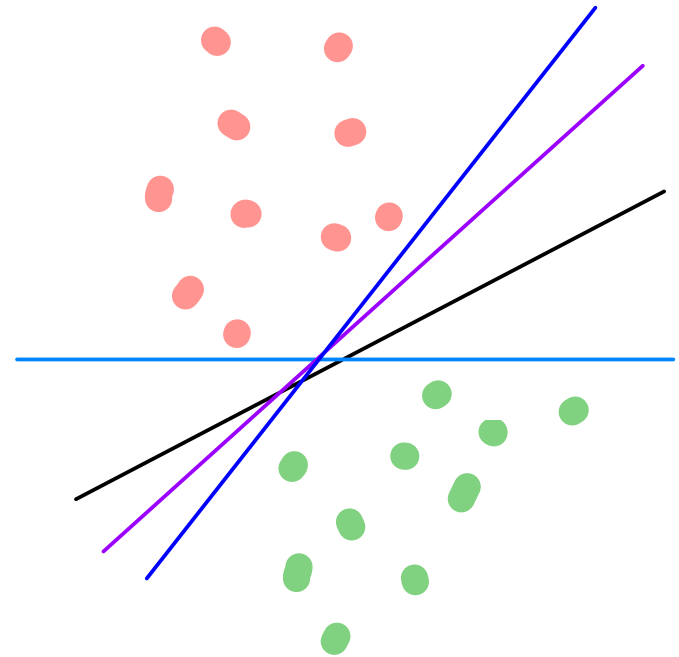
SVM-1

Support Vector Machines

Spam / NonSpam

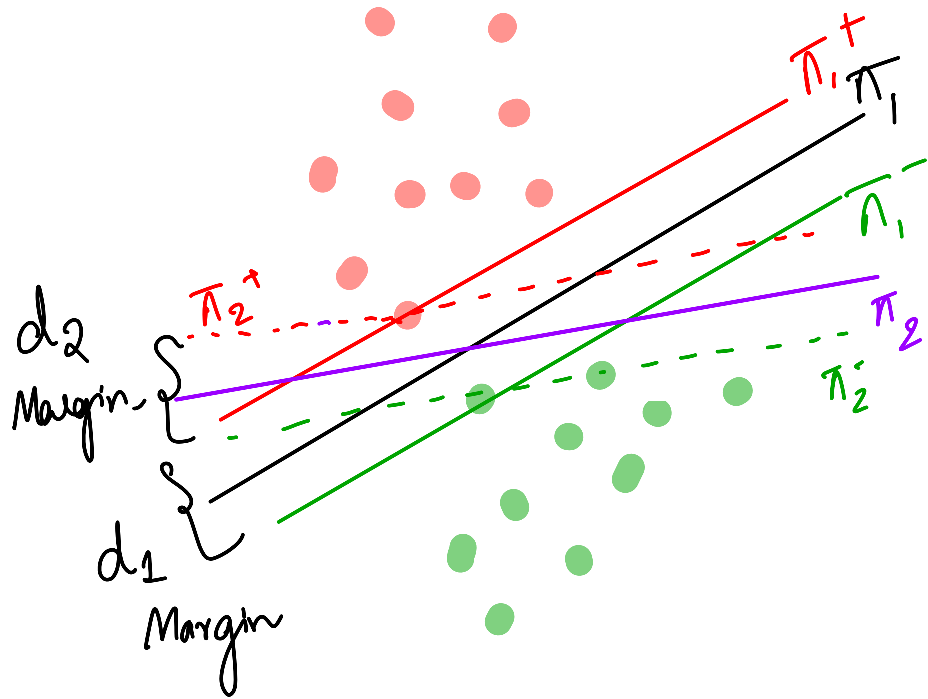
* Theoretically it is most powerful algorithm.
* Practically ⇒ NOT frequently used these day
* 1900's , 2000's
* Mathematically heavy



Q̲ Which hyperplane is best?

