Supervised

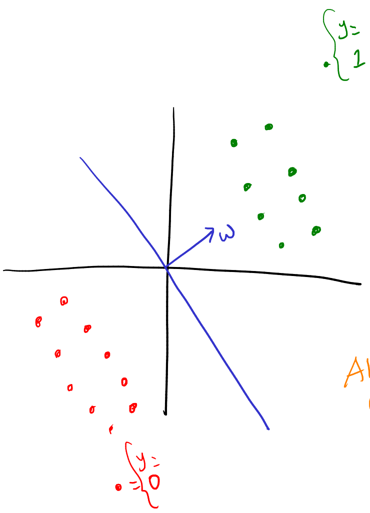
Binary classification

Generative
- NAIVE BAYES
- Gaussian discriminant Analysis

Discriminative.
- K-NN
- Decision trees

Discriminative models for classification

How to model $P(y=1/x)$?

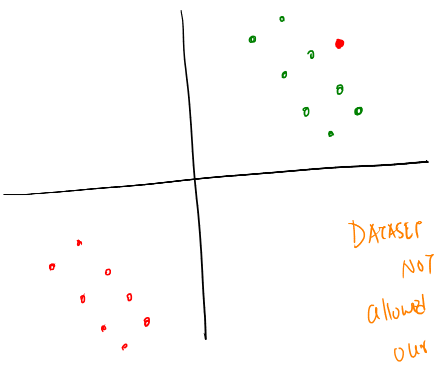


ALLOWED under our model!

SIMPLEST ASSUMPTION

$$P(y=1/x) = \begin{cases} 1 & \text{if } w^T x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

LINEAR SEPARABILITY ASSUMPTION



DATASET is NOT allowed under our model.

Goal.

$$\min_{h \in \mathcal{H}} \sum_{i=1}^{n} \mathbb{1}(h(x_i) \neq y_i)$$

NP - HARD for a general dataset even if $\mathcal{H}$ is just linear hypotheses.

- How about with extra "Linear separability" assumption.

LINEAR SEPARABILITY ASSUMPTION

$$\exists w \in \mathbb{R}^d \quad s.t \quad \text{sign}(w^T x_i) = y_i \quad \forall i \in [n]$$

PERCEPTRON [Rosenblatt, 1950's]

Input: $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ $\quad x_i \in \mathbb{R}^d$, $\quad y_i \in \{+1, -1\}$

ITERATION

$w^0 = 0 \in \mathbb{R}^d \quad [0 \; 0 \cdots 0]$

until convergence

- Pick (x_i, y_i) pair from the dataset
- IF $\text{sign}(w^{t^T} x_i) = y_i$

  do nothing

  else

  $\boxed{w^{t+1} = w^t + x_i y_i}$ ← UPDATE RULE.

  $x_i \in \mathbb{R}^d$, $y_i \in \{\pm 1\}$

  end.

end.

UPDATE RULE

if mistake $\quad w^{t+1} = w^t + x_i y_i$

MISTAKE TYPE 1

Predicted $= 1$ ← $(w^{t^T} x_i) \geq 0$

Actual $= -1$ ← $y_i = -1$

$$w^{t+1} = w^t + x_i y_i$$

$$(w^{t+1})^T x_i = (w^t + x_i y_i)^T x_i$$

$$= \underbrace{w^{t^T} x_i}_{\geq 0} + \underbrace{\underbrace{y_i}_{-1} \underbrace{\|x_i\|^2}_{> 0}}_{\text{Negative}}$$

