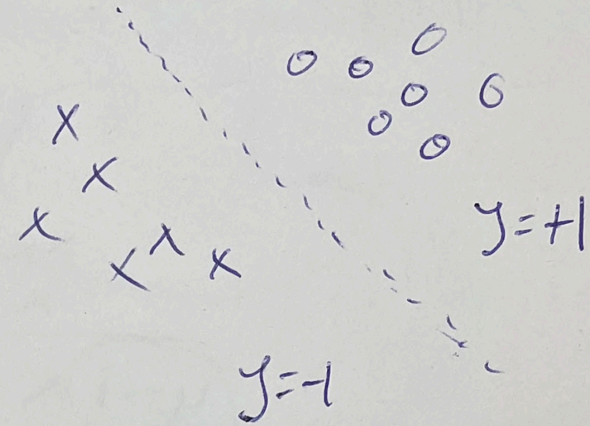
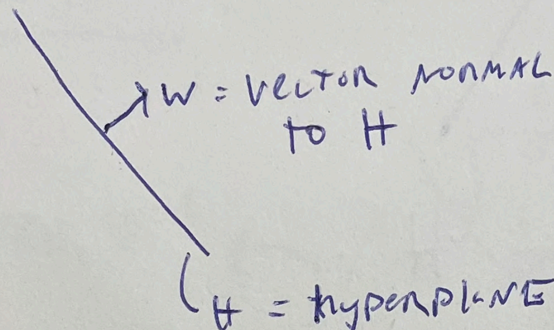
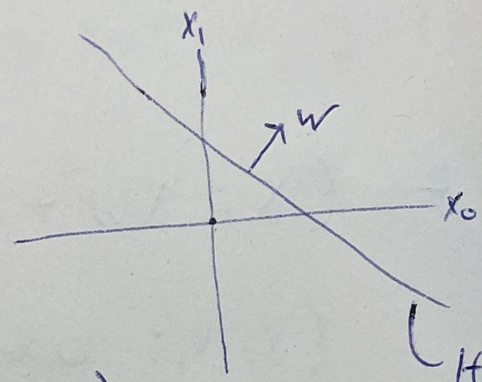
DECISION BOUNDARY CLASSIFICATION



HYPERPLANE:





$$H : \{ X : XW^T + b = 0 \}$$

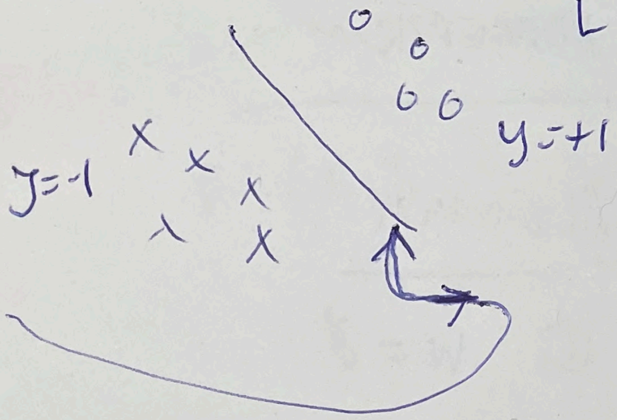
↳ $\{ x_0, x_1 : w_0 x_0 + w_1 x_1 + b = 0 \}$

ORIGINAL SPACE	NEW SPACE
$X : [x_0, x_1]$	$X' : [1 \; x_0 \; x_1]$
$W : [w_0, w_1]$	$W' : [b \; w_0 \; w_1]$