A̲ Choose π with highest margin "d"

MARGIN
MAXIMISING
CLASSIFIERS



$\pi \qquad w^T x + b = 0$

~~{0, 1}~~ $\qquad$ +ve $\qquad \{+1, -1\}$

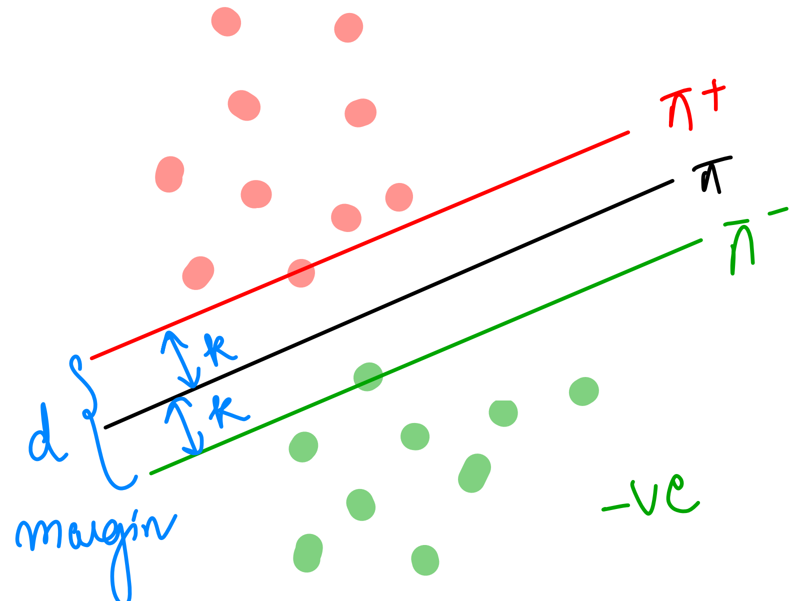
$\pi^+ : w^T x + b = k \qquad w^T x + b - k = 0$

$\pi^- : w^T x + b = -k \qquad w^T x + b + k = 0$

Margin $\Rightarrow d(\pi^+, \pi^-)$

$$d(\pi^+, \pi^-) = \frac{|b + k - (b - k)|}{\|w\|}$$

$$\boxed{d(\pi^+, \pi^-) = \frac{2k}{\|w\|}}$$



$d = \dfrac{2k}{\|w\|}$ } ← MAXIMISE

k = Constt.

Case 1

$k = 1$

$$d = \frac{2 \times 1}{\|w\|} = \frac{2}{\|w\|}$$

$$\underset{w}{\arg\max} \ \frac{2}{\|w\|}$$

Case 2

$k = 10$

$$d = \frac{2 \times 10}{\|w\|} = \frac{20}{\|w\|}$$

$$\underset{w}{\arg\max} \ \frac{20}{\|w\|}$$

By default, we will take $k = 1$ for simplicity of calculation.

GOAL = Maximise Margin $\Rightarrow \dfrac{2}{||w||}$

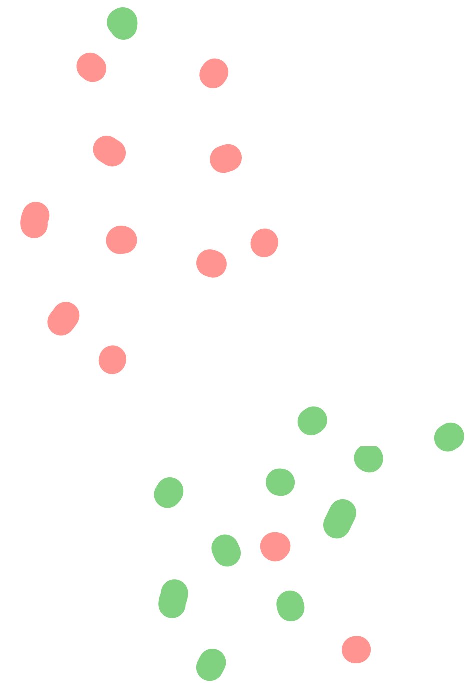
s.t. all points are lying in correct region

This is known as HARD MARGIN CLASSIFIER

Q When will this fail?

