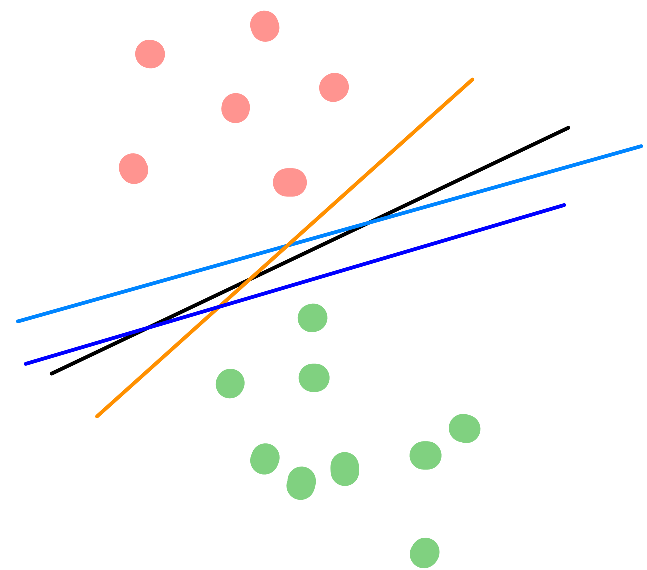
SVM-1

Support Vector Machines (SVM)

* Theoretically speaking
  most powerful algorithm

* Practically → Not used frequently
  these day



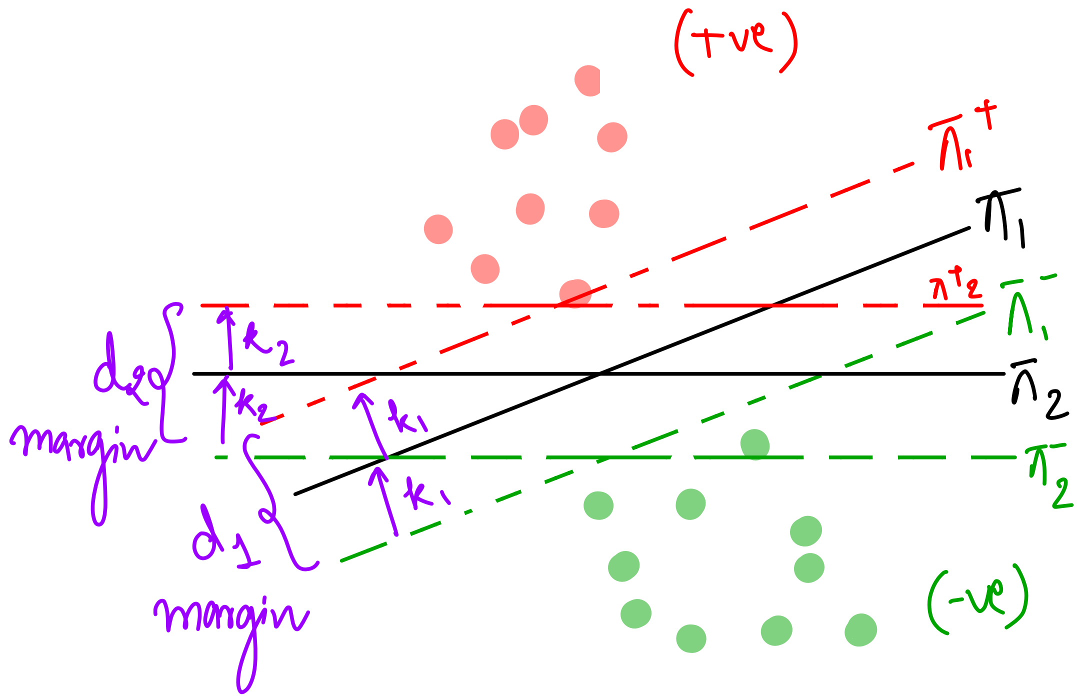
* Mathematically very heavy

* 90's, 2000's.

Q which π to choose?