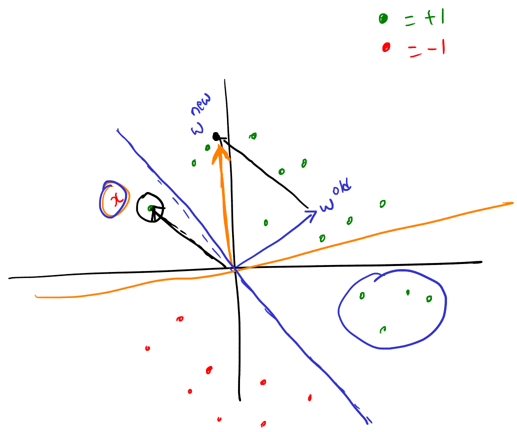
MISTAKE TYPE 2

Pred $= -1$ ← $(w^{t^T} x_i) < 0$

Act $= +1$ ← $y_i = +1$

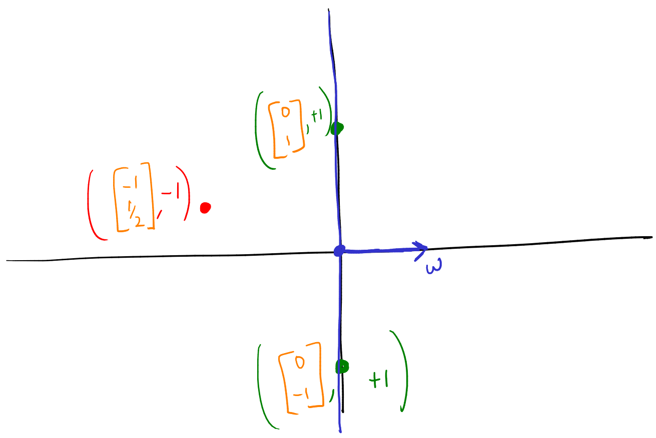
$$(w^{t+1})^T x_i = \underbrace{w^{t^T} x_i}_{< 0} + \underbrace{\underbrace{y_i}_{+1} \|x_i\|^2}_{> 0}$$

⇒ update rule pushes w in the "right" direction for x_i



$$w^{new} = w^{old} + x \cdot \underset{\sim}{(1)}_{y}$$

- Fixing w for one x might affect decision for other data points.
- So need more careful arguement for convergence.

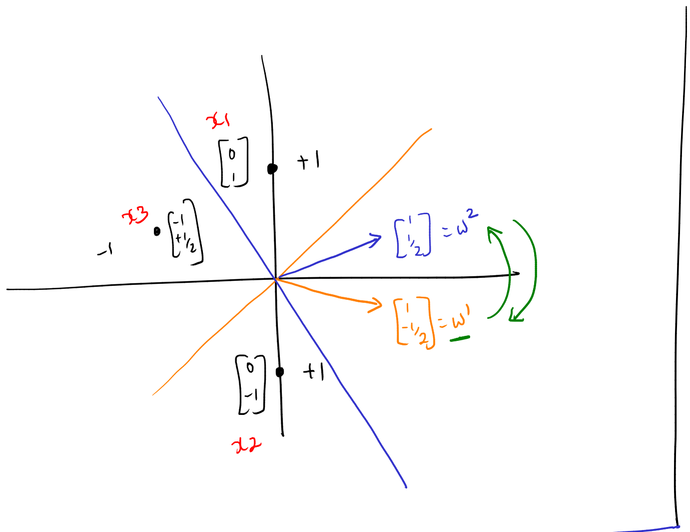


Is this a Linearly separable dataset?

$$\left\{ \begin{bmatrix}0\\1\end{bmatrix}, +1 \quad \begin{bmatrix}0\\-1\end{bmatrix}, +1 \quad \begin{bmatrix}-1\\1/2\end{bmatrix}, -1 \right\}$$

Is there a $w \in \mathbb{R}^2$ s.t. $w^T x_i \geq 0 \Rightarrow y_i = +1$

$w^T x_i < 0 \Rightarrow y_i = -1$



PERCEPTRON

$$w^0 = \begin{bmatrix}0 & 0\end{bmatrix}$$

$w^{0T} x_1 = 0$; $w^{0T} x_2 = 0$; $\boxed{w^{0T} x_3 = 0 \quad \hat{y}_3 = +1}$

