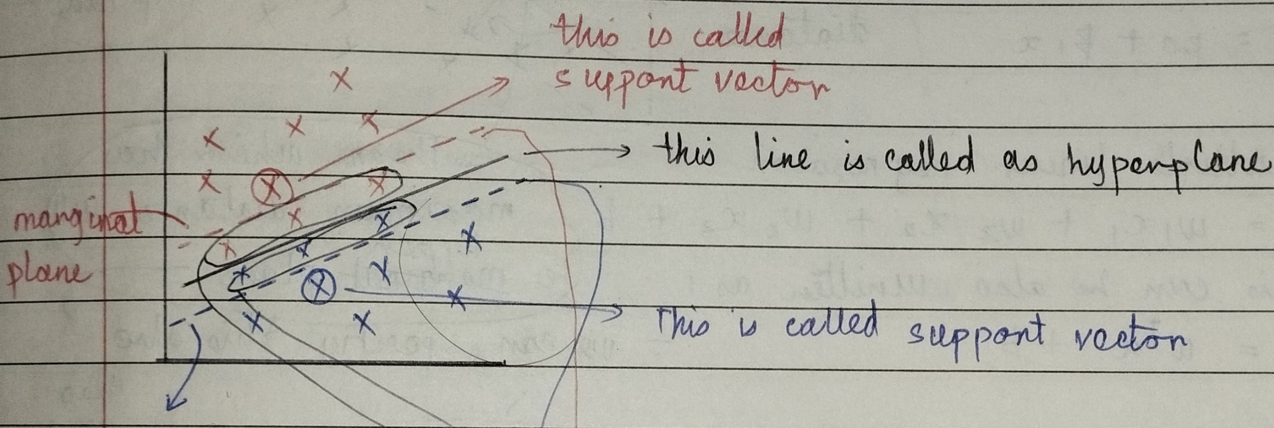
Support Vector Machines (SVM)

Tasks :- 1) Classification 2) Regression

Types :- 1) Support Vector Classifier 2) Support Vector Regressor

→ A support vector machine (SVM) is a powerful and versatile machine learning model, capable of performing linear or non-linear classification, regression.

→ SVM shine with small to medium-sized nonlinear datasets (i.e hundreds to thousands of instances), especially for classification tasks. However, they don't scale very well to large datasets.



If we have data points inside the marginal plane then this is called as soft marginal plane

↓

If outside the marginal plane then it is called as hard marginal plane

Error are there

difficult to exist in real-world datasets

Aim :-

* best-fit line equation

$y = mx + c$

$y = \theta_0 + \theta_1 x$

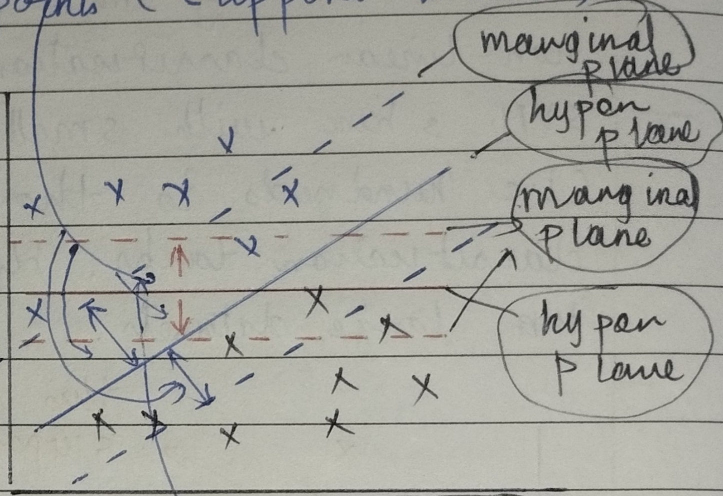
$y = \beta_0 + \beta_1 x$

$y = w_1 x_1 + w_2 x_2 + w_3 x_3 + b \rightarrow \boxed{y = w^T x + b}$

Date

Aim of support Vector Machines :-

→ Find a hyperplane (decision boundry) that best divides your data into classes while maximizing the distance between the marginal planes :- / the planes that pass through the nearest data points (support vectors) from each class.



