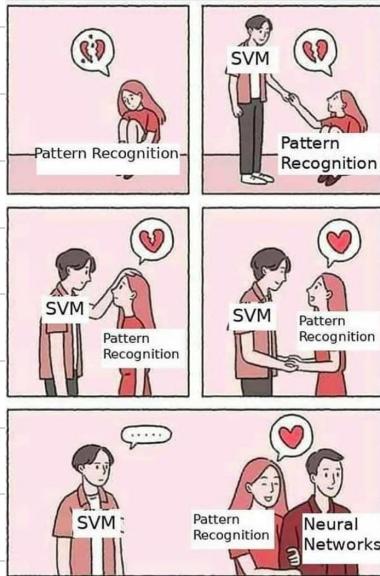


SVM-1

Session - 11

Sep 17, 2025



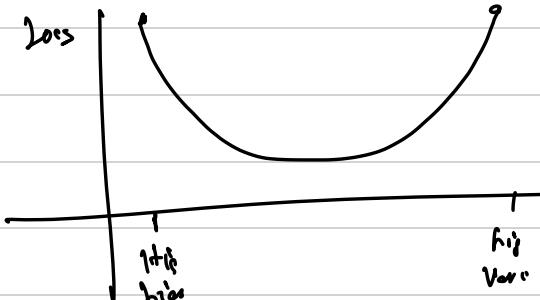
Motivational Speaker: "You can achieve anything you want. You have a tiger in you"

The tiger in me:



AGENDA

- ① Support Vector Machine - SVM
- ② Geometric Intuition
- ③ Hard margin
- ④ Soft margin
- ⑤ Hyper-parameter, Hinge Loss



Support Vector Machine

Example

Problem Statement :

Imagine yourself as a Data Scientist at Google.



Task :

You've been asked to come up with model to classify emails as :

SPAM or HAM



80+ - Human

fall 2012 - SVM
60+

Kernel
Linear

Non-linear
RBF

Image Not.
Aim Not
 $\rightarrow 20+$
Neural Network
Area

SVM has two flavours:

1. Linear
2. Non-linear

Topics we've covered today

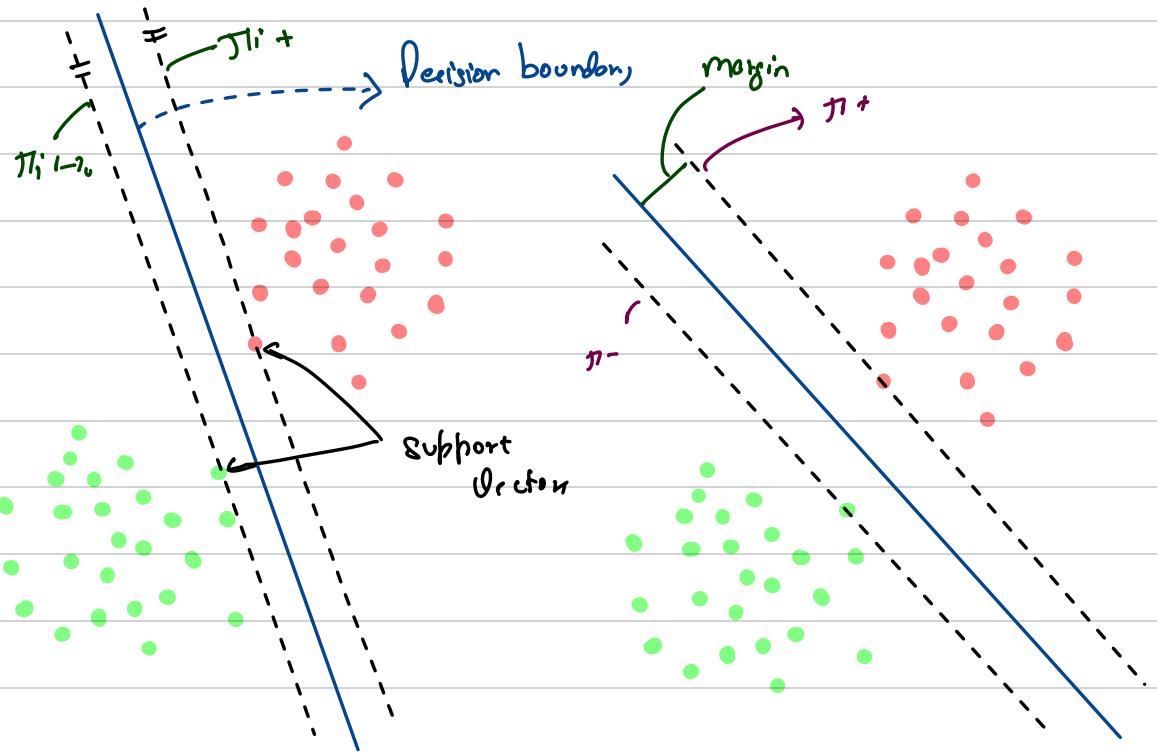
1. Hard-Margin
2. Soft-Margin
3. Hyper-parameters in SVM
4. Hinge-Loss
5. Primal Dual Form