$\hat{y}_1 = +1$ $\quad \hat{y}_2 = +1$

$$w^1 = w^0 + x_3 y_3 = \begin{bmatrix}0\\0\end{bmatrix} + \begin{bmatrix}-1\\1/2\end{bmatrix} \times -1$$

$$= \begin{bmatrix}1\\-1/2\end{bmatrix}$$

$$w^2 = w^1 + x_1 y_1 = \begin{bmatrix}1\\-1/2\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} \cdot 1$$

$$w^2 = \begin{bmatrix}1\\1/2\end{bmatrix}$$

$\leftarrow$

$$w^3 = w^2 + x_2 y_2 = \begin{bmatrix}1\\1/2\end{bmatrix} + \begin{bmatrix}0\\-1\end{bmatrix} \times 1$$

$$w^3 = \begin{bmatrix}1\\-1/2\end{bmatrix} = w^1$$

ASSUMPTION

① $\boxed{\text{LINEAR SEPARABILITY with } \gamma\text{-MARGIN}}$

A Dataset $\{(x_1,y_1), \dots (x_n,y_n)\}$

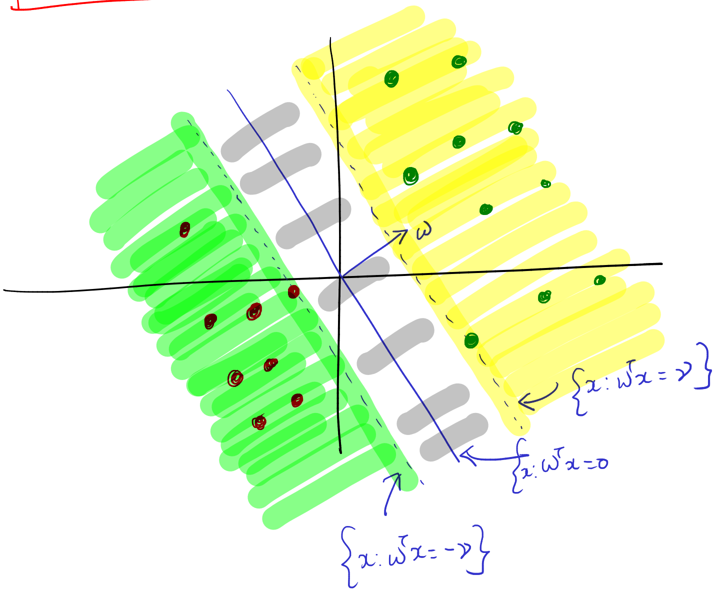
is Linearly Separable

with γ-margin

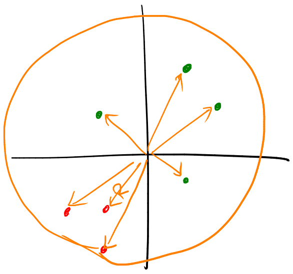
if $\exists\, w^* \in \mathbb{R}^d$ s.t

$({w^*}^T x_i)\, y_i \geq \gamma \quad \forall i$

for some $\gamma > 0$

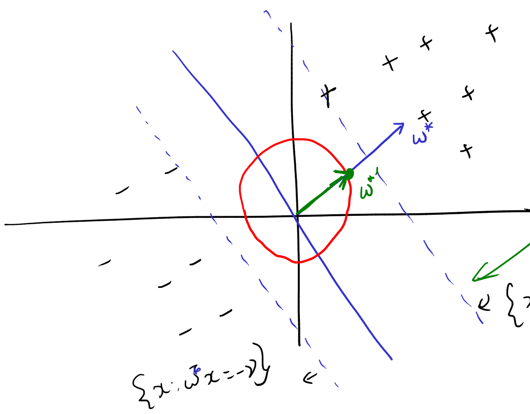


② RADIUS ASSUMPTION



$\forall i \in D \quad \|x_i\|_2 \leq R \quad$ for some $R > 0$.

③ Without loss of generality, $\boxed{\text{assume } \|w^*\| = 1}$



$\{x : {w^{*\prime}}^T x = \gamma'\}$

$\{x : {w^*}^T x = \gamma\}$

$\frac{{w^*}^T x}{\|w^*\|} = \frac{\gamma}{\|w^*\|}$

$\gamma' = \frac{\gamma}{\|w^*\|}$

$\underbrace{\left(\frac{w^*}{\|w^*\|}\right)^T}\, x = \gamma'$

$({w^*}')^T x = \gamma'$

ANALYSIS OF "MISTAKES" OF PERCEPTRON

- Observe that an update happens only when a mistake occurs.