best-fit line :- equation

$y = mx + c$

$y = \theta_0 + \theta_1 x$

$y = \beta_0 + \beta_1 x$

Aim = maximum distance

Multiple linear regression

$y = w_1 x_1 + w_2 x_2 + w_3 x_3 + b$

→ This can be also written as :

$y = w^T x + b$

The one which has maximum distance will be ~~marginal~~ hyper plane

— we can specify this also

also called as decision boundry

$y = mx + c$ same as $ax + by + c = 0$

→ Both of them are equation of straight line

$y = mx + c$

~~$ax + by + c = 0$~~

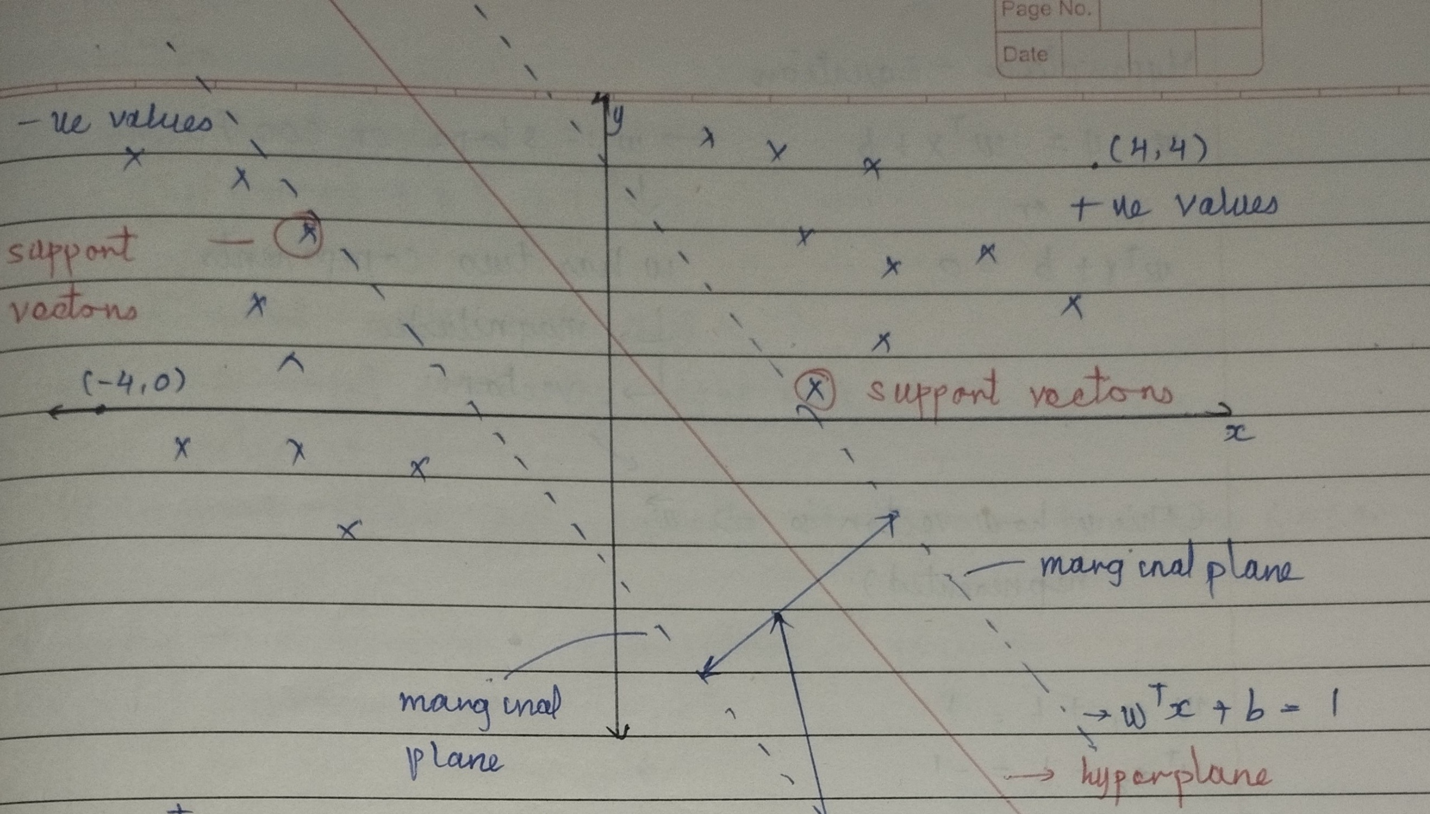
~~$y = -\frac{ax}{b}$~~

$ax + by + c = 0$

$by = -ax - c$

$y = -\frac{a}{b}x - \frac{c}{b}$

here $m = -\frac{a}{b}$ $\quad c = -\frac{c}{b}$



$3x + 2y \pm 4 = 0$:- $ax + by + c = 0$ $\qquad w^T x + b = 0$

① $3(-4) + 2(0) + 4 = 0$

$-12 + 0 + 4 = 0$

$-12 + 4 = 0$ $\qquad w^T x + b = -1$

$-8 = 0$ — negative

② $3(4) + 2(4) + 4 = 0$

$12 + 8 + 4 = 0$ $\qquad$ this as our goal it should or must maximum distance as (difference)

$24 = 0$ — positive value

How we calculate the distance

$y = mx + c$

$y = mx + b$:- $y = w_1 x_1 + b$

Single Linear Regression

↓

Multiple :- $y = w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots\ldots + w_n x_n + b$

↓

This can be represented as ⟶ $y = w^T x + b$

(This is decision boundary or Hyperplane Equation)

w :- weight (slope) - m

x :- input (input) - x

b :- bias (intercept) - c

Page No.
Date

Hyperplane - Equation

$y = w^T x + b$ → w :- slope or coefficient
↓
or
w has two components

$w^T x + b = 0$
↳ magnitude
↳ vectors

(This is how vector is represented) ⇒ $\vec{w}$

$w^T x_1 + b = 1$
$w^T x_2 + b = -1$
(−) (−) (+)

$$\frac{w^T (x_1 - x_2)}{||w||} = \frac{2}{||w||}$$

