Session - 11

SVM - 1

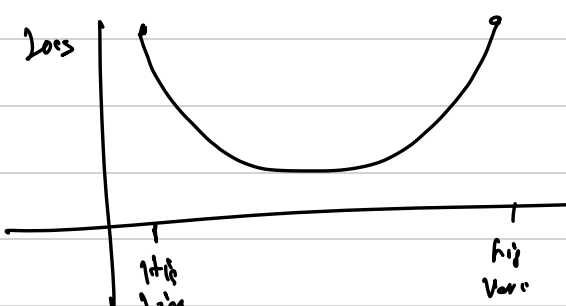
Sep 17, 2025

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

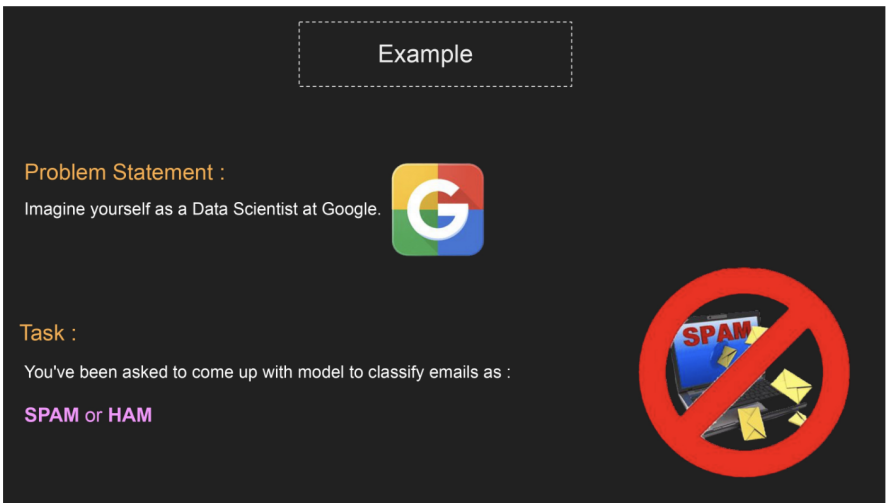
The tiger in me:

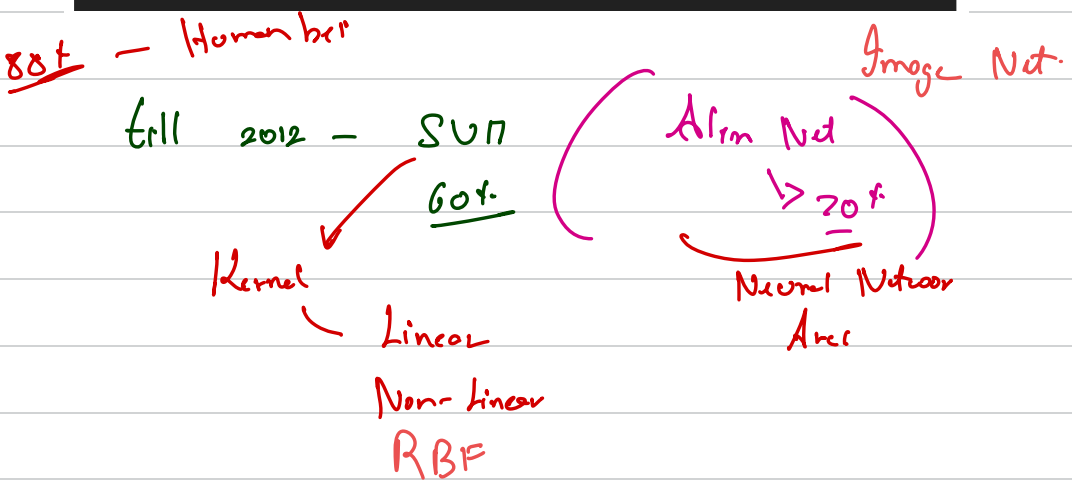
AGENDA

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



Support Vector Machine





SVM has two flavours:
1. Linear
2. Non-linear

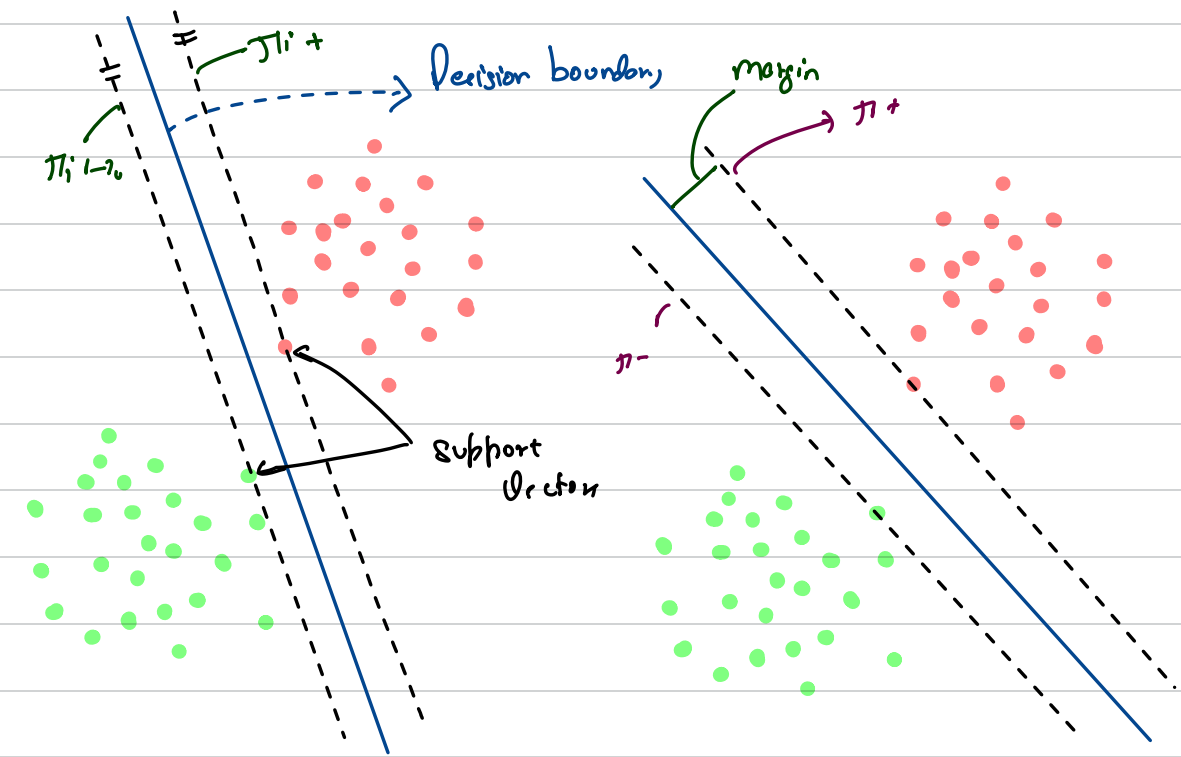
Topics we've for today
1. Hard-Margin
2. Soft-Margin
3. Hyper-parameters in SVM
4. Hinge-Loss
5. Primal Dual Form — Next class

