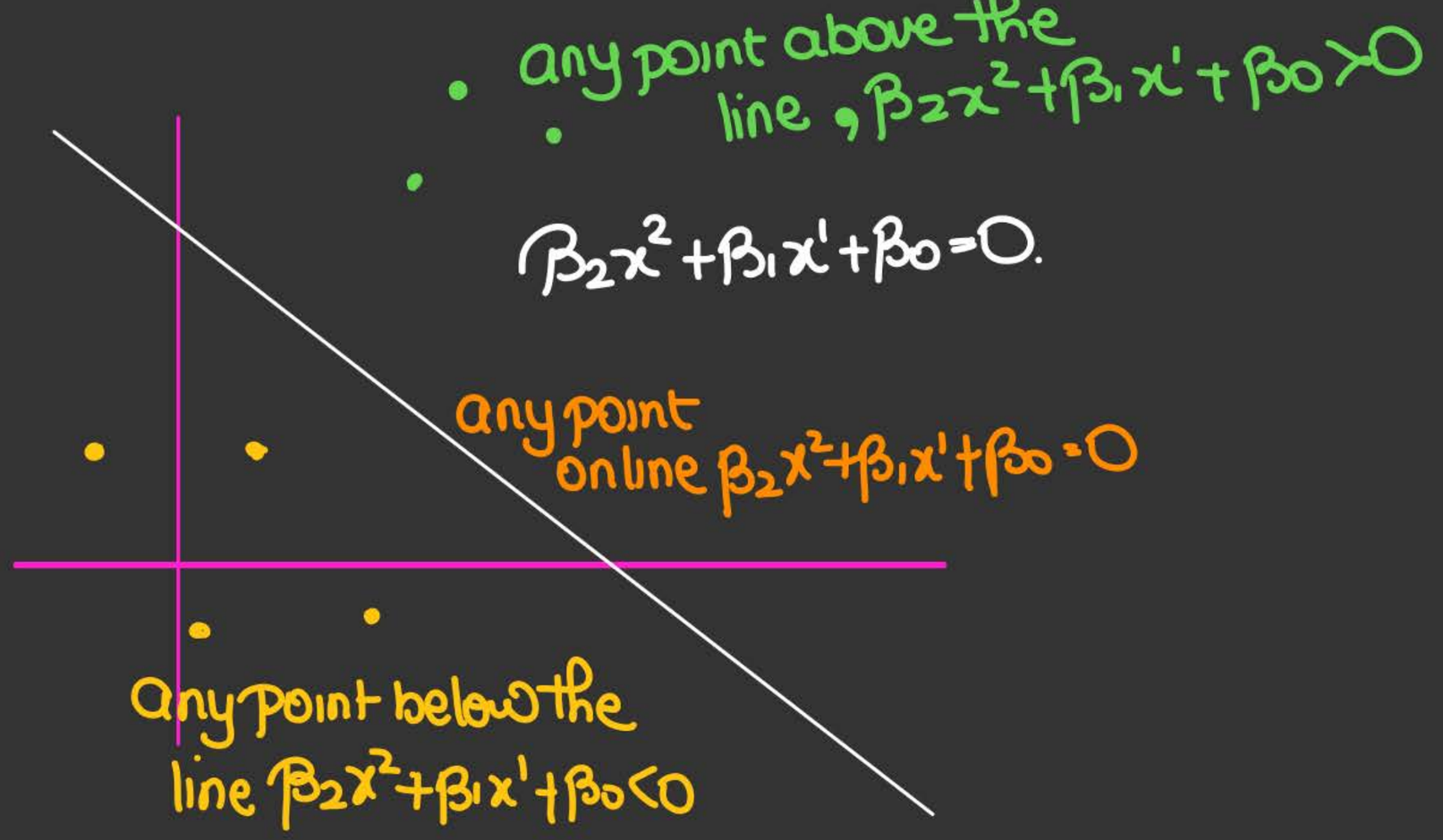
- Classifier is a straight line (2D) that seperate the two class data points