$$XW^T + b = X'W'^T$$

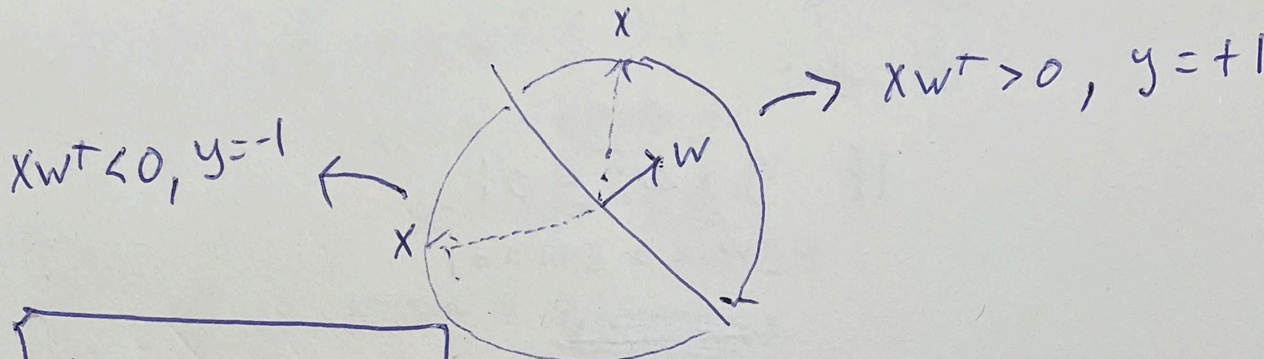
↳ TRICK TO "GET RID OF" THE BIAS b.

PERCEPTRON



Goal is to LEARN H

DECISION RULE:



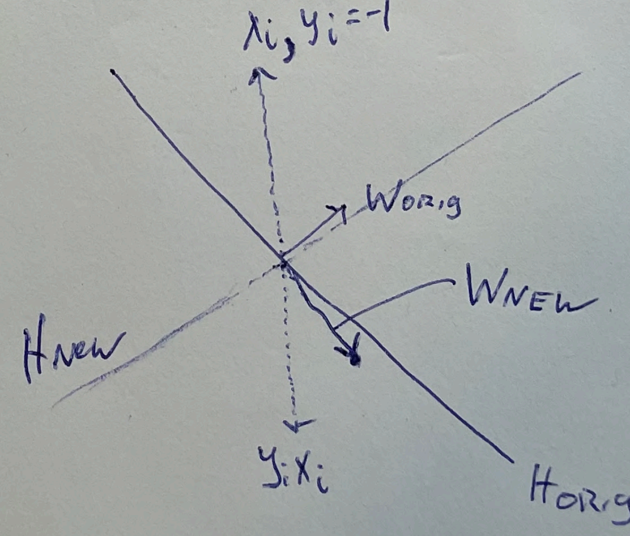
$$\boxed{\hat{y} = \text{sign}(XW^T)}$$

ERROR FUNCTION: ~~ERROR FUNCTION~~

$$\text{ERROR}(y_i, X_i; w) = \begin{cases} 1 & \text{if } y_i X_i W^T \leq 0 \\ 0 & \text{ELSE} \end{cases}$$

UPDATE RULE:

$$\boxed{W = W + y_i X_i}$$



PERCEPTRON ALGORITHM

$D = \{x_1, y_1 \dots x_m, y_m\}$

1. $\vec{w} = \vec{\emptyset}$
2. while TRUE:
   - (a) ERROR = 0
   - (b) $\forall i$ do:

     If $y_i x_i w^T \le 0$:

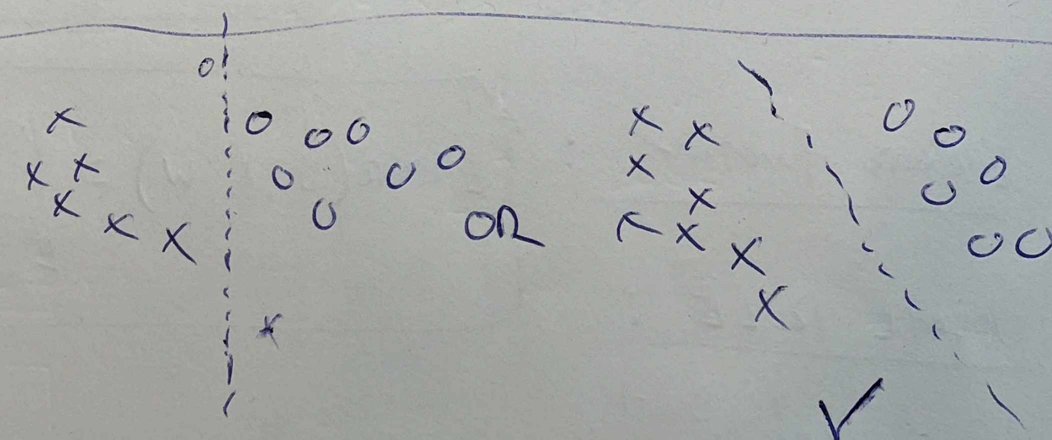
     ERROR = ERROR + 1

     $w = w + y_i x_i$
   - (c) If ERROR = 0:

     Break;

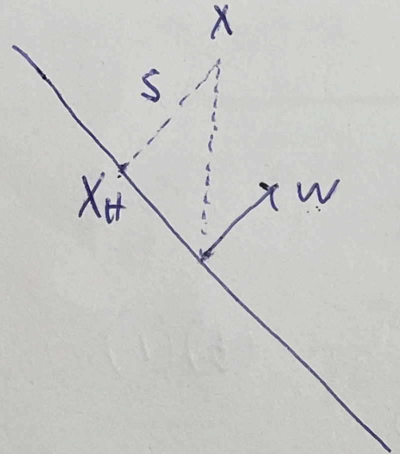
WEAKNESS OF THE PERCEPTRON ALGORITHM

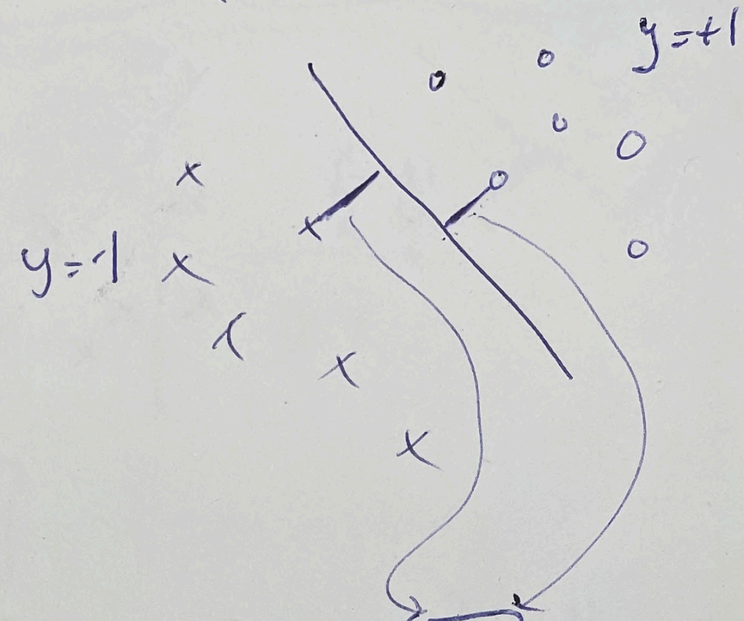
↳



SUPPORT VECTOR MACHINE (SVM)

- SVM IS A MAXIMUM MARGIN CLASSIFIER.





MARGIN = $\boxed{S}$

① $X_H = X - S$

$\rightarrow X_H W^T = 0$ BECAUSE X_H IS ON H.

$\rightarrow (X-S)W^T = 0$

② $S = \alpha W$

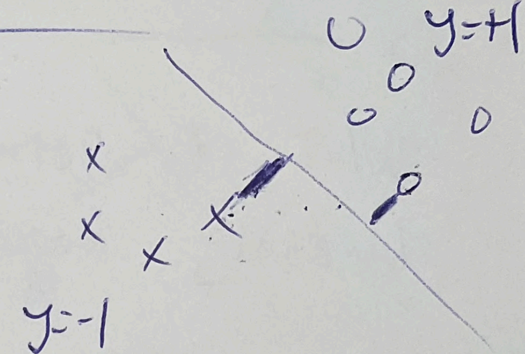
$\rightarrow \quad (X - \alpha W) W^T = 0$

$\rightarrow \quad \alpha = \dfrac{XW^T}{WW^T} = \dfrac{XW^T}{\|W\|_2^2}$

③ $\|S\|_2 = \|\alpha W\|_2 = \alpha \|W\|_2 = \dfrac{XW^T}{\|W\|_2^2} \|W\|_2$

$$\boxed{\|S\| = \frac{XW^T}{\|W\|_2}}$$

SVM - STARTING OBJECTIVE



FROM LAST PAGE: $\boxed{\|s\|_2 = \frac{XW^T}{\|W\|_2}}$

STARTING OBJECTIVE:

$$\max_{W} \left[\min_{i} \frac{X_i W^T}{\|W\|_2} \right]$$

PROBLEM #1: THIS APPLIES A DIFFERENT RULE DEPENDING ON THE SIGN OF $X_i W^T$.