A When the data is non linearly separable

Hard Margin Classifier will fail



GOAL: Maximise Margin. s.t. all points are correctly classified.

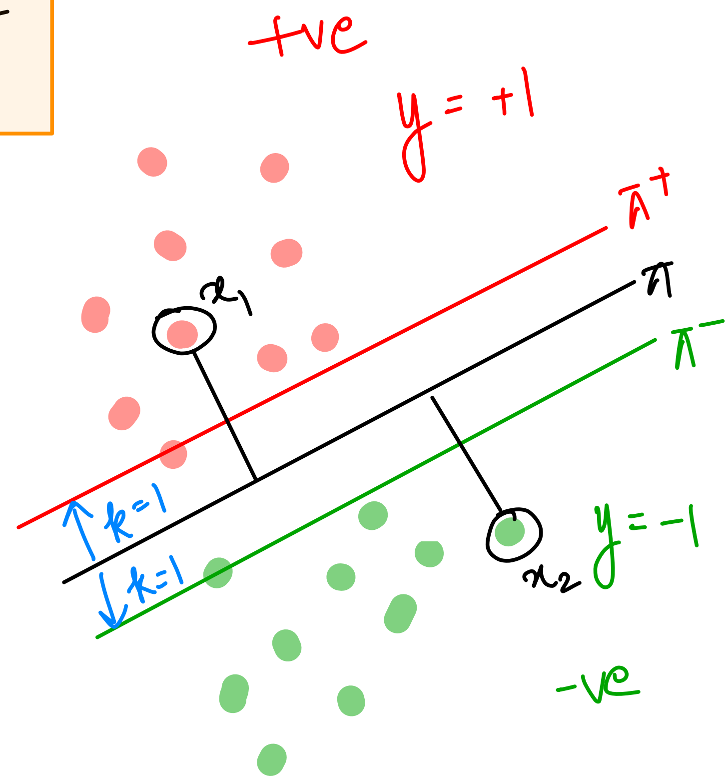
$$d = \frac{2}{\|w\|}$$

$$\boxed{\underset{w}{\arg\max} \frac{2}{\|w\|}} \rightarrow \boxed{\underset{w}{\arg\min} \frac{\|w\|}{2}}$$

$$\underset{w}{\arg\min} \frac{\|w\|}{2} \quad s.t \quad (w^T x + b) y \geq 1$$

x_1 +ve point: $\underbrace{(w^T x + b)}_{+ve} \underbrace{y}_{+ve} \geq 1$

x_2 -ve point: $\underbrace{(w^T x + b)}_{-ve} \underbrace{y}_{-ve} \geq 1$



$x_1 \cdot \underbrace{(w^T x + b)}_{(0.5)} \underbrace{y}_{(+1)} \Rightarrow 0.5$