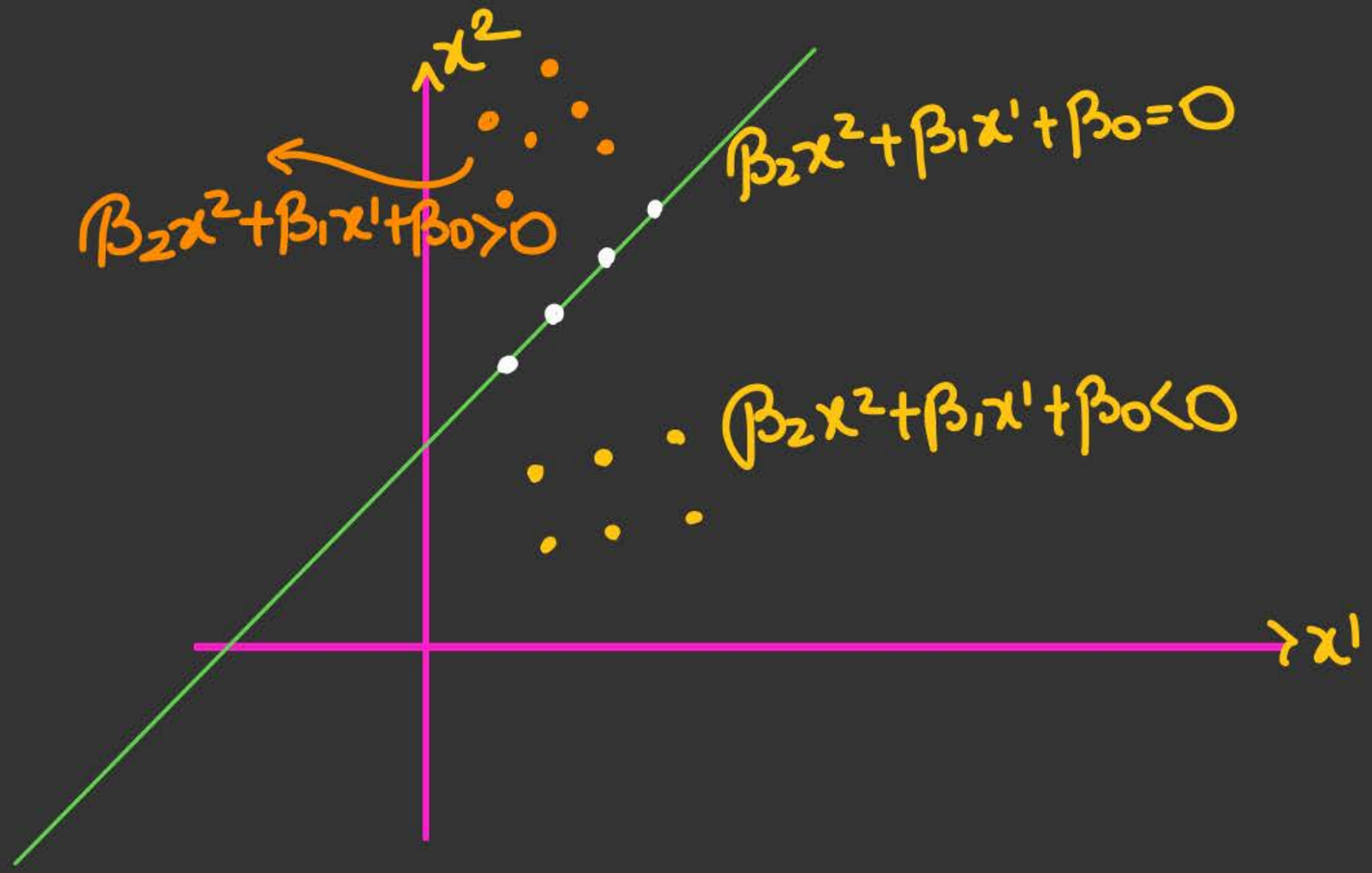
* Now we need a classifier

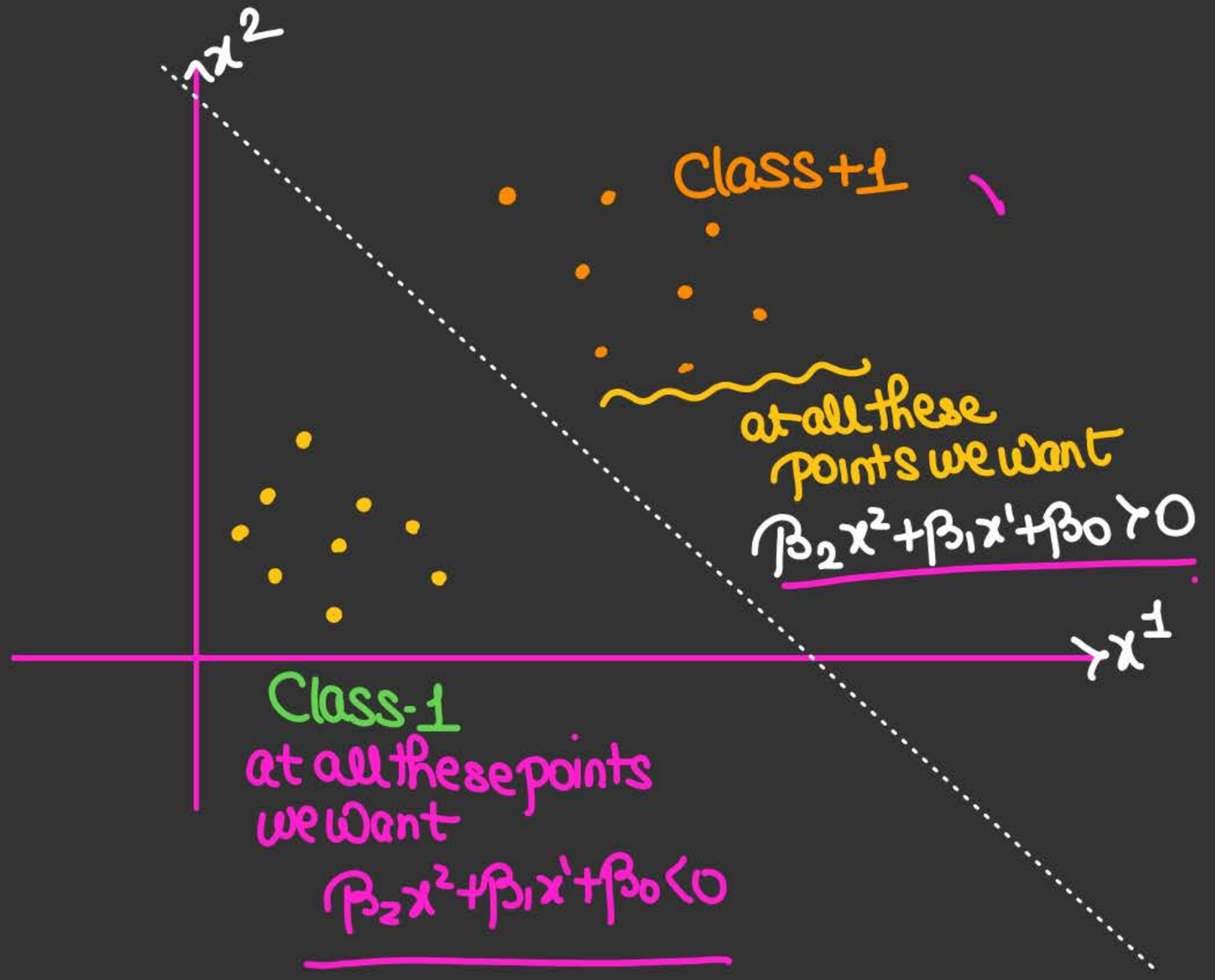
Equation of classifier

Equation of line ⇒ $\beta_1 x^1 + \beta_2 x^2 + \beta_0 = 0$


Maths ⇒ anypoint above the line, $x^2+3x'-5>0$
x^2
(1,6)
(0,5)
$x^2=-3x'+5$ ⇒ $x^2+3x'-5=0$
Line
x'
at any Point below the line
$x^2+3x'-5<0$
(0,0)
$\left(\frac{5}{3},0\right)$
$x'=0$
$x^2=5$
$x^2=0$
$x'=5/3$
anypoint on line ⇒ follow $x^2+3x'-5=0$


any point above the line, $\beta_2 x^2 + \beta_1 x' + \beta_0 > 0$
$\beta_2 x^2 + \beta_1 x' + \beta_0 = 0$.
any point on line $\beta_2 x^2 + \beta_1 x' + \beta_0 = 0$
any point below the line $\beta_2 x^2 + \beta_1 x' + \beta_0 < 0$