A Choose π which has highest margin

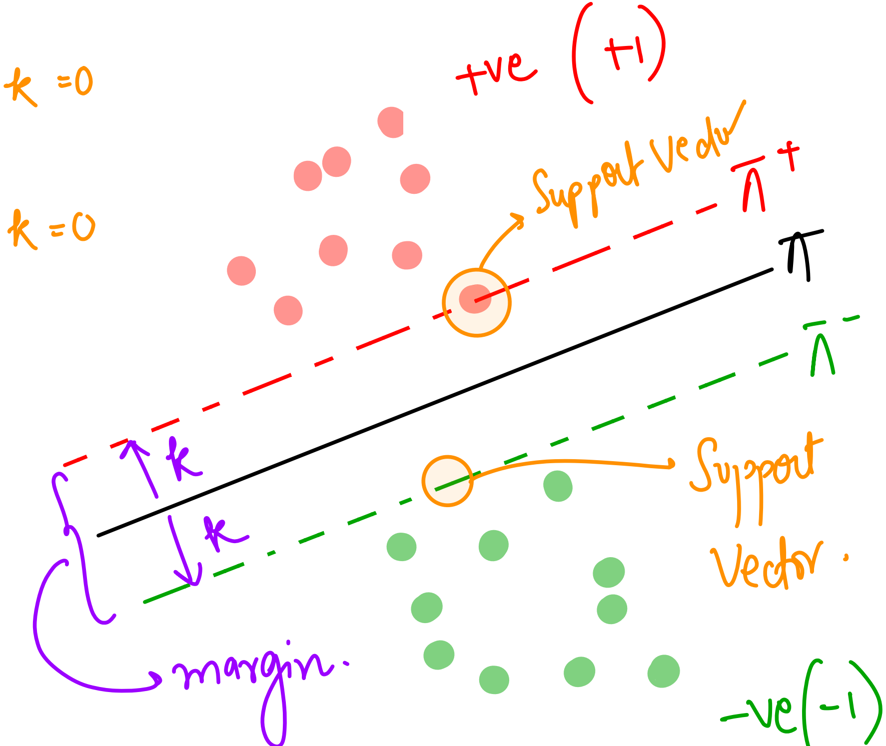
MARGIN MAXIMISING CLASSIFIER



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

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

$$\pi^{-1} : w^T x + b = -k \qquad w^T x + b + k = 0$$



margin $\text{dist}(\pi^{+}, \pi^{-})$

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

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

$d = \text{margin} = \dfrac{2k}{\|w\|} \Rightarrow k = \text{Constt.}$

Case 1

$k = 1$

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

$$\underset{w}{\operatorname{arg\,max}} \left(\frac{2}{\|w\|} \right)$$

Case 2

$k = 10$

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

$$\underset{w}{\operatorname{arg\,max}} \left(\frac{20}{\|w\|} \right)$$

By default, we take $k = 1$ to simplify mathematical calculation.

<u>Goal</u>: maximise margin $= \dfrac{2}{\|w\|}$

Such that all the points are correctly classified.

This is known as HARD MARGIN CLASSIFIER

<u>Goal:</u> Maximise Margin s.t. all pts are correctly classified.

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

$$\Rightarrow \boxed{\underset{w}{\operatorname{argmax}} \frac{2}{\|w\|}} \Rightarrow \boxed{\underset{w}{\operatorname{argmin}} \frac{\|w\|}{2}}$$