) - Next class

SPAM / HAM Classification

SPAM $\rightarrow 1$
HAM $\rightarrow -1$

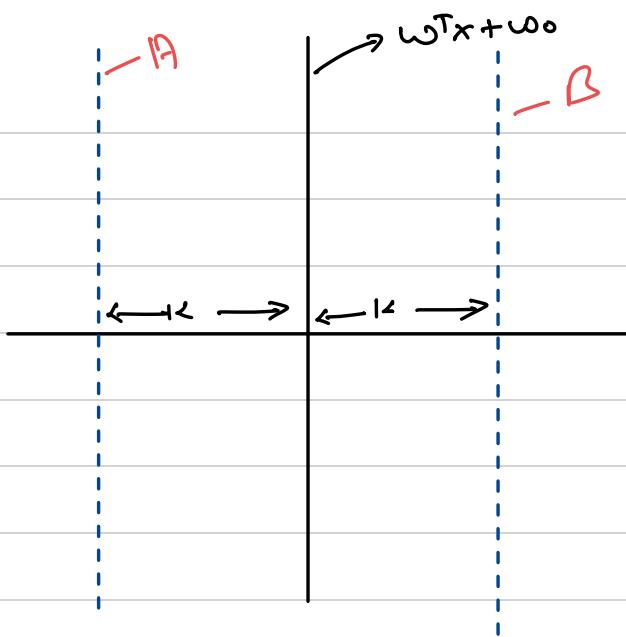
Margin Maximizing Classifier



Assume we have :

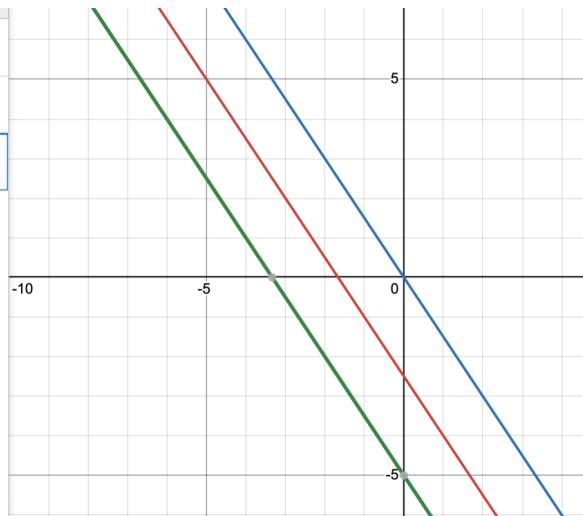
- +ve and -ve data points
- π is a margin-maximizing hyperplane
- π^+ is the +ve hyperplane parallel to π
- π^+ is touching the closest +ve points to π
- π^- is the -ve hyperplane parallel to π
- π^- is touching the closest -ve points to π
- Margin is the dist. between π^+ and π^-

