Generalized Discriminant f^n

① Linear case -

$$g(x) = w^T x + b_{x1}$$

$$A = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \\ b \end{bmatrix}, \quad y = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \\ 1 \end{bmatrix}$$

$$g(x) = A^T y$$

② Non linear case

$$y = \phi(x) \quad \text{Transform into higher dim space.}$$

$$x^2 = \begin{bmatrix} x_0^2 \\ x_1^2 \\ \vdots \\ x_n^2 \\ 1^2 \end{bmatrix}$$

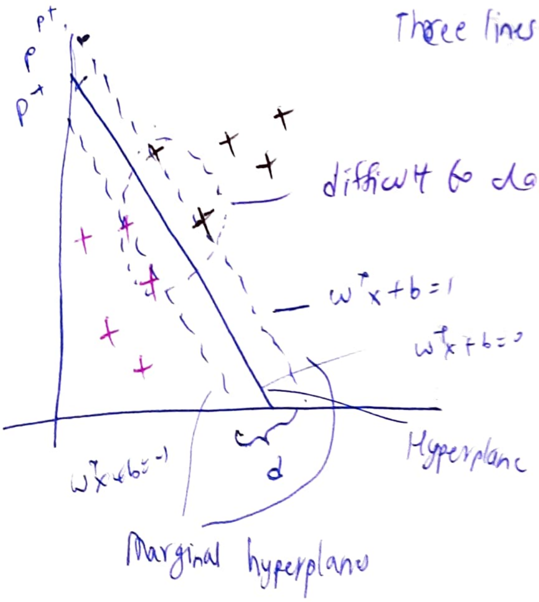
$$g(x) = A^T Y$$

why this ?

In some higher dim space, non linear data will be linear.

Derivation of SVM

Three lines are ||el



① Eqns are chosen for simplicity
② Equidistant 1 & -1

For any brown pt, $w^Tx + b \geq 1$, $y = 1$
pink pt, $w^Tx + b \leq -1$, $y = -1$

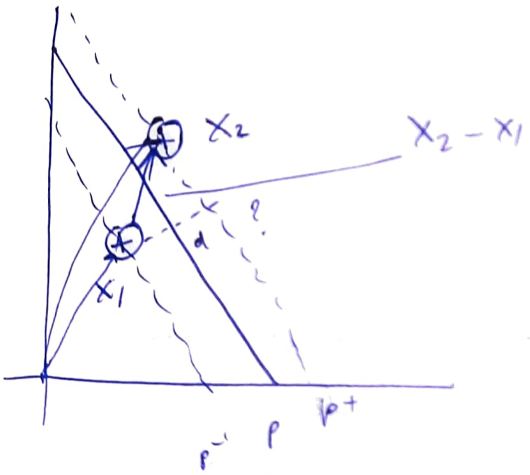
$\Rightarrow$ multiply with y,

$$y(w^Tx + b) \geq 1$$

equal on multiplying with $y = 1$, $y = -1$.

For the support vectors,

$$y(w^Tx + b) = 1$$



Project $(x_2 - x_1)$ along dotted to find d.
Dot product

Vector w is $\perp$ to all of them.

$$d = (x_2 - x_1) \cdot \frac{w}{\|w\|}$$

$$= \frac{x_2^T w - x_1^T w}{\|w\|} \quad - ①$$

Now for x_2,

$y(w^Tx + b) = 1$, $y = 1$

$w^Tx + b = 1 \Rightarrow w^Tx_2 = 1 - b \quad - ②$

For x_1, $y = -1$

$w^Tx_1 = -1 - b \quad - ③$

$$(AB)^T = B^T A^T \quad \text{--- (4)}$$

Eqⁿs (2) & (3), transposing similar to 4

$$d = \frac{(1-b)-(-1+b)}{\|w\|} = \frac{2}{\|w\|}$$

svm optimization - Hard margin (strict constraint)

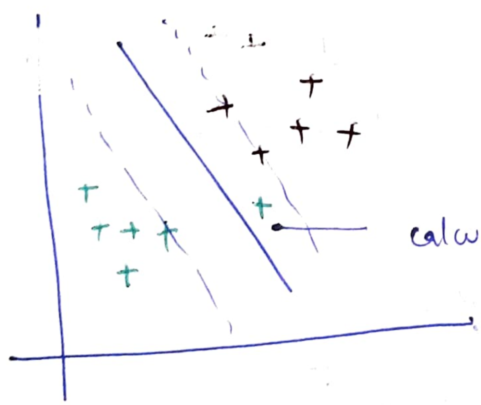
$$\max \frac{2}{\|w\|} \rightarrow \min \frac{\|w\|}{2} \quad \left(\max x = \min \frac{1}{x}\right)$$

$$\text{s.t. } y(wx+b) \geq 1$$

sol^n exists if

1. Linearly separable
2. No pts. in margin region

soft margin



calculate constraint for this pt.

$y = -1$ is o/p for green pt. on other side.

$w^T x + b$ is +ve.

$$y(w^T x + b)$$

$$= -ve \times +ve$$

$$= -ve < 1$$

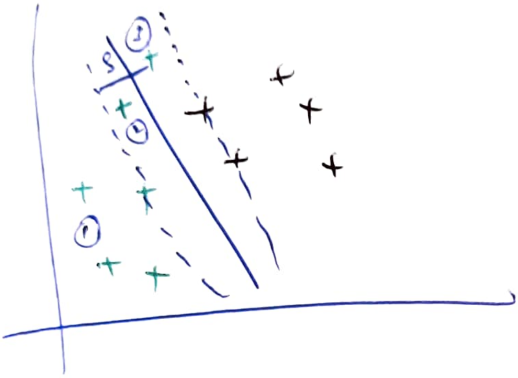
∴ constraint fails.

Actual constraint should be ≥ 1

Allow some slack.

$$\rightarrow y(w^T x + b) \geq 1 - \xi$$

$$\xi = \begin{cases} 0 & \text{correctly classified} \quad (1) \\ [0, 1] & \text{, correct class, but penalty} \quad (2) \\ > 1 & \text{, misclassified.} \quad (3) \end{cases}$$



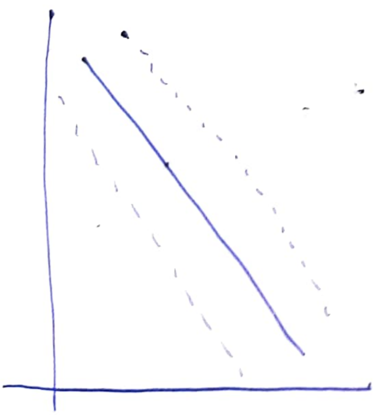
ξ is dist. from marginal plane of the ground truth class.

$$\xi \geq 1 - y(w^T x + b)$$

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

$$L_1 \text{ svm} = \min \frac{\|w\|}{2} + \underbrace{c}_{\text{Hyper param}} \underbrace{\sum_{i=1}^{n} |\xi|}_{\substack{\text{classification error} \\ \text{(Hinge loss)}}}$$

$$L_2 \text{ svm} : \min \frac{\|w\|}{2} + c \sum_{i=1}^{n} \xi^2$$



$$\min \left(\text{margin error} \right) + \underbrace{c}_{\text{controls the importance}} \left(\text{classification error} \right)$$

Hence, called as "soft margin".