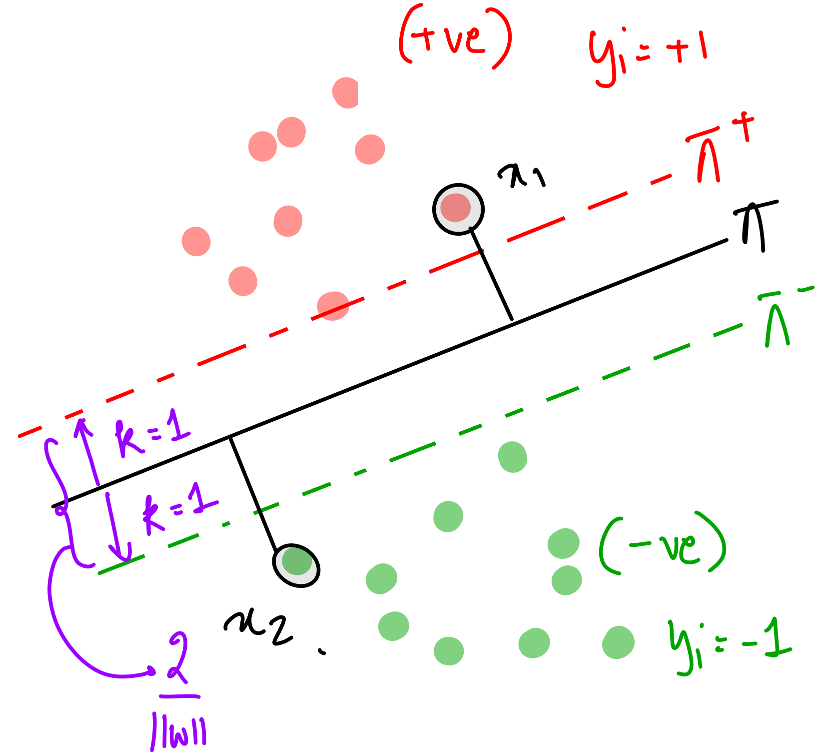
$$\boxed{\underset{w}{\operatorname{argmin}} \frac{\|w\|}{2} \quad \text{s.t} \quad (w^T x + b) y_i \geq 1}$$

$$\underset{w}{\arg\min}\ \frac{\|w\|}{2} \quad s\ t \quad (w^Tx+b)\,y_i \ge 1$$

(x_1) $\quad \underbrace{\overbrace{(w^Tx+b)}^{\text{distance}}}_{+ve}\ \underbrace{y_i}_{+ve} \ge 1$

+ve point

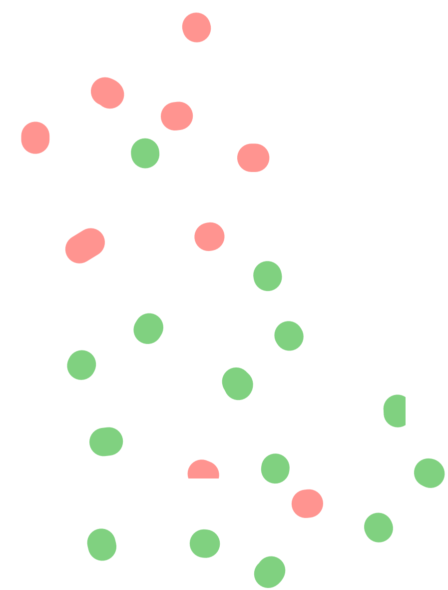
(x_2) $\quad \underbrace{(w^Tx+b)}_{-ve}\ \underbrace{y_i}_{-ve} \ge 1$



Q When will HARD MARGIN CLASSIFIER fail?

A When the data is not linearly/Clearly seperable, hard margin classifier fails

Lets allow some misclassification.