$1 - 0.5 \quad \varepsilon_1$

$x_2 = \underbrace{(w^T x + b)}_{(-0.5)} \underbrace{y}_{(+1)} \Rightarrow -0.5$

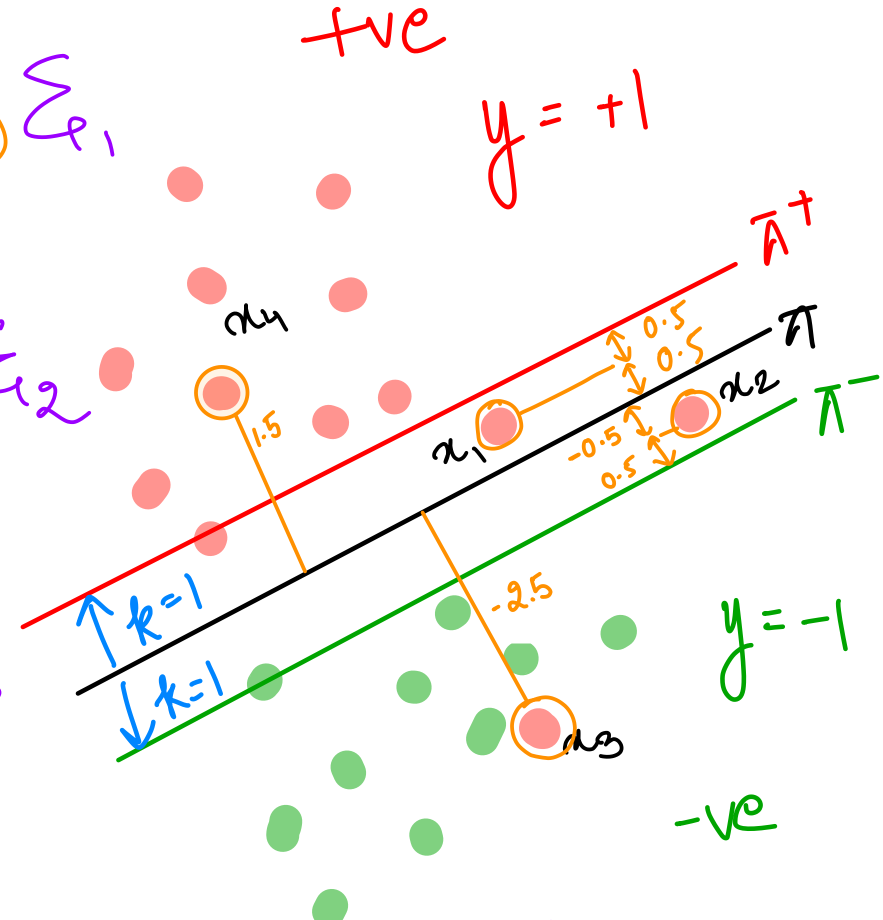
$1 - 1.5 \quad \varepsilon_2$

$x_3 : \underbrace{(w^T x + b)}_{(-2.5)} \underbrace{y}_{(+1)} \Rightarrow -2.5$

$1 - 3.5 \quad \varepsilon_3$

$x_4 \cdot \underbrace{(w^T x + b)}_{(+1.5)} \underbrace{y}_{(+1)} \Rightarrow +1.5$

$1 - (-0.5)$



$\varepsilon_i = 0$ for correctly classified point | $\varepsilon_i = 0$
points lying beyond π^+

Soft Margin Classifier

$$\arg\max_{w} \frac{2}{\|w\|} \Rightarrow \arg\min_{w} \frac{\|w\|}{2}$$

MSE + λ Reg

Regularisation Const

$$\arg\min_{w} \underbrace{\frac{\|w\|}{2}}_{A} + \frac{C}{n} \sum_{i=1}^{n} \underbrace{\epsilon_i}_{B}$$