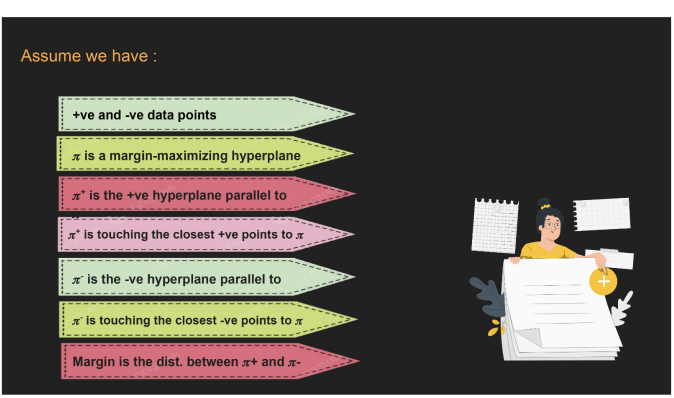
SPAM / HAM CLASSIFICATION

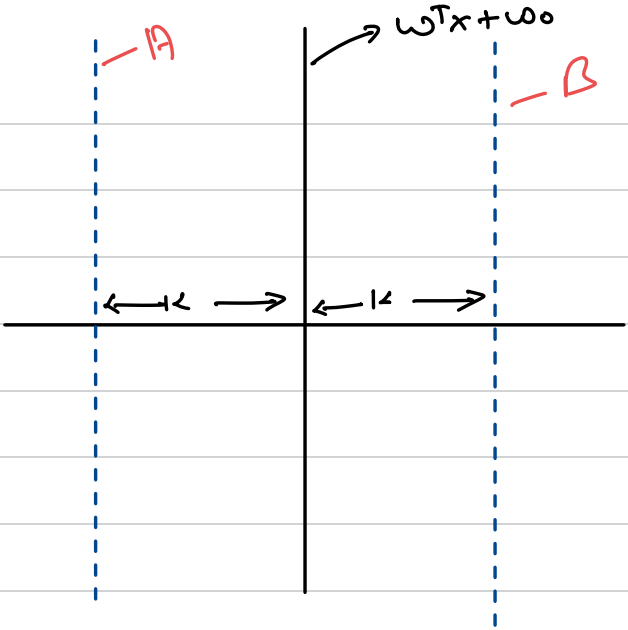
Spam → 1

Ham → -1

Margin Maximizing Classifier

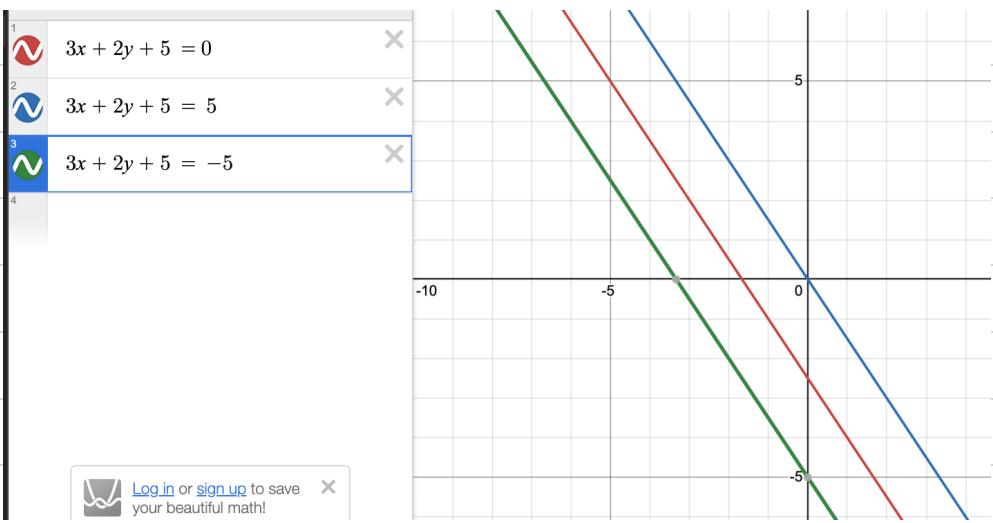


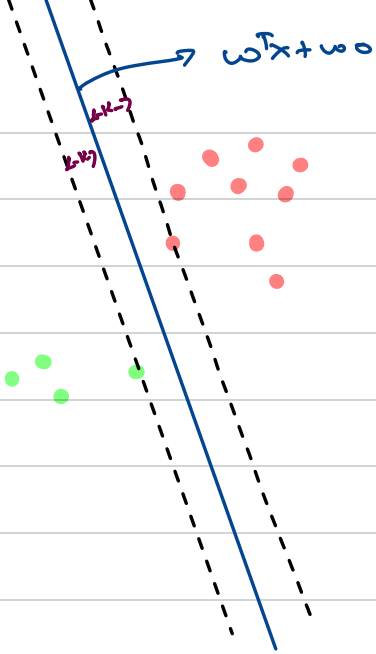




$A = w^T x + w_0 = -1$

$B = w^T x + w_0 = +1$





$$\pi_{i+} = w^T x + w_0 = k$$

$$\pi_{i-} = w^T x + w_0 = -k$$

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

$$ax + by + c_1 = 0$$

$$ax + by + c_2 = 0$$

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

Objective: $\underset{w}{\text{argmax}} \ \frac{2k}{\|w\|}$

Case - 1