$x_1 \cdot \ (w^T x + b)\, y_i \Rightarrow 0.5$

$(+0.5)\ (+1) \Rightarrow 1 - 0.5 \quad \varepsilon_1$

$x_2 : \ (w^T x + b)\, y_i \Rightarrow -0.5$

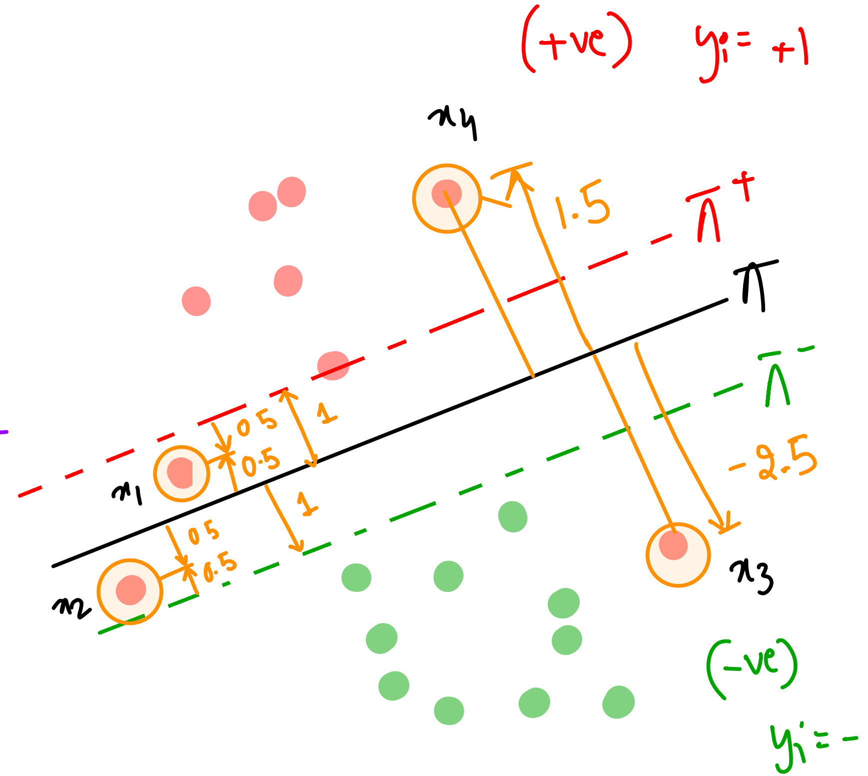
$(-0.5)\ (+1) \Rightarrow 1 - 1.5 \quad \varepsilon_2$

$x_3 : \ (w^T x + b)\, y_i \Rightarrow -2.5$

$(-2.5)\ (+1) \Rightarrow 1 - 3.5 \quad \varepsilon_3$

$x_4 \cdot \ (w^T x + b)\, y_i \Rightarrow +1.5$

$(+1.5)\ (+1) \rightarrow 1 - (-0.5) \quad \varepsilon_4 = 0$



for correctly classified point : data points above π^+

Soft margin Classifier

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

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

HP

Errors.

$C \uparrow$ Overfit

$C \downarrow$ Underfit

$$\underbrace{MSE}_{\text{over}} + \lambda \underbrace{Reg}_{\text{under}}$$

Regularisation Constt.