C → HP

errors

MSE + $\lambda \sum |w|$

$\sum w^2$

A: Underfitting

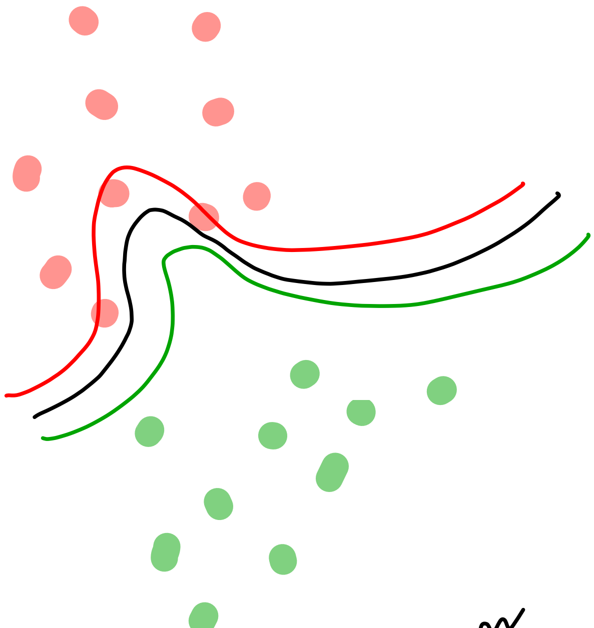
B: Overfitting

MSE: A + λ B: Reg

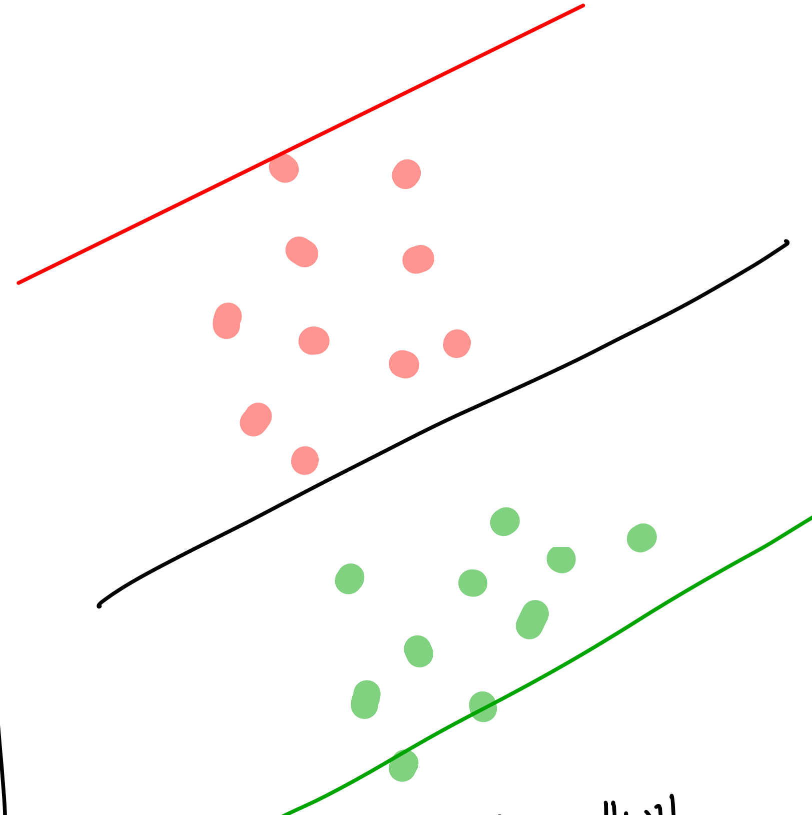
$C = 1/\lambda$

$C\,A + B$

error: A, $\|w\|_2$: B



focus is on Min $\sum_{i=}^{m} \varepsilon_i$



focus is on Min $\frac{\|w\|}{2}$

$$\arg\min_{w} \frac{\|w\|}{2} + \frac{C}{n}\sum_{i=1}^{n} \varepsilon_i \qquad \forall i\ 1 \to n$$