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

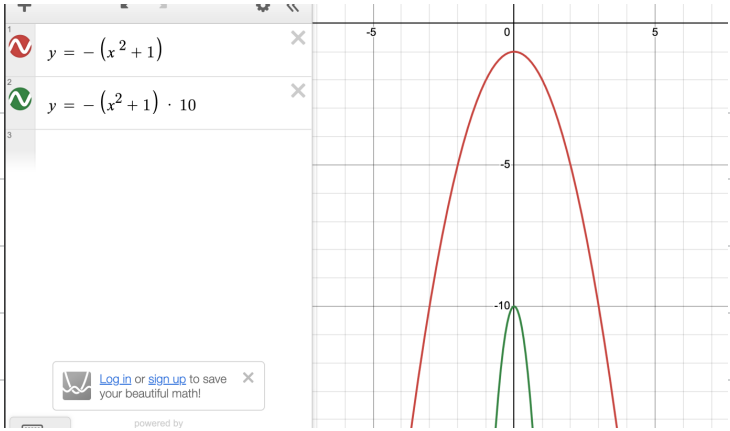
$\text{argmax}_w = w_1$

Case - 2

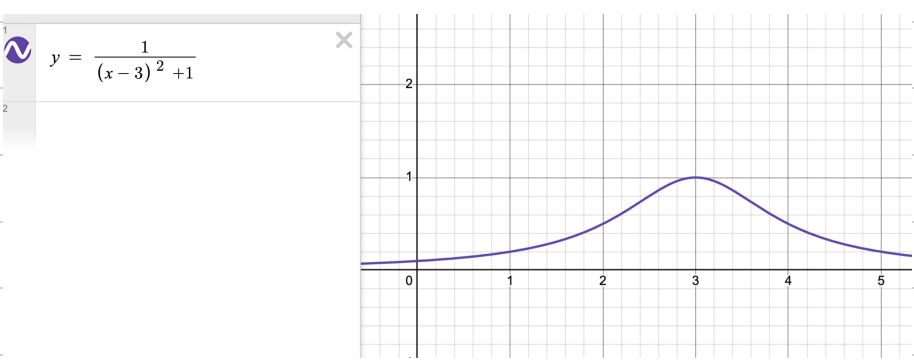
$$y_2 = 10 \times \frac{2k}{\|w\|}$$

$\text{argmax}_w = w_2$

$w_1 = w_2$



$y = \dfrac{1}{(x-3)^2+1}$; argmax of x ; $y = 1$ for $x = 3$



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

$K = 1$;

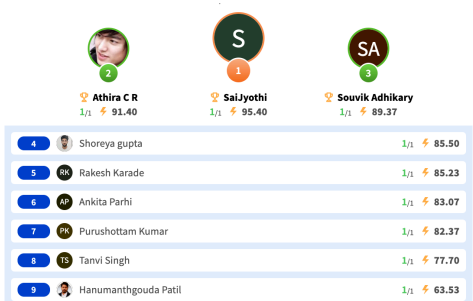
now objective = $\frac{2}{||w||}$

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



...

- π+ : w^T * x + b = 40
- π- : w^T * x + b = -50

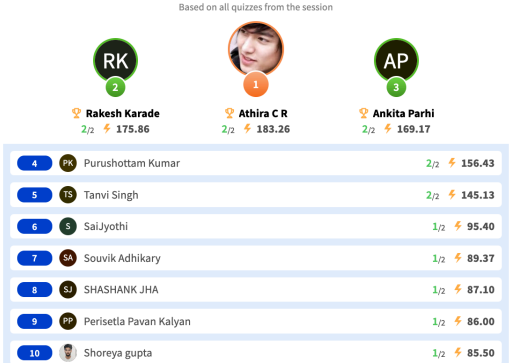
then margin will be:

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



$ax + by + c_1 = 0$

$ax + by + c_2 = 0$

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

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

Case - 1

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

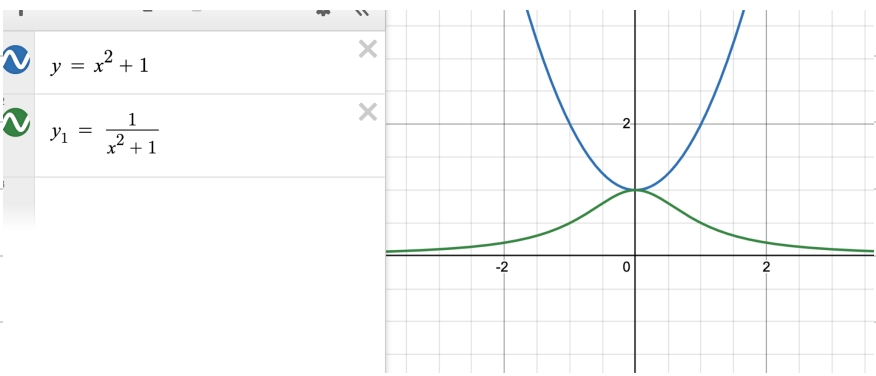
$\text{argmax}_w = w_1$

Case - 2

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

$\text{argmin}_w = w_2$

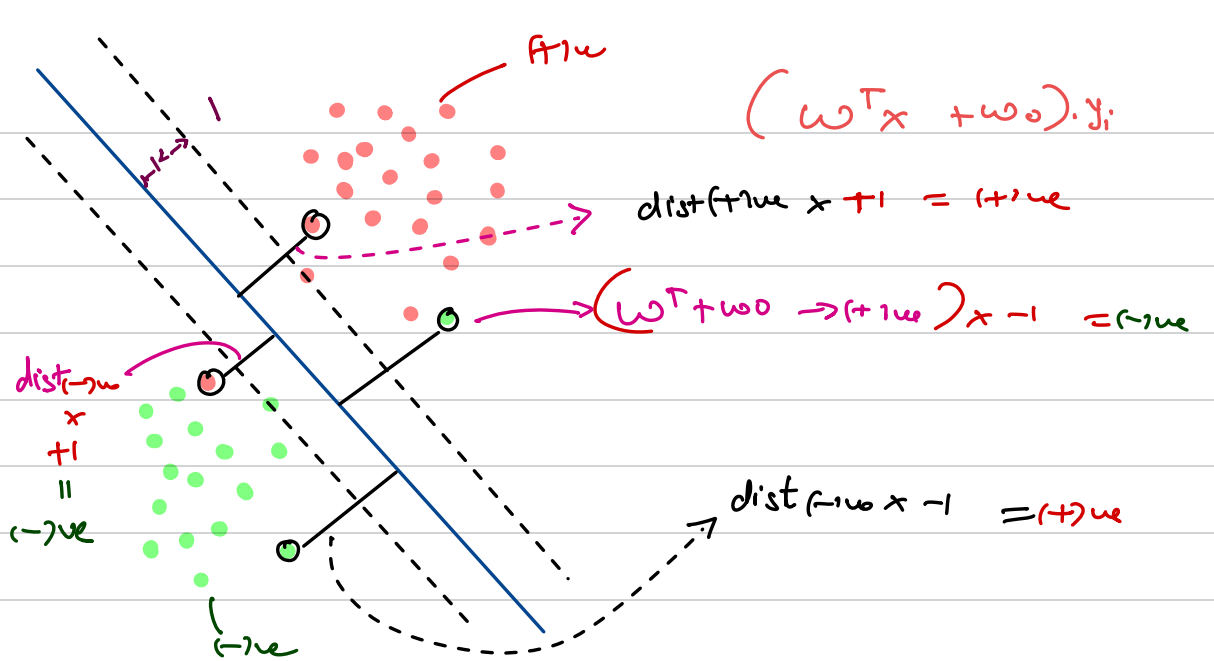
Again $w_1 = w_2$



Hard Margin Classifier

$$\text{argmin} \frac{\|w\|}{2} \quad \text{S.T.} \quad (w^T x + w_0) \cdot y_i \geq 1$$

All points are correctly classified

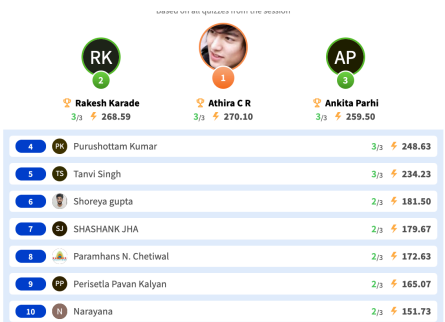


What do you mean by a hard margin?