- Say w^{ℓ} is the current guess and a mistake happens w.r.t (x, y)

$$w^{\ell+1} = w^{\ell} + x \cdot y$$

$$\|w^{\ell+1}\|^2 = \|w^{\ell} + x \cdot y\|^2$$

$$= (w^{\ell} + x \cdot y)^T (w^{\ell} + x \cdot y)$$

$$= \|w^{\ell}\|^2 + \underbrace{2 \cdot (w^{\ell^T} x) y}_{\leq 0} + \underbrace{\|x\|^2 \cdot \underbrace{y^2}_{=+1}}_{\leq R^2}$$

because mistake.

$$\|w^{\ell+1}\|^2 \leq \underbrace{\|w^{\ell}\|^2}_{\downarrow} + R^2$$

$$\|w^{\ell+1}\|^2 \leq \left(\|w^{\ell-1}\|^2 + R^2\right) + R^2$$

$$\vdots$$

$$\|w^{\ell+1}\|^2 \leq \underbrace{\|w^{0}\|^2}_{=0} + \ell R^2$$

mistakes

$$\Rightarrow \boxed{\|w^{\ell+1}\|^2 \leq \ell R^2} \qquad (1)$$

$$(w^{\ell+1})^T w^* = (w^{\ell} + x \cdot y)^T w^*$$

$$= w^{\ell^T} w^* + \underbrace{(w^{*^T} x) y}_{\geq \gamma}$$

$$(w^{\ell+1})^T w^* \geq w^{\ell^T} w^* + \gamma$$

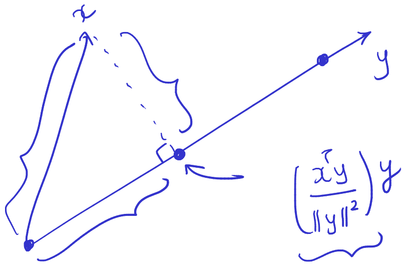
$$\geq \left(w^{\ell-1^T} w^* + \gamma\right) + \gamma$$

$\vdots$

$$(w^{\ell+1})^T w^* \geq \underbrace{w^{0^T} w^*}_{0} + \ell\gamma$$

$$\Rightarrow \boxed{(w^{\ell+1})^T w^* \geq \ell\gamma} \quad \text{— (2).}$$

For any x, y



$$\underbrace{\left(\frac{x^T y}{\|y\|^2}\right) y}$$

$$\left\| \left(\frac{x^T y}{\|y\|^2}\right) y \right\|^2 \leq \|x\|^2$$

[Pythagoras]

$$(x^T y)^2 \frac{\|y\|^2}{\|y\|^4} \leq \|x\|^2$$

Cauchy-Schwartz $\Rightarrow \boxed{(x^T y)^2 \leq \|x\|^2 \|y\|^2}$

From (2)

$$\ell\gamma \leq (w^{\ell+1})^T w^*$$

$$\Rightarrow \ell^2\gamma^2 \leq \left((\underbrace{w^{\ell+1}}_{x})^T \underbrace{w^*}_{y}\right)^2 \leq \|w^{\ell+1}\|^2 \underbrace{\|w^*\|^2}_{1}$$

[From C.S]

$$\Rightarrow \boxed{\|w^{\ell+1}\|^2 \geq \ell^2\gamma^2} \quad \text{— (3).}$$

$$\ell^2\gamma^2 \underset{\uparrow \text{ From (3)}}{\leq} \|w^{\ell+1}\|^2 \underset{\uparrow \text{ From (1)}}{\leq} \ell R^2$$

$$\Rightarrow \ell^2\gamma^2 \leq \ell R^2$$

$\ell \leftarrow$ # mistakes.

$$\Rightarrow \boxed{\ell \leq R^2/\gamma^2} \leftarrow$$ RADIUS MARGIN BOUND

$\Rightarrow$ # mistakes is bounded [because $\gamma > 0$].

$\Rightarrow$ PERCEPTRON CONVERGES !