$$\text{s.t } (w^T x + b)\, y \geq 1 - \varepsilon_i$$

(I) $C \uparrow \Rightarrow$ Overfitting

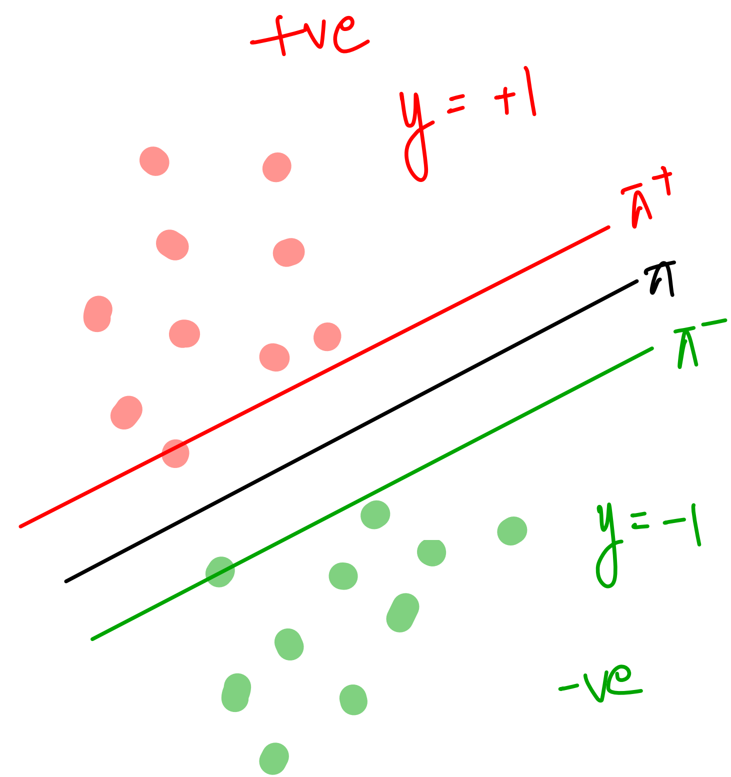
(II) $C \downarrow \Rightarrow$ Underfitting

x_q = Query point

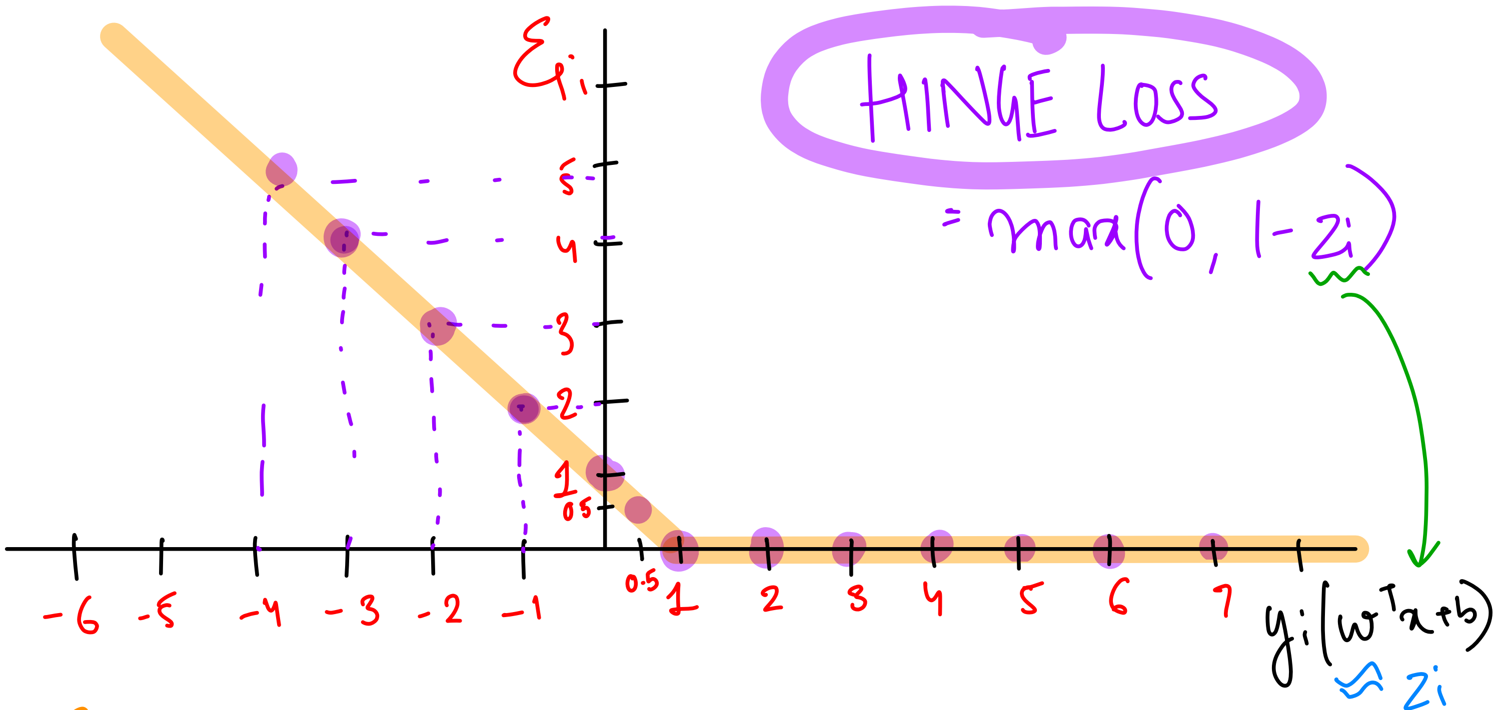
$(w^T x + b) \rightarrow$ +ve sign $\rightarrow$ +ve class

