Perceptron

$w^0 = 0 \in \mathbb{R}^d$

until convergence

Pick (x_i, y_i)

If mistake

$w_{t+1} = w_t + x_i y_i$

else

do nothing.

end.