0 users have participated

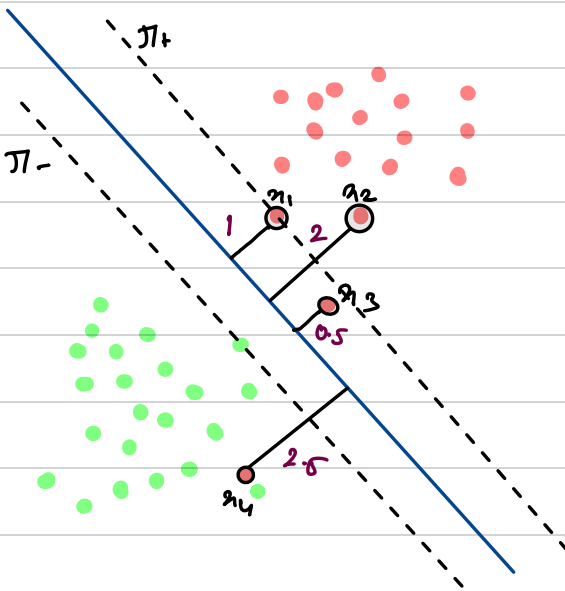
✓	A	The SVM allows no error in classification.	0%
	B	The SVM allows some error in classification.	0%
	C	The SVM allows high error in classification.	0%

End Quiz Now



Soft Margin Classifier

$$\text{argmin} \quad \frac{||w||}{2} \quad \text{S.T.} \quad \underbrace{(w^T x + w_0) \cdot y_i \geq 1 - \varepsilon_i}_{\text{All points are correctly classified}}$$



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

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

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

$= 0.5$

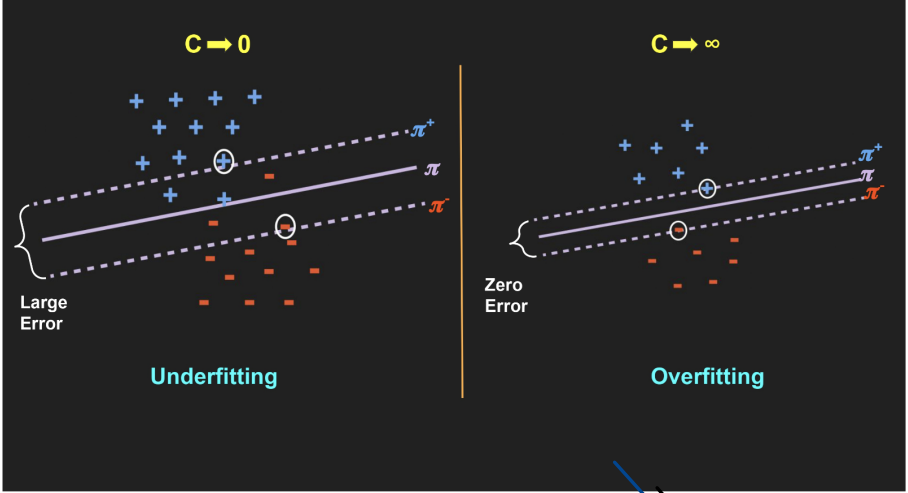
$x_4 \rightarrow \text{Error}(x_4) = 1 - (-2.5)$

$= 3.5 \quad \leftarrow \varepsilon_{x_4}.$

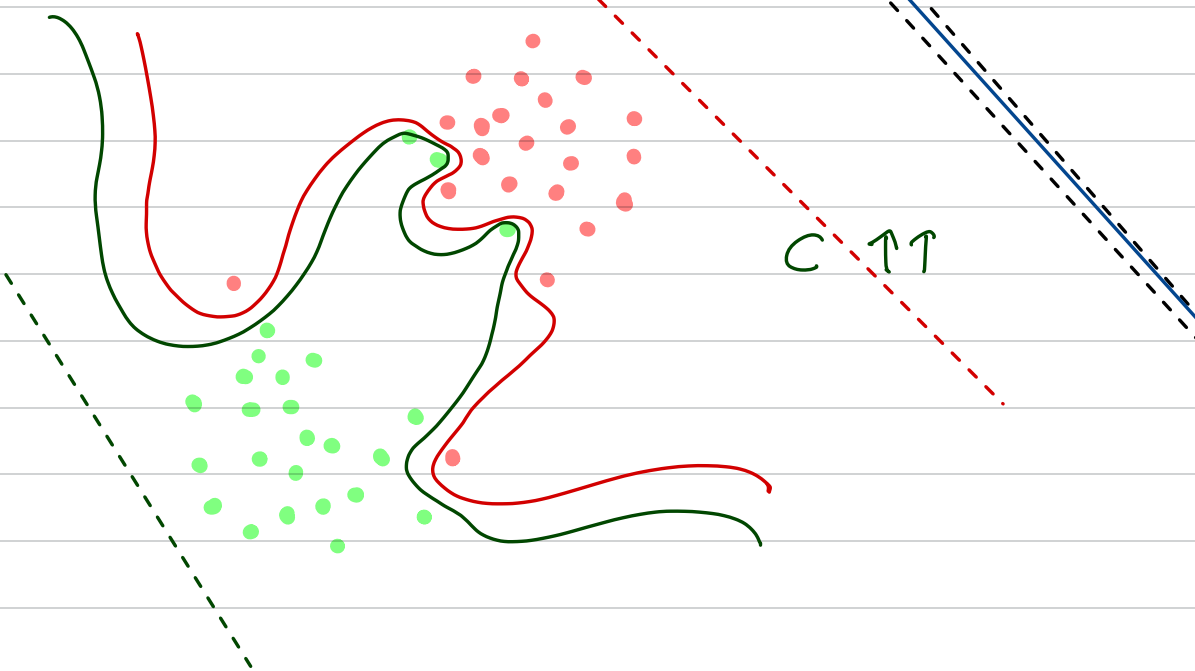
$$\underbrace{\text{argmin} \quad \frac{||w||}{2} + \frac{C}{N} \times \sum_{i=1}^{N} \varepsilon_i}_{\text{Loss function.}}$$