⇒ Maximize → if we maximize this then our distance is high

with this formula we can get vectors value ($\vec{w}$)

Aim :- We need to maximize by changing w & b

$(w, b) = \frac{2}{||w||}$ ⇒ with this we can find marginal plane distance

2 by magnitude

(optimisation function)

constraint :- such that y $\begin{cases} +1 \text{ when } w^T x + b \geq 1 \\ -1 \text{ when } w^T x + b \leq -1 \end{cases}$

For all accurate datapoints

$y(i) \times (w^T x_i + b) \geq 1$

SVM - Constraint

The main constraint is :

$$\boxed{y_i (w^T x_i + b) \geq 1}$$

This constraint makes sure that :

- Each data point is on the right side of the margin (not inside it)
- All support vectors are exactly at margin boundries (= 1 or -1)

$\underset{(w, b)}{\text{Maximize}} = \frac{2}{\|w\|} \Rightarrow \underset{(w, b)}{\text{Minimize}} = \frac{\|w\|}{2}$

↓

if we minimize → our marginal plane distance will increase

This is our cost (loss) function

Cost function of ~~SVM~~ SVM

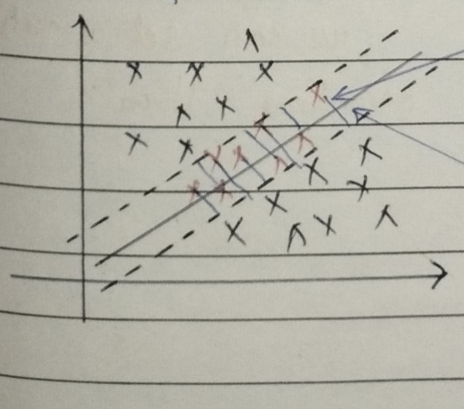
This is our hyperparamter

★ ← $\underset{(w, b)}{\text{Minimize}} = \frac{\|w\|}{2} + C_i \sum_{i=1}^{n} \xi_i$ { How many points we want to or can avoid misclassification }

↓ works to maximize the margin (makes the boundary as far apart as possible)

C :- constant

$C \sum \xi_i$:- adds penalties for misclassified or incorrectly placed points (ξ_i are called slack variables and measure how much a point violates the margin)



C = 8 → means this much error is okay ~~but~~ (allowed)

Page No.
Date

overlapping :- misclassification

$$(w, b) = \min \frac{\|w\|}{2} + C_i \sum_{i=1}^{n} \xi_i$$

How many errors?