SOLUTION: $$\max_{W} \left[\min_{i} \frac{y_i X_i W^T}{\|W\|_2} \right]$$

↳ THIS RESTRICTS W TO BE A SEPARATING HYPERPLANE! #PERCEPTRON

PROBLEM 2: THIS DOESN'T YIELD UNIQUE SOLUTION FOR W.

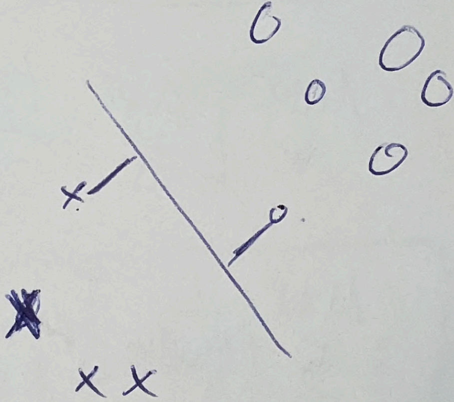
EXAMPLE:

$$\left.\begin{array}{l} X=[1,1] \\ W=[1,1] \end{array}\right] \rightarrow OBJ = \frac{2}{\sqrt{2}} = \sqrt{2}$$

$$\left.\begin{array}{l} X=[1,1] \\ W=[2,2] \end{array}\right] \rightarrow OBJ = \frac{4}{\sqrt{8}} = \sqrt{\frac{16}{8}} = \sqrt{2}$$

SVM CONT...

$$\max_{w}\left[\min_{i} \frac{y_i x_i w^T}{\|w\|_2}\right]$$



↳ $$\max_{w} \frac{1}{\|w\|_2}\left[\min_{i} y_i x_i w^T\right]$$

↳ $$\max_{w} \frac{1}{\|w\|_2}\left[1\right] \quad \text{s.t.} \quad \min_{i} y_i x_i w^T = 1$$

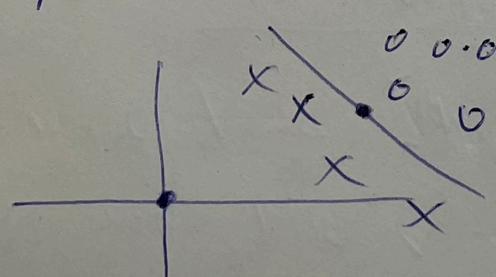
↳ $$\max_{w} \frac{1}{\|w\|_2} \quad \text{s.t.} \quad \forall_i \; y_i x_i w^T \geq 1$$

↳ $$\min_{w} \|w\|_2 \quad \text{s.t.}$$

↳ $$\boxed{\min_{w} \|w\|_2^2 \quad \text{s.t.} \quad \forall_i \; y_i x_i w^T \geq 1}$$

↳ $\boxed{\text{Quadratic Program}}$

↳ But wait!!!



SVM CONT...

WE NEED TO TAKE BIAS BACK OUT OF W!

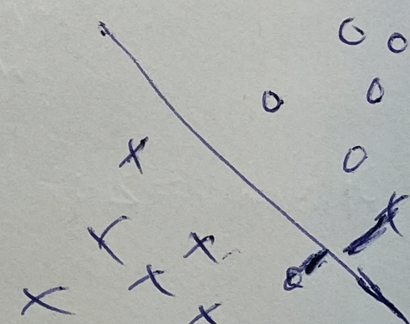
$$\boxed{\max_{W,b} \|W\|_2^2 \quad \text{S.T. } \forall_i \; y_i(x_i W^T + b) \geq 1}$$

FINAL FORM OF QP THAT WE ARE INTERESTED IN FOR LINEAR TEXT CLASSIFICATION!

$$\min_{W,b} \|W\|_2^2 + C\sum_{i=1}^{M} \epsilon_i$$

$$\text{ST. } \forall_i \; y_i(x_i W^T + b) \geq 1 - \epsilon_i$$

$$\epsilon_i = \begin{cases} 1 - y_i(x_i W^T + b) & \text{if } y_i(x_i W^T + b) < 1 \\ 0 & \text{ELSE} \end{cases}$$



SVM W/ SOFT CONSTRAINTS TO HANDLE NON-LINEARLY SEPERABLE DATA.

Slightly