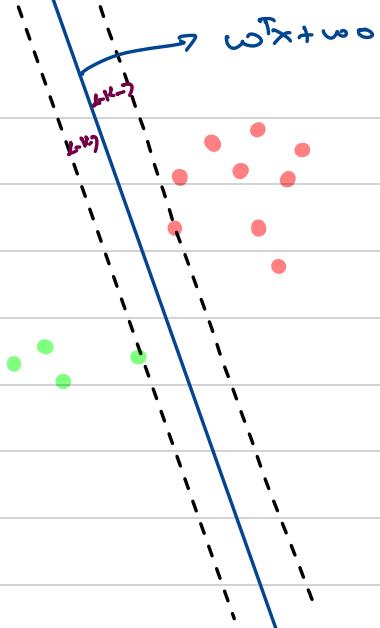




1	<input checked="" type="radio"/> $3x + 2y + 5 = 0$	X
2	<input checked="" type="radio"/> $3x + 2y + 5 = 5$	X
3	<input checked="" type="radio"/> $3x + 2y + 5 = -5$	X
4		

 Log in or sign up to save your beautiful math! X





$$\rho_{i+} = w^T x_i + w_0 = K$$

$$\rho_{i-} = w^T x_i + w_0 = -K$$

Distance b/w ρ_{i+} & ρ_{i-} = $\frac{2K}{\|w\|}$

$$a_1x + b_1 + c_1 = 0$$

$$a_2x + b_2 + c_2 = 0$$

$$\frac{|c_2 - c_1|}{\|w\|}$$

Objective :

$$\underset{\omega}{\operatorname{argmax}} \quad \frac{2K}{\|w\|}$$

Case - 1

$$y_1 = \frac{2K}{\|w\|}$$

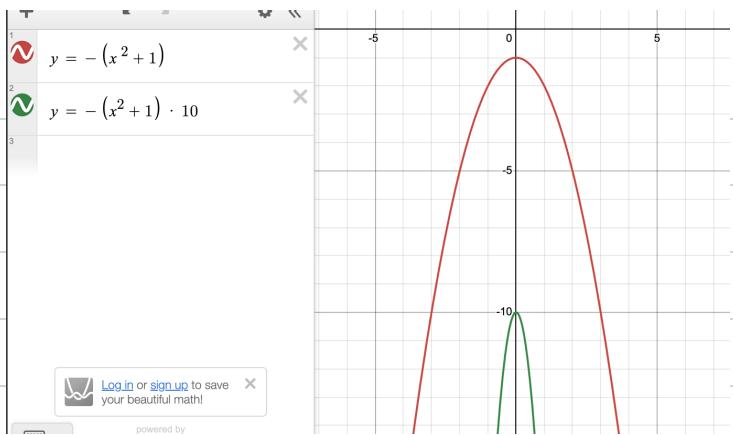
$$\underset{\omega}{\operatorname{argmax}} \omega = \omega_1$$