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

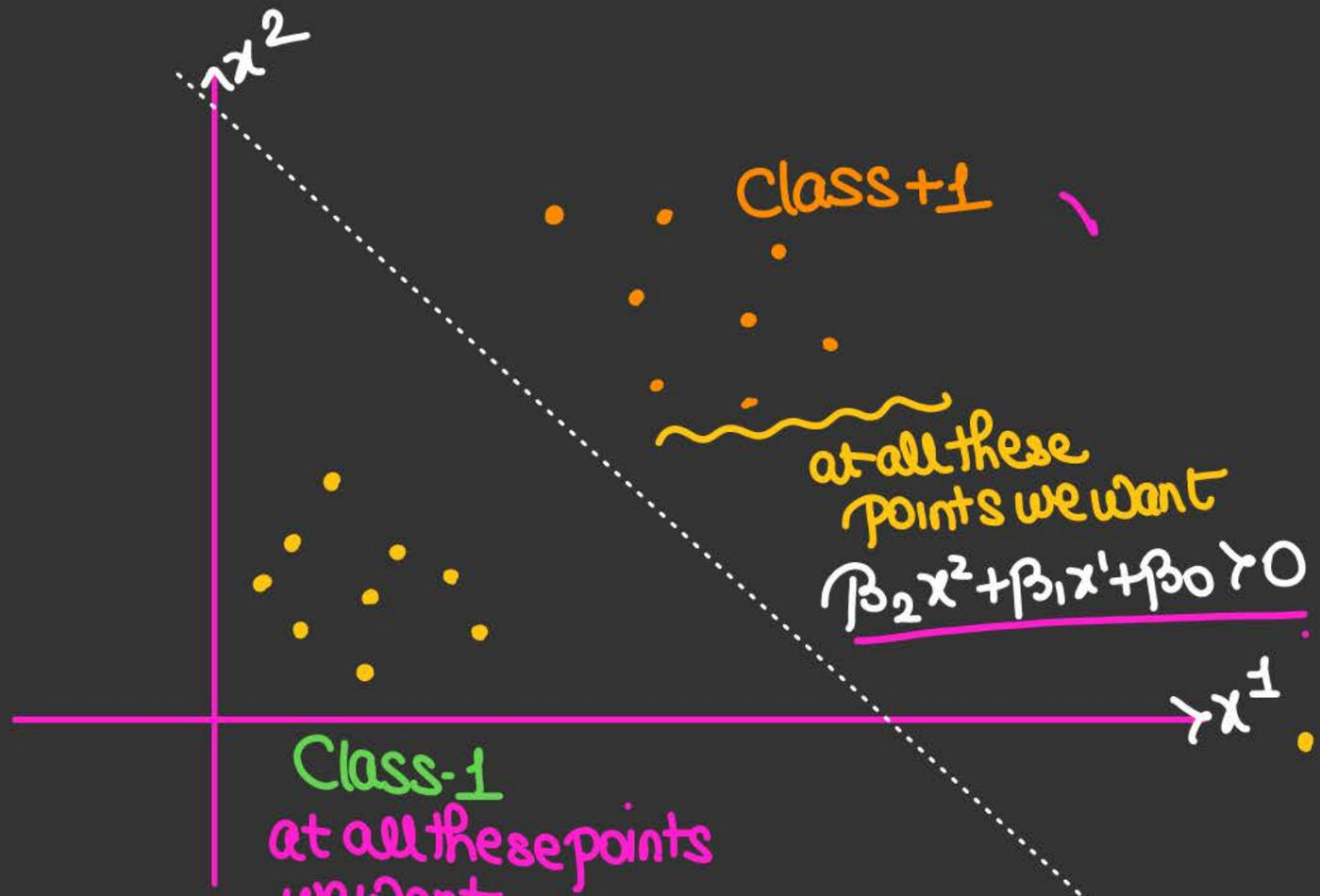


we want a line
$\beta_2 x^2 + \beta_1 x^1 + \beta_0$

Such that all

Class +1 points $\Rightarrow$ above the line

Class -1 points $\Rightarrow$ below the line



we want a line
$\beta_2 x^2 + \beta_1 x^1 + \beta_0$

Such that all
Class +1 points $\Rightarrow$ above the line
Class -1 points $\Rightarrow$ below the line

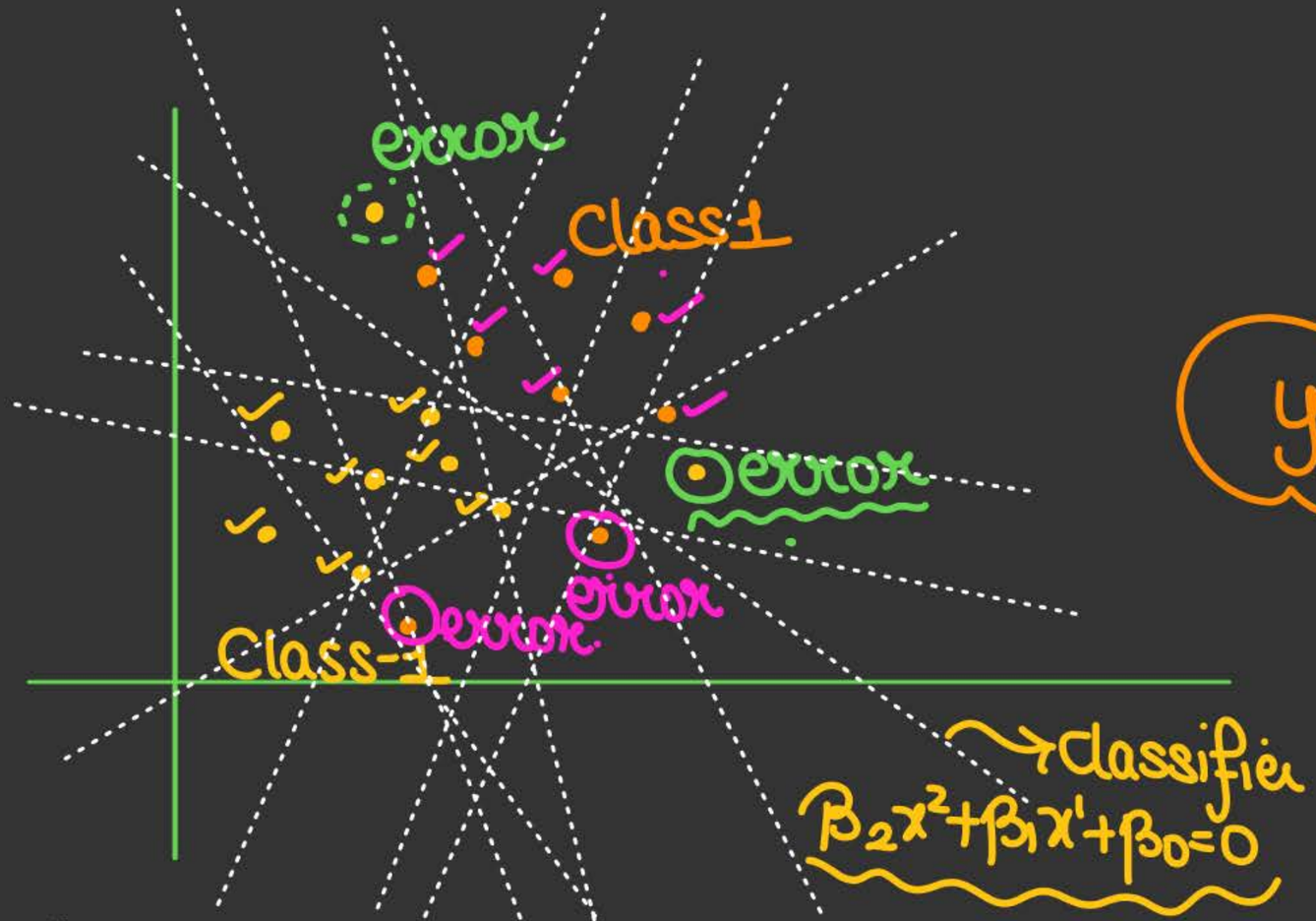
- So if y_i for a point +1, then point shd be above the line $\beta_2 x_i^2 + \beta_1 x_i^1 + \beta_0 > 0$
- if y_i for a point -1, then point shd be below the line $\beta_2 x_i^2 + \beta_1 x_i^1 + \beta_0 < 0$

So we want that Classifier shd be such that

for all datapoint $y_i(\beta_2 x_i^2 + \beta_1 x_i^1 + \beta_0) > 0$

So $y_i(x\beta) > 0$ $\forall$.