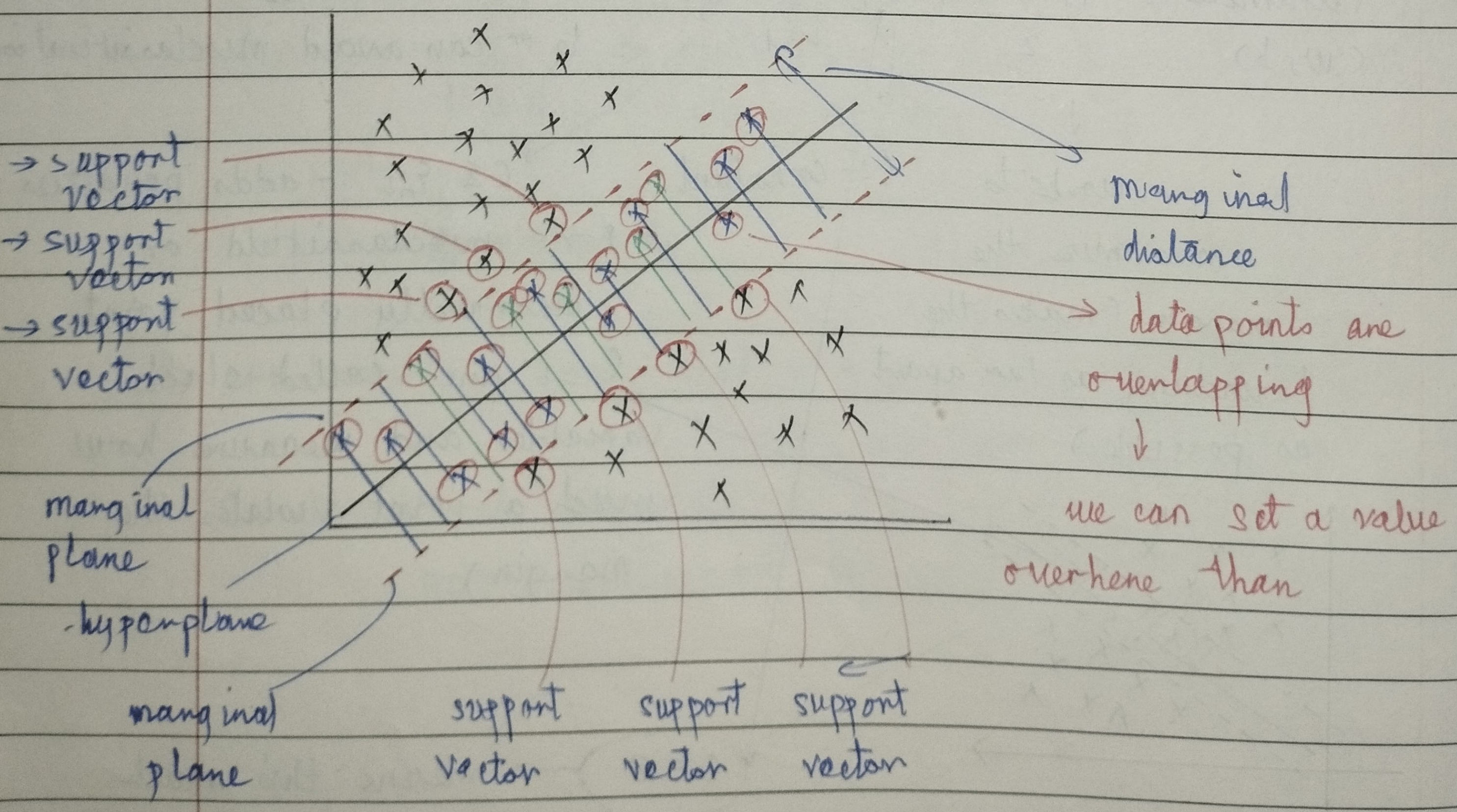
Value of the error

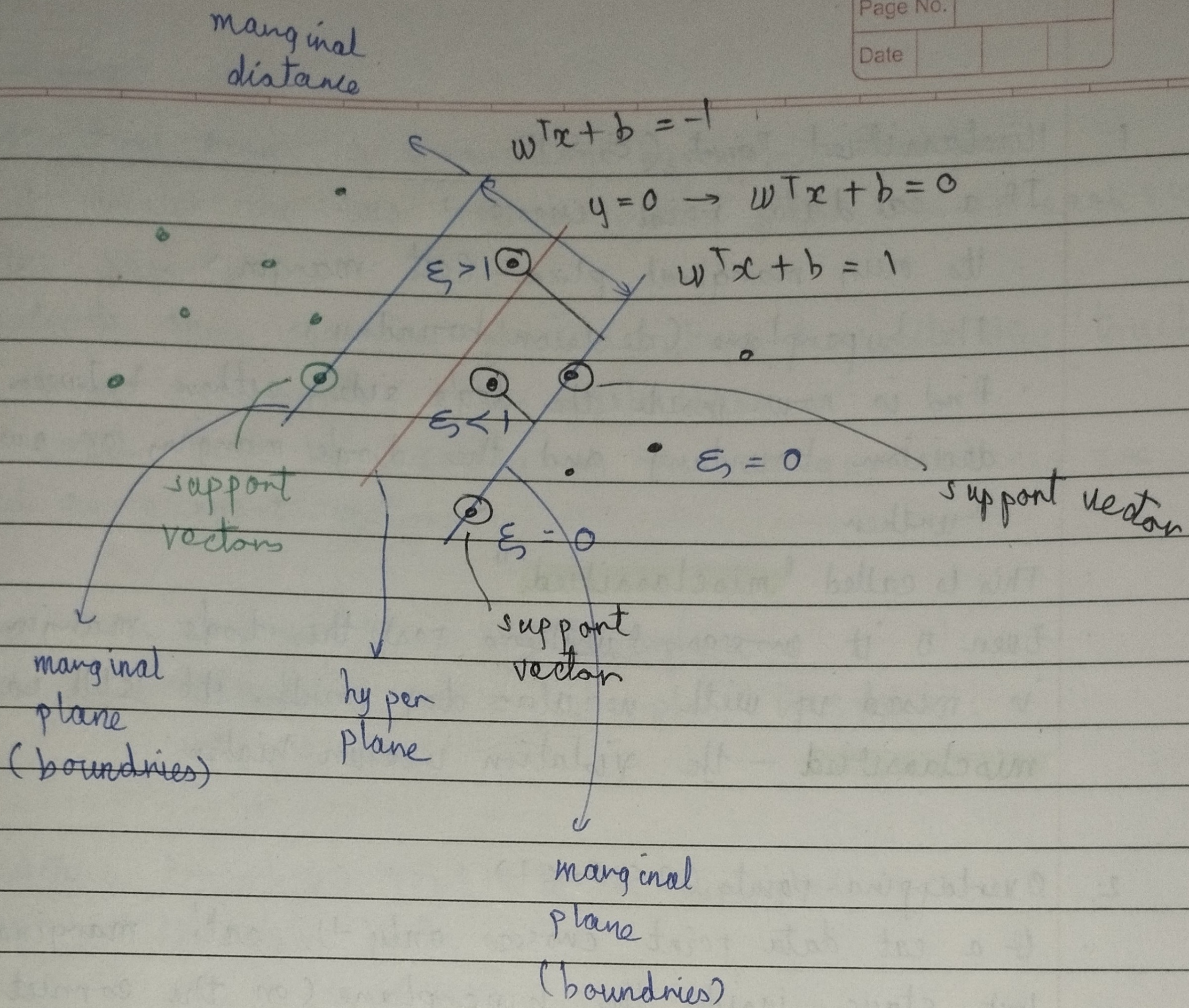
→ How many errors we can consider (can be inside the marginal distance)

→ This also helps us from overfitting the model

→ For eg $c = 5$ it's okay to have this much errors. I don't have to change the hyperplane (decision boundary) with respect to marginal plane

$\Sigma \xi_i$ represents the total margin violations (errors) & C controls the penalty for these errors





- $\xi < 1$:- Points that are inside the margin boundaries but still on the correct side of the hyperplane

→ These are called as overlapping points or margin violators

- $\xi > 1$: Points that are on the wrong side of the hyperplane (misclassified)

→ These are truly misclassified points

1. ~~Misclassified :- If you have data point that actually belongs to "cat" but it appears on the "dog" side of the decision boundary (hyperplane), this point is misclassified. In terms of the SVM slack variable $\xi > 1$ for this point~~
2. ~~Overlapping (Margin Violator) :- If a "cat" data point has crossed into the margin of~~