Case - 2

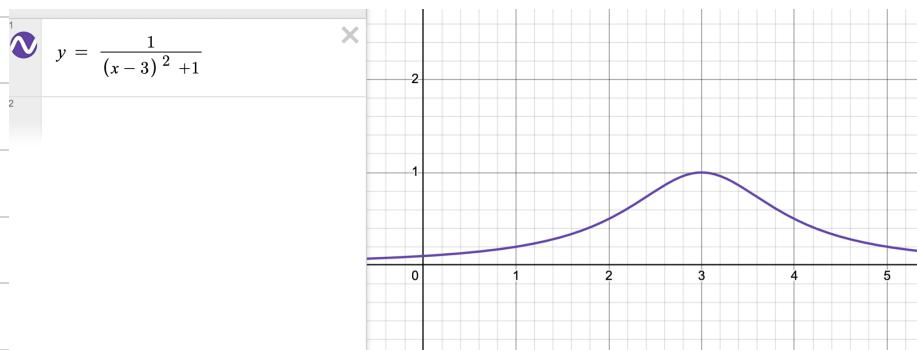
$$y_2 = 10 \times \frac{2K}{\|w\|}$$

$$\underset{\omega}{\operatorname{argmax}} \omega = \omega_2$$

$$\omega_1 = \omega_2$$



$$y = \frac{1}{(x-3)^2 + 1} ; \text{ argument of } n ; y=1 \text{ for } x=3$$



Let's assume in the objective $\frac{2}{\|w\|}$

$$1^T = 1 ;$$

$$\text{now objective} = \frac{2}{\|w\|}$$

What do you mean by generalization in terms of SVM?

1 user has participated

- A How far the hyperplane is from the training datapoints 0%
- B How accurately the SVM can predict outcomes for unseen data 100%
- C How accurately the SVM classifies training datapoints 0%

[End Quiz Now](#)



Athira C R
1/1 91.40



SaiJyothi
1/1 95.40



Souvik Adhikary
1/1 89.37

4		Shreya gupta	1/1 85.50
5		Rakesh Karade	1/1 85.23
6		Anikta Parhi	1/1 83.07
7		Purushottam Kumar	1/1 82.37
8		Tanvi Singh	1/1 77.70
9		Hanumanthgouda Patil	1/1 63.53

- "
- $\pi_+ : w^T x + b = 40$
 - $\pi_- : w^T x + b = -50$
 - then margin will be:

0 users have participated

- A $10/\|w\|$ 0%
- B $40/\|w\|$ 0%
- C $50/\|w\|$ 0%
- D $90/\|w\|$ 0%

[End Quiz Now](#)

Based on all quizzes from the session



Rakesh Karade
2/2 175.86



Athira C R
2/2 183.26



Ankita Parhi
2/2 169.17

4		Purushottam Kumar	2/2 156.43
5		Tanvi Singh	2/2 145.13
6		SaiJyothi	1/2 95.40
7		Souvik Adhikary	1/2 89.37
8		SHASHANK JHA	1/2 87.10
9		Periseta Pavan Kalyan	1/2 86.00
10		Shreya gupta	1/2 85.50

$$a_n t + b y + c_1 = 0$$

$$a_n t + b y + c_2 = 0$$

$$\frac{|c_2 - c_1|}{\|w\|}$$

$$\frac{|-50 - 40|}{\|w\|} = \frac{(-90)}{\|w\|} = \frac{90}{\|w\|}$$

Case - 1

$$y_1 = \frac{2}{\|\omega\|}$$

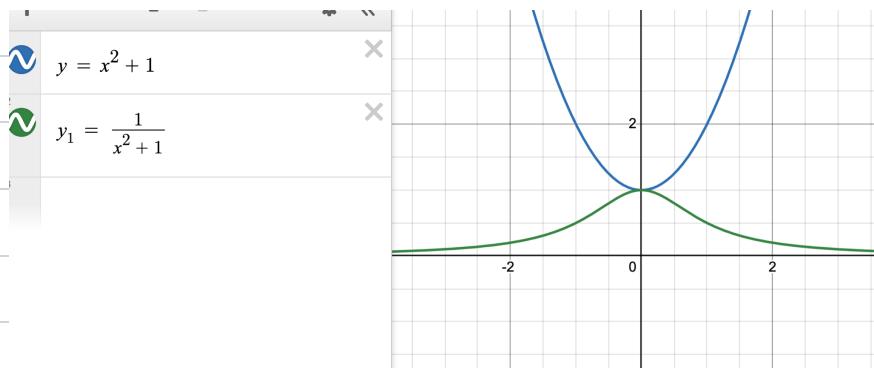
$$\operatorname{argmax} \omega = \omega_1$$