$\downarrow$

-ve signed $\rightarrow$ -ve class.



At inference.



1. $\varepsilon_e = 0$
2. $(w^T x + b) y \geq 1 - \varepsilon_{e_1}$

$$z_i \geq 1 - \varepsilon_{e_1} \Rightarrow \varepsilon_{e_1} \geq 1 - z_i$$

Log Reg $\{0, 1\}$

Loss function $\rightarrow$ LOG LOSS

$$\underbrace{y}_{\text{actual}} \log \underbrace{\hat{y}}_{\text{predicted}} + (1-y) \log (1-\hat{y})$$

<u>LOSS</u>. $\displaystyle\sum_{i=1}^{n} \log\left(1+\exp\left(-y_i\left(w^T x+b\right)\right)\right)$

SVM $\{-1, 1\}$

Loss function : $\rightarrow$ HINGE LOSS

$$\sum_{i=1}^{n} \varepsilon_i \Rightarrow \max(0, 1-z_i)$$

$$z_i \rightarrow (w^T x+b)\, y$$

$$\max\left(0, 1-(w^T x+b)\, y\right)$$

primal form

$$\arg\min_{w} \frac{\|w\|}{2} + \frac{C}{n}\sum_{i=1}^{n} \varepsilon_i \quad \text{st } (w^T x + b) y > 1 - \varepsilon_i$$

$$\forall\, i\ 1 \to n$$

Primal Dual form

dual form

$$\arg\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\text{s.t. } 0 \le \alpha \le C \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

1. $x_i \rightarrow \alpha_i$, earlier f's $\rightarrow$ w's

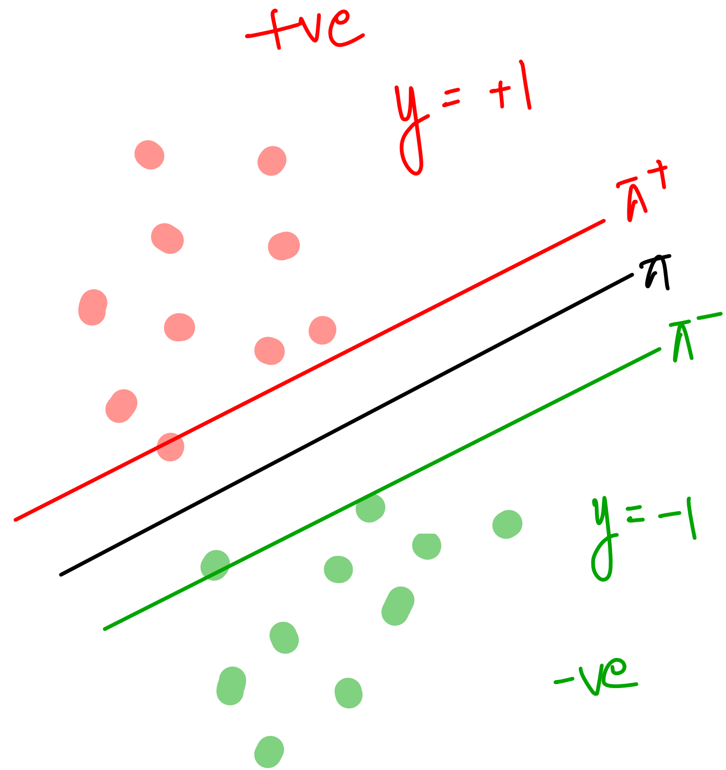
2. All x's occur in the form of $x_i^T x_j$