$x = [1 \; x^1 \; x^2 \; \dots]$ $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \end{bmatrix}$

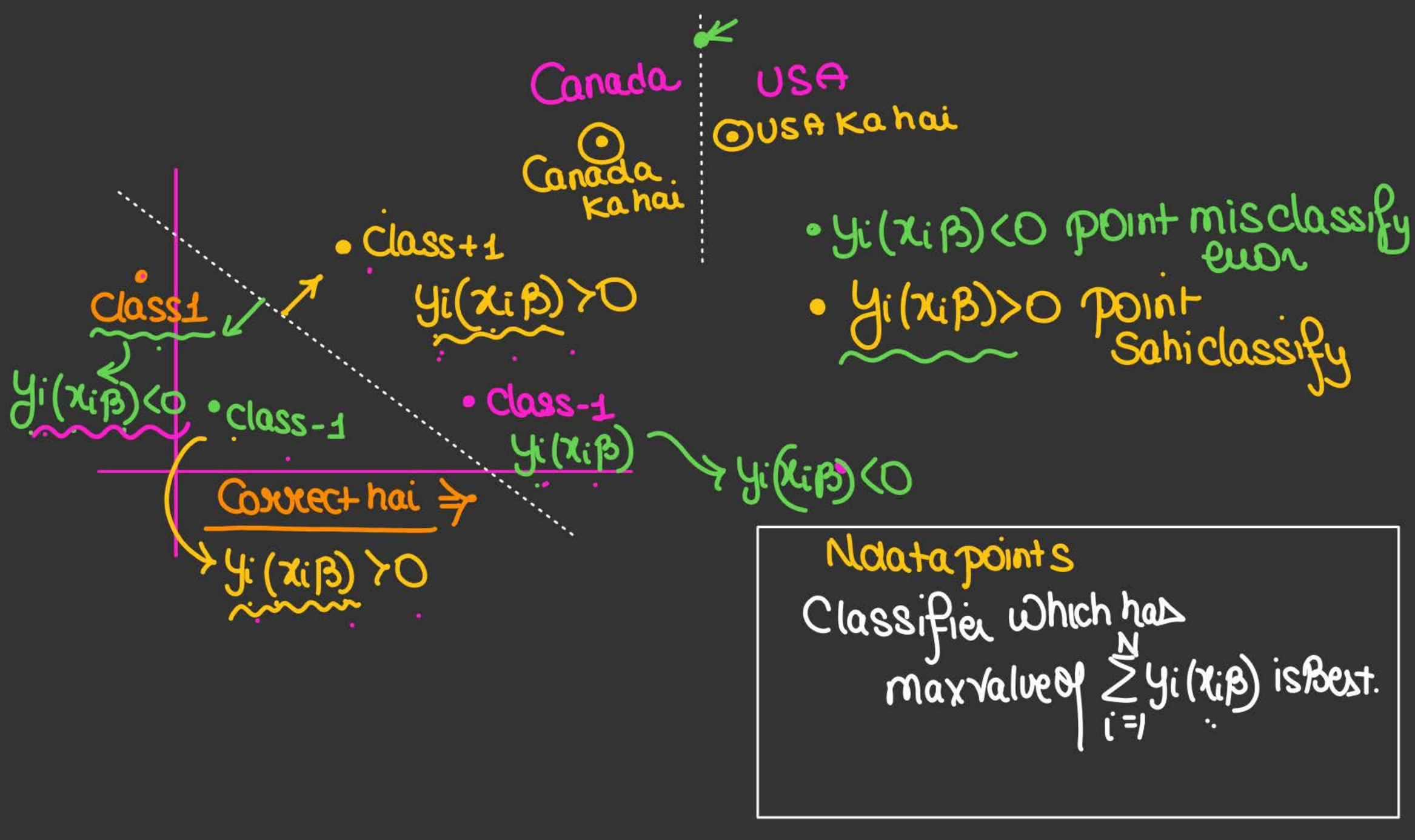


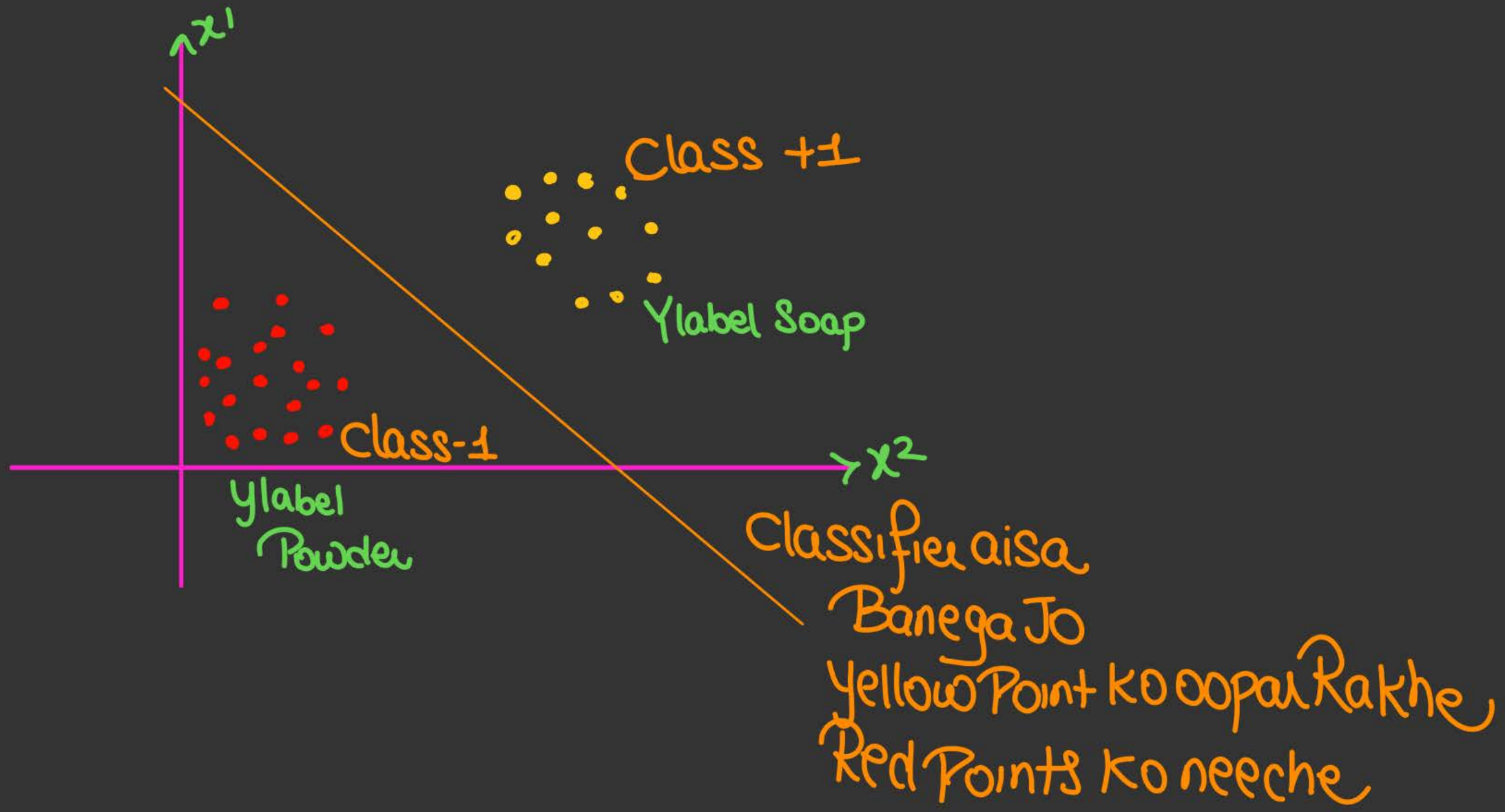
$y_i(x_i\beta) > 0$

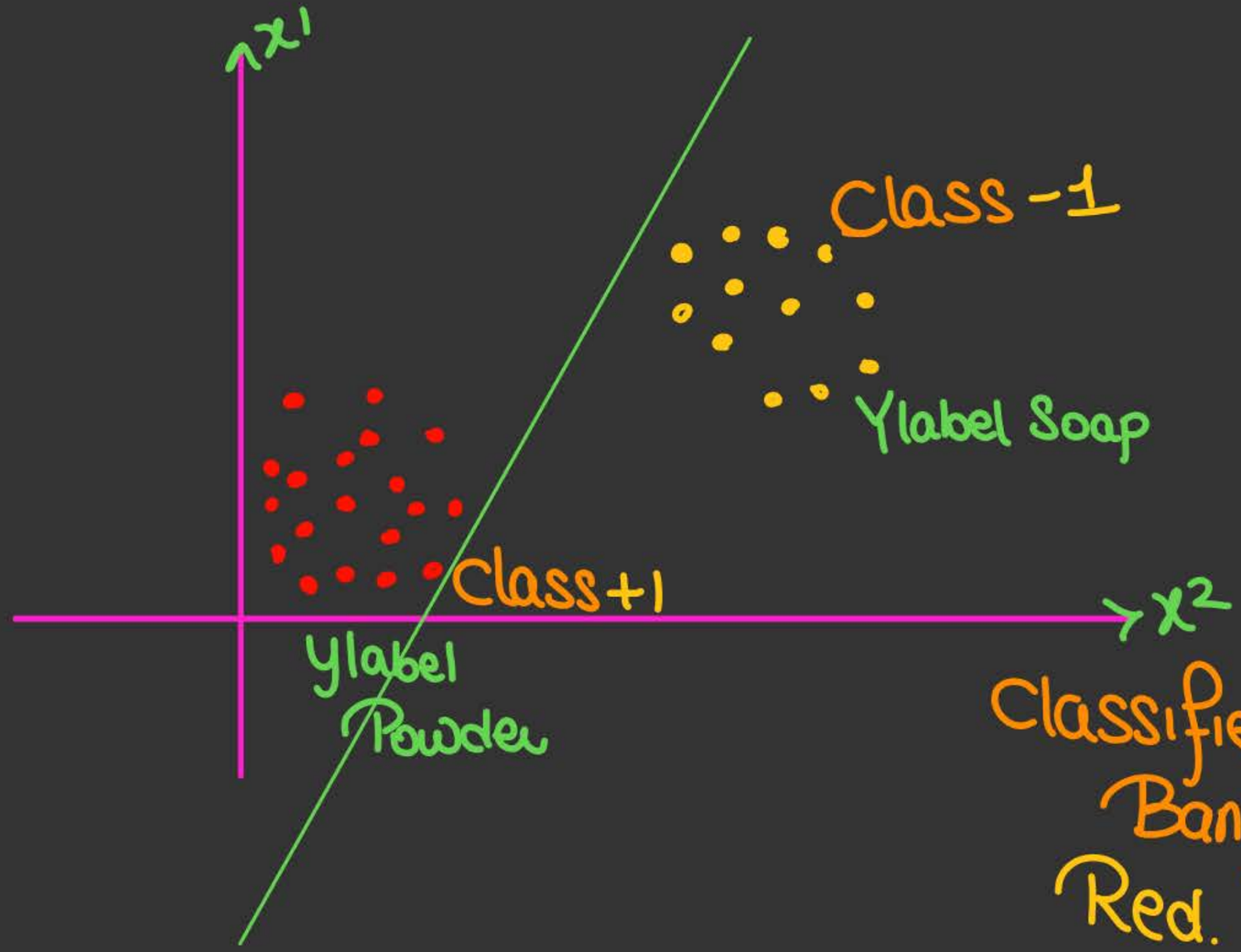
- So we have the training data in 2D we can draw ∞ lines, ∞ classifier
- Best classifier
  max accuracy
  min error

→ Classifier
$\beta_2 x^2 + \beta_1 x' + \beta_0 = 0$

- Classifier sahi : $y_i(x_i\beta) > 0$
- $y_i(x_i\beta) < 0$


Canada
USA
USA ka hai
Canada ka hai
Class+1
$y_i(x_i\beta)>0$
Class1
$y_i(x_i\beta)<0$
Class-1
Class-1
$y_i(x_i\beta)$
$y_i(x_i\beta)<0$
Correct hai
$y_i(x_i\beta)>0$
$y_i(x_i\beta)<0$ point misclassify
$y_i(x_i\beta)>0$ point Sahi classify
N data points
Classifier which has
max value of $\sum_{i=1}^{N} y_i(x_i\beta)$ is Best.


x1
Class +1
Ylabel Soap
Class-1
x2
Ylabel
Powder
Classifier aisa
Banega Jo
Yellow Point ko oopar Rakhe
Red Points ko neeche



Classifier aisa