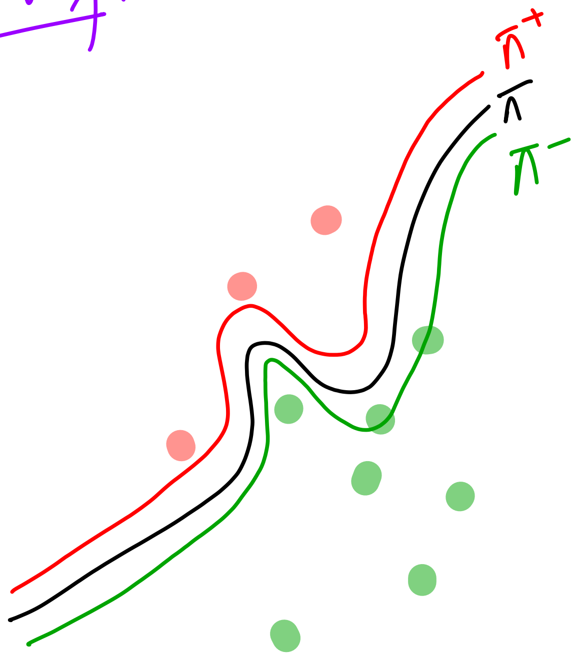
$$C = \frac{1}{\lambda}$$

Loss Reg

$$\underbrace{1A}_{MSE} + \underbrace{\lambda B}_{Reg}$$

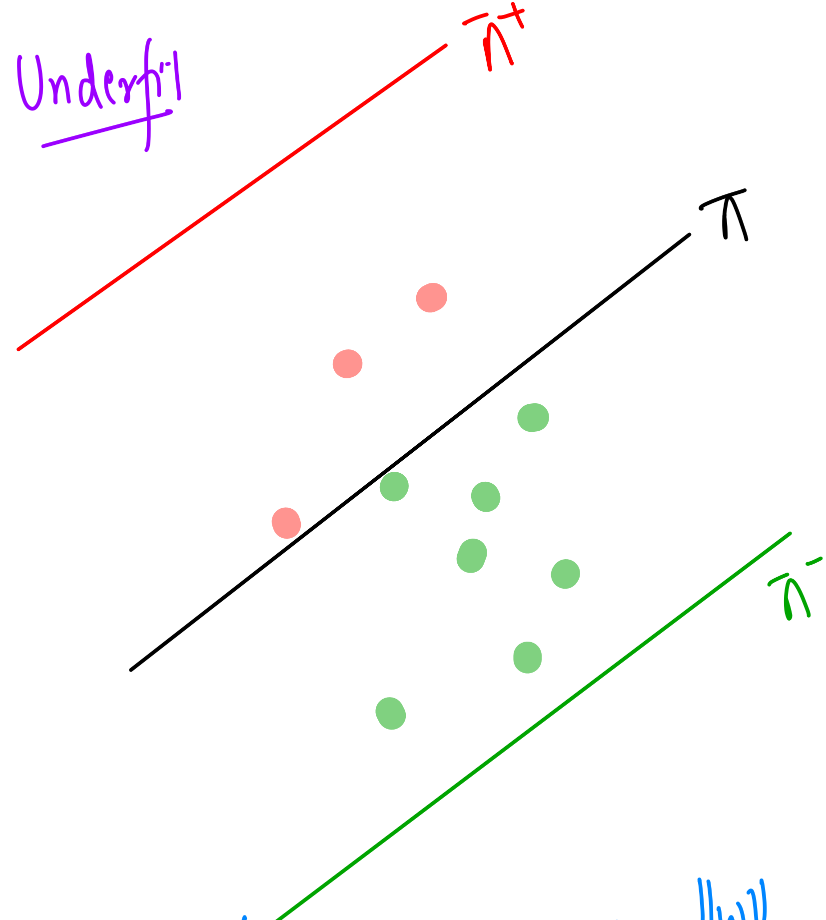
$$C \underbrace{A}_{\sum \epsilon_i} + 1 \underbrace{B}_{\|w\|/2}$$

Overfit



Only focus on $\min \sum \varepsilon_i$

Underfit



Only focussing on $\min \frac{\|w\|}{2}$

$$\underset{w}{\operatorname{argmin}} \quad \frac{\|w\|}{2} + \frac{C}{n}\sum_{i=1}^{n} \xi_i \quad \text{s t} \quad (w^T x + b) y_i \geq 1 - \xi_i$$

$$\forall \, 1 \rightarrow n$$

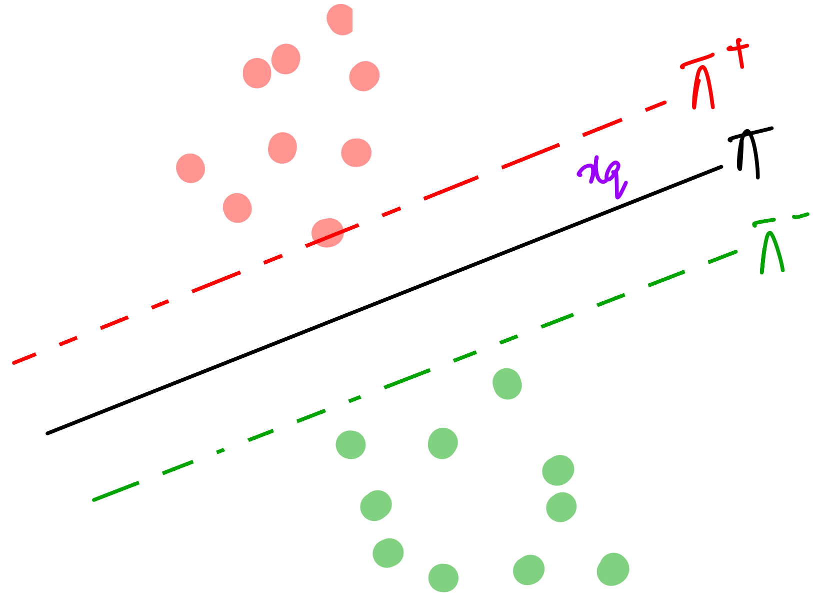
$x_q \Rightarrow$ query point

$$\underbrace{(w^T x + b)}$$

$\longrightarrow$ +ve value $\Rightarrow$ +ve class

$\downarrow$

-ve value $\Rightarrow$ -ve class .



1. Hinge loss
2. Primal Dual
3. Kernel → poly + radial
4. Code

next class