$C \rightarrow$ 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

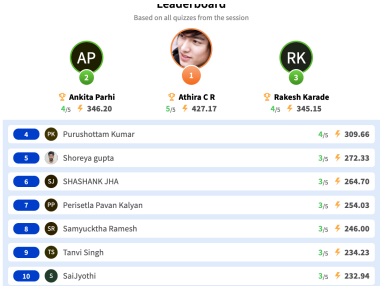
x1,x2,x3 are -ve datapoints which are 0.2, 3.0, 1.0 at unit distance below the π-,
what will be their respective ζi?

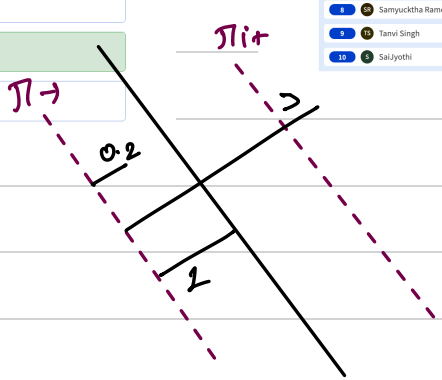
3 options

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





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%

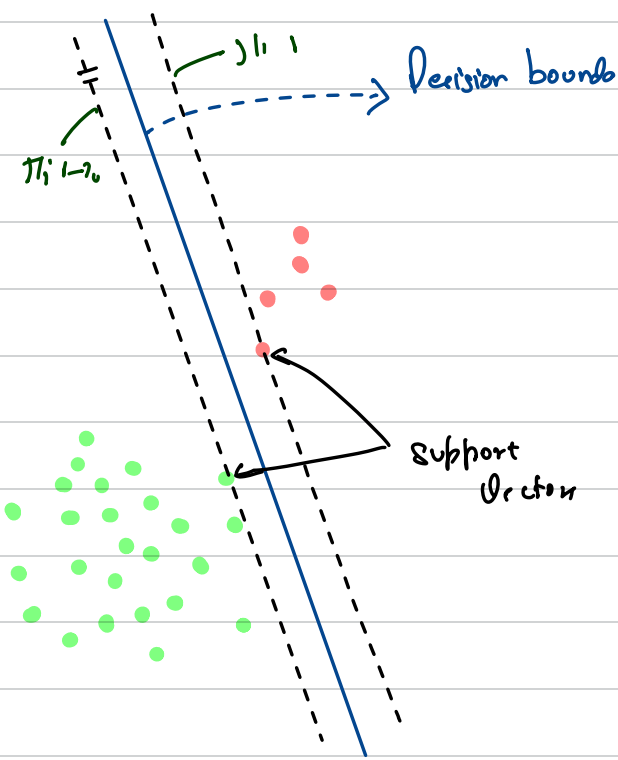
End Quiz Now

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.



$L = \sum (y_i - \hat{y})^2 + \lambda \times w2$ λ - This → underfits