Case - 2

$$y_2 = \frac{\|\omega\|}{2}$$

$$\operatorname{argmin} \omega = \omega_2$$

Again $\omega_1 = \omega_2$

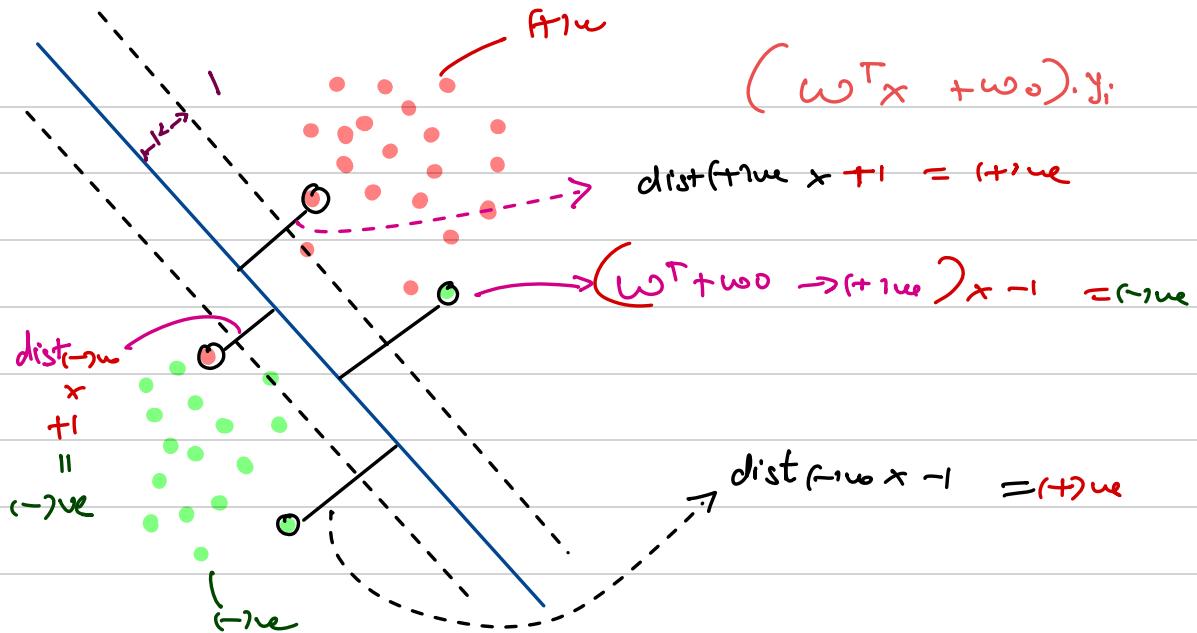


Hard Margin Classifier

$$\operatorname{argmin} \frac{\|\omega\|}{2}$$

S.T. $(\underbrace{\omega^T x_i + \omega_0}_{\text{All points are}}) \cdot y_i \geq 1$

correctly classified



What do you mean by a hard margin?

0 users have participated

- A The SVM allows no error in classification. 0%
- B The SVM allows some error in classification. 0%
- C The SVM allows high error in classification. 0%

[End Quiz Now](#)

	Rakesh Karade 3/3 ✓ 268.59		Athira CR 3/3 ✓ 270.10		Ankita Parhi 3/3 ✓ 259.50
4 PK Purushottam Kumar	3/3 ✓ 248.63				
5 TS Tanvi Singh	3/3 ✓ 234.23				
6 SG Shoreya gupta	2/3 ✓ 181.50				
7 SJ SHASHANK JHA	2/3 ✓ 179.67				
8 PH Paramhans N. Chetiwai	2/3 ✓ 172.63				
9 PP Perisetta Pavan Kalyan	2/3 ✓ 165.07				
10 N Narayana	2/3 ✓ 151.73				