Banega Jo
Red. Point ko oopar Rakhe
yellow Points ko neeche

Training and Testing of classifier

$$\beta_0 + \beta_1 x^1 + \beta_2 x^2 + \dots - \beta_D x^D = 0.$$

Classifier

D dimension data

- dimension of classifier ⇒ Classifier is a hyperplane of D dimension.
- No. of variables D+1

we need to find β's that give best classifier

Best classifier $\sum_{i=1}^{N} y_i (x_i \beta)$ is maximum

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

4 data point

$$\max\left(y_1\left(\beta_0+\beta_1 x_1^1+\beta_2 x_1^2\right)+ y_2\left(\beta_0+\beta_1 x_2^1+\beta_2 x_2^2\right)- - -\right]$$

gradient-descent nahi lagega.

How to do Testing

- we have the classifier

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

we have a new test point $x_t = (x_t^1, x_t^2, \dots x_t^D)$

How to find the class for test point ⇒

we have to check that test point is above or below the classifier

find $\beta_0 + \beta_1 x_t^1 + \beta_2 x_t^2 + \dots \beta_D x_t^D$

if > 0 point above classifier ⇒ +1

if < 0 " below " ⇒ -1

Linear Classification

Linear Regression of an Indicator Matrix

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

Linear class.

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

B_0 is also called Bias

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.8 \\ -1.2 \end{bmatrix}$$

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] → [1 3 5]

Point B: (−2, 4] → [1 −2 4]

Point C: [1, −1] → [1 1 −1]

Point D: [−4, −3] → [1 −4 −3]

$A\beta = -0.5 + 2.4 - 6 = -ve$ Class-1

$B\beta = -0.5 - 1.6 - 4.8 = -ve$ Class-1

$C\beta = -0.5 + 0.8 + 1.2 = +ve$ Class 1

$D\beta = -0.5 - 3.2 + 3.6 = -ve$ Class -1

Linear classifier

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

Dataset of 4 points:

$$(1,2):+1, \quad (2,1):-1, \quad (3,2):+1, \quad (0,0):-1$$

Classifier: H.W

$$f(x_1, x_2) = x_1 - x_2$$

Predict class as +1 if $f \geq 0$, else -1.