3. $$f(x_q) = \sum_{i=1}^{n} \alpha_i y_i x_i^T x_q$$

4. $\alpha = 0$ for all non support vectors

   $\alpha > 0$ for all support vectors.

		w_1	w_2	w_3	w_4	w_5	
		f_1	f_2	f_3	f_4	f_5	y
α_1	x_1						
α_2	x_2						
α_3	x_3						
α_4	x_4						
α_5	x_5						
α_6	x_6						
α_7	x_7						



$n \quad \underline{x_i} \Rightarrow O(n^2)$

$$f(x_q) = \sum_{i=1}^{n} \underbrace{\alpha_i}_{\text{learnt during training}} \; \underbrace{y_i}_{\text{actual label}} \; \underbrace{x_i^T}_{\text{datapoint } x_i} \; \underbrace{x_q}_{\text{query point.}}$$

1 Million $\Rightarrow$ 30 ~~Support~~

x_i

1000 $\Rightarrow$ 3 Support vector

x_i

$$f(x_q) = \alpha_1 y_1 x_1^T x_q + \alpha_2 y_2 x_2^T x_q + \alpha_3 y_3 x_3^T x_q$$

Quiz time!

Quiz Ended!

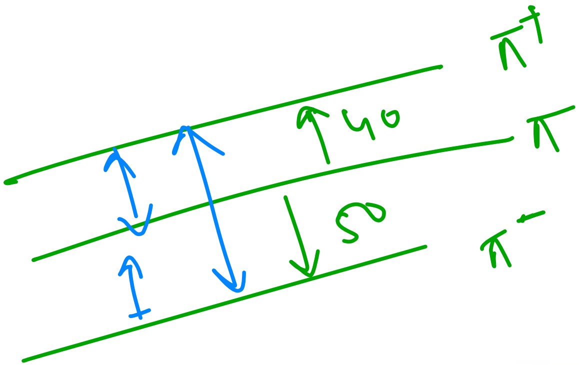
What do you mean by generalization in terms of SVM?

18 users have participated

A	How far the hyperplane is from the training datapoints	33%
B	How accurately the SVM can predict outcomes for unseen data	50%
C	How accurately the SVM classifies training datapoints	17%

Quiz time!

Quiz Ended!



If,

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

then margin will be:

$$w^T x + b = k$$

$$w^T x + b = -k$$

90

$$\frac{90}{\|w\|}$$

22 users have participated

A	10/|w|	32%
B	40/|w|	5%
C	50/|w|	9%
D	90/|w|	55%

Quiz time!

Quiz Ended!

What do you mean by a hard margin?

22 users have participated

A	The SVM allows no error in classification.	82%
B	The SVM allows some error in classification.	5%
C	The SVM allows high error in classification.	14%

Quiz time!

Quiz Ended!

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 ξi?

21 users have participated

A	0.8, -2.0, 0.0	43%
B ✓	0.2, 3.0, 1.0	29%
C	0.8, 2.0, 0.0	29%