Soft Margin Classifier

$$\underset{\omega}{\operatorname{argmin}} \frac{\|\omega\|}{2}$$

S.T. $(\omega^T x_i + \omega_0) \cdot y_i \geq 1 - \varepsilon_i$

All points are correctly classified

$$x_1 \rightarrow \text{Error}(x_1) = 0$$

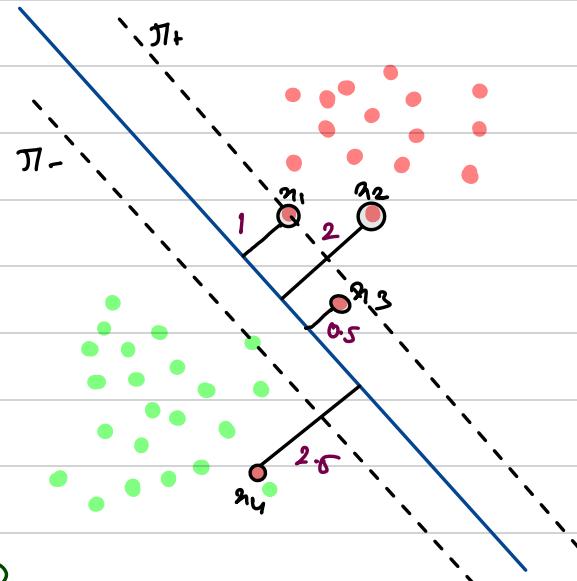
$$x_2 \rightarrow \text{Error}(x_2) = 0$$

$$x_3 \rightarrow \text{Error}(x_3) = 1 - \varepsilon_{x_3}$$

$\frac{1}{2}$
 0.5

$$x_4 \rightarrow \text{Error}(x_4) = 1 - (-2.5)$$

$= 3.5$ $\sum \varepsilon_{x_i}$

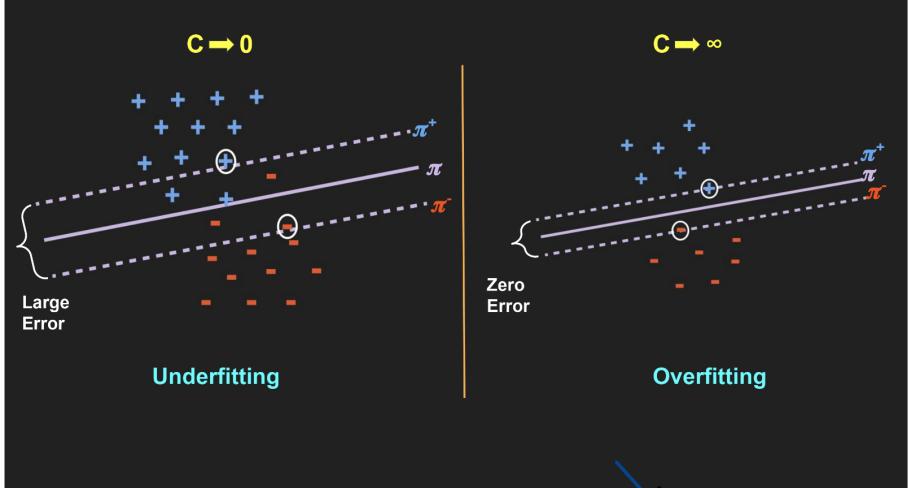


$$\underset{\omega}{\operatorname{argmin}} \frac{\|\omega\|}{2} + \frac{C}{N} \times \sum_{i=1}^N \varepsilon_i$$