How many points are misclassified?

(1,1) ⇒ 2(1)+3(1)−6 ⇒ (−1) neeche

(3,0) ⇒ 2(3)+3(0)−6 ⇒ 0 on the line

For the hyperplane

$$f(x_1, x_2) = 2x_1 + 3x_2 - 6$$

check whether the points $(1, 1)$ and $(3, 0)$ lie on the **same side** or on **opposite sides** of the boundary.

Classifier:

$$f(x_1, x_2) = x_1 + 2x_2 + b$$

For point $(1, 1)$, classification changes from **+1 to -1** when bias b is decreased.

Find the **critical value of** b at which the point lies exactly on the boundary.

↳ (1,1) On the line ⇒ $f(1,1)=0$, $1+2+b=0$

$b=-3$

x^1	x^2	y
1.	4.	1.
4.	2.	1.
3.	6.	-1
6	4	-1

Process

Which is a better classifier

1. $\beta = \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}^T$

2. $\beta = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$

$\sum y_i(x_i\beta)$ for both Cases

Better.

1. $\left[1(1+(-3)+4) + 1(1+(-12)+2) + (-1)(1-9+6) - 1(1-18+4)\right] = 8$

2. $\left[1(1+3+4) + 1(1+12+2) - 1(1+9+6) - 1(1+18+4)\right] = -16$

Classifier with max $\sum y_i (x_i \beta)$

is Better.

• Linear Classification

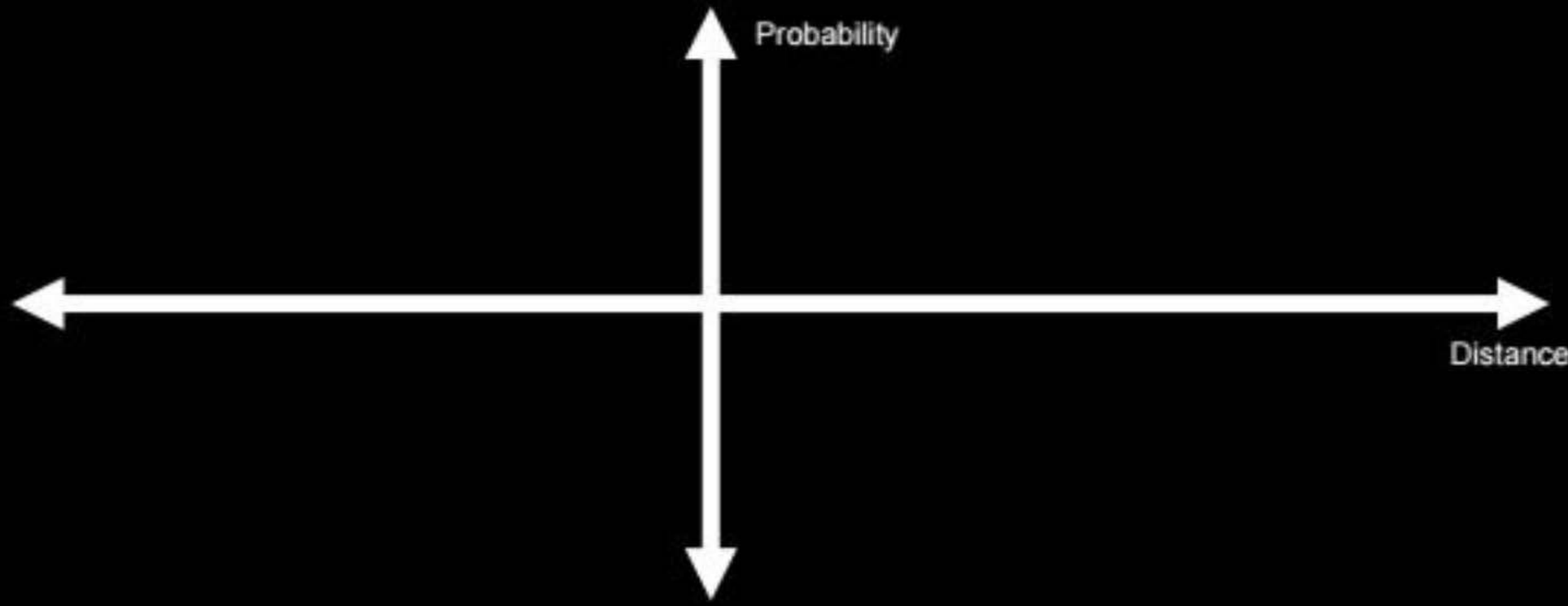
Problem of outliers

• 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

• 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

• 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

Linear Classification

This is called sigmoid...

Also called Sigmoid Function

Linear Classification

Logistic Regression

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

Linear Classification

Logistic Regression

- Logistic regression cannot solve XOR problem...

Linear Classification

Logistic Regression

2 class case

The loss function...

Linear Classification

Logistic Regression

2 class case

We find the parameters by rule ...

Linear Classification

Logistic Regression

Now calculation
Probability is easy...

Linear Classification

Logistic Regression

Simple decision rule in
2 class case

THANK - YOU