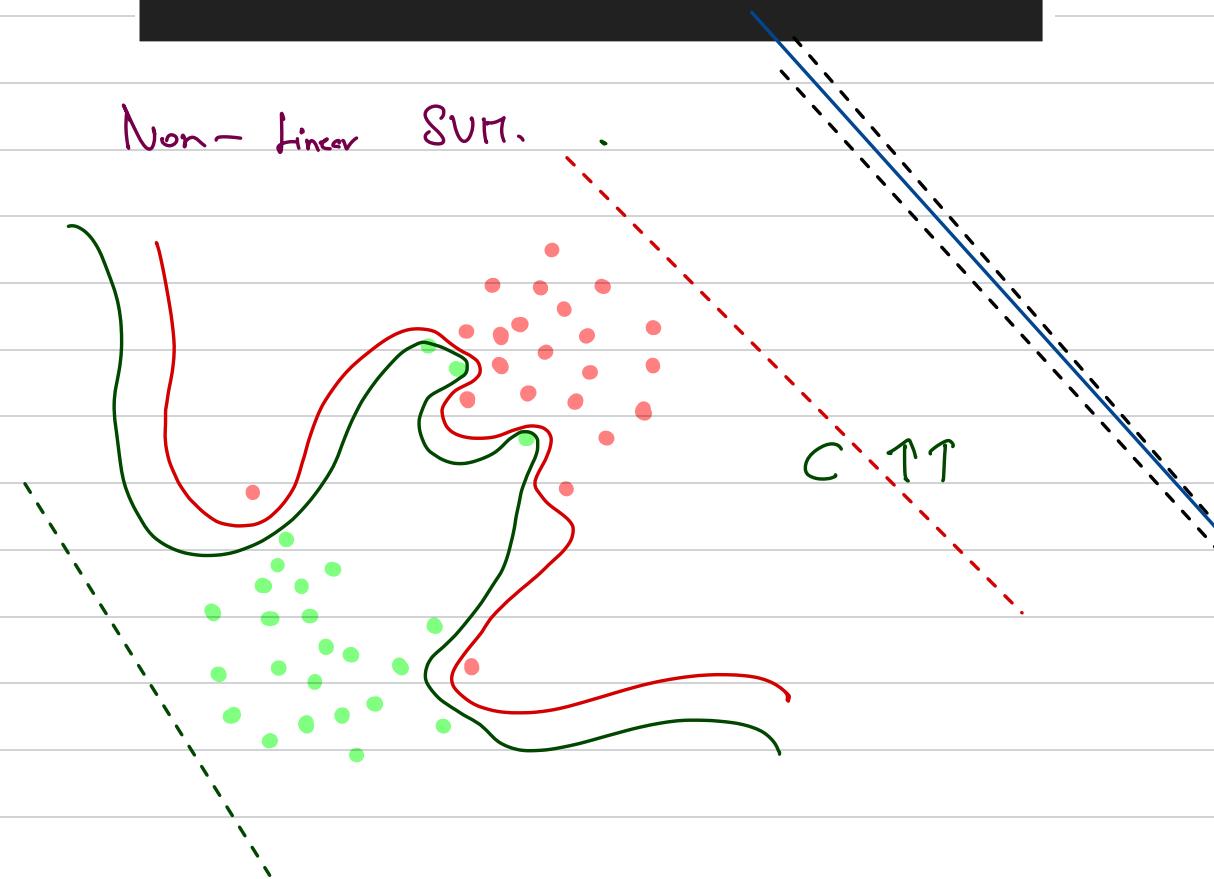
Loss function.

$C \rightarrow$ v. high. overfit

$C \rightarrow$ too low - underfit



Non - Linear SVM.



What would happen when you use very large value of C?

0 users have participated

A We can still classify training data correctly for given value of C.

0%



B We cannot classify training data correctly for given value of C.

0%

C Can't say for sure

0%

[End Quiz Now](#)

Rank	User	Score
1	Purushottam Kumar	4/4 309.66
2	Shreya gupta	3/5 272.33
3	SHASHANK JHA	3/5 264.70
4	Periseta Pavan Kalyan	3/4 254.03
5	Tanvi Singh	3/4 234.23
6	Samyuktha Ramesh	2/4 185.96
7	Souvik Adhikary	2/4 178.77

x_1, x_2, x_3 are -ve datapoints which are 0.2, 3.0, 1.0 at unit distance below the π , what will be their respective ξ ?

3 options

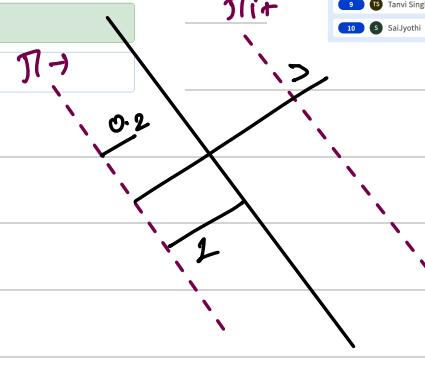
Active Duration(Most preferred: 30 seconds)

Appears for 90 Secs

A 0.8, -2.0, 0.0

B 0.2, 3.0, 1.0

C 0.8, 2.0, 0.0



Leaderboard
Based on all quizzes from the session



Rank	User	Score
4	Purushottam Kumar	4/5 309.66
5	Shreya gupta	3/5 272.33
6	SHASHANK JHA	3/5 264.70
7	Periseta Pavan Kalyan	3/4 254.03
8	Samyuktha Ramesh	3/5 246.00
9	Tanvi Singh	3/5 234.23
10	Saljuyyothi	3/5 232.94

SVM will be impacted if there's an imbalance in the no. of datapoints belonging to each class?

0 users have participated

A True

0%

B False

0%

[End Quiz Now](#)



Rank	User	Score
4	Periseta Pavan Kalyan	4/5 348.53
5	Anikta Parhi	4/5 346.20
6	Purushottam Kumar	4/5 309.66
7	Paramhans N. Chetiwal	3/5 265.33
8	SHASHANK JHA	3/5 264.70
9	Snehal Adhikary	3/5 261.58
10	Samyuktha Ramesh	3/5 246.00

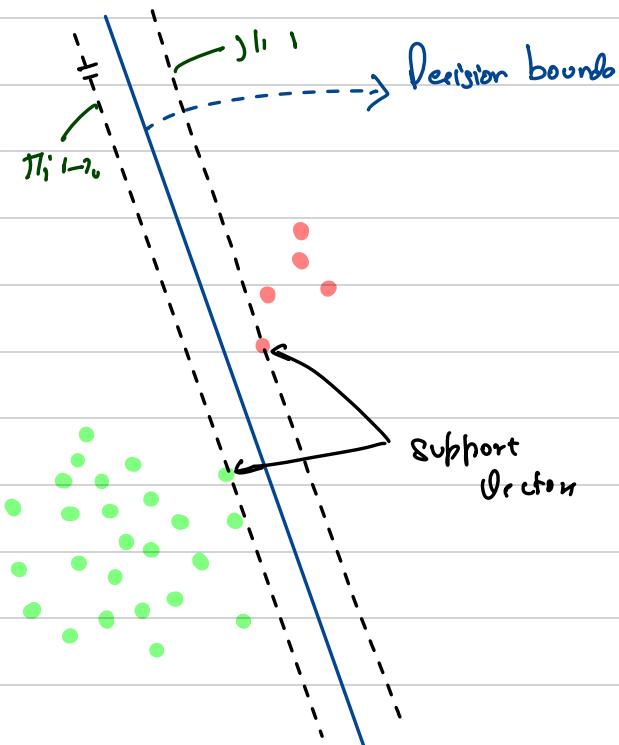
Are SVMs affected by class imbalance?

Only a few data points contribute to the **Hinge loss** (ξ_i) .

These points are called **Support Vectors**.

Hence, SVM is only affected by **imbalance in no. of support vectors** for each class.

NOTE: The balance in no. of support vectors can't be guaranteed.



$$L = \sum (y_i - f)^2 + \lambda \|\omega\|^2 \quad \lambda = \frac{C}{n} \rightarrow \text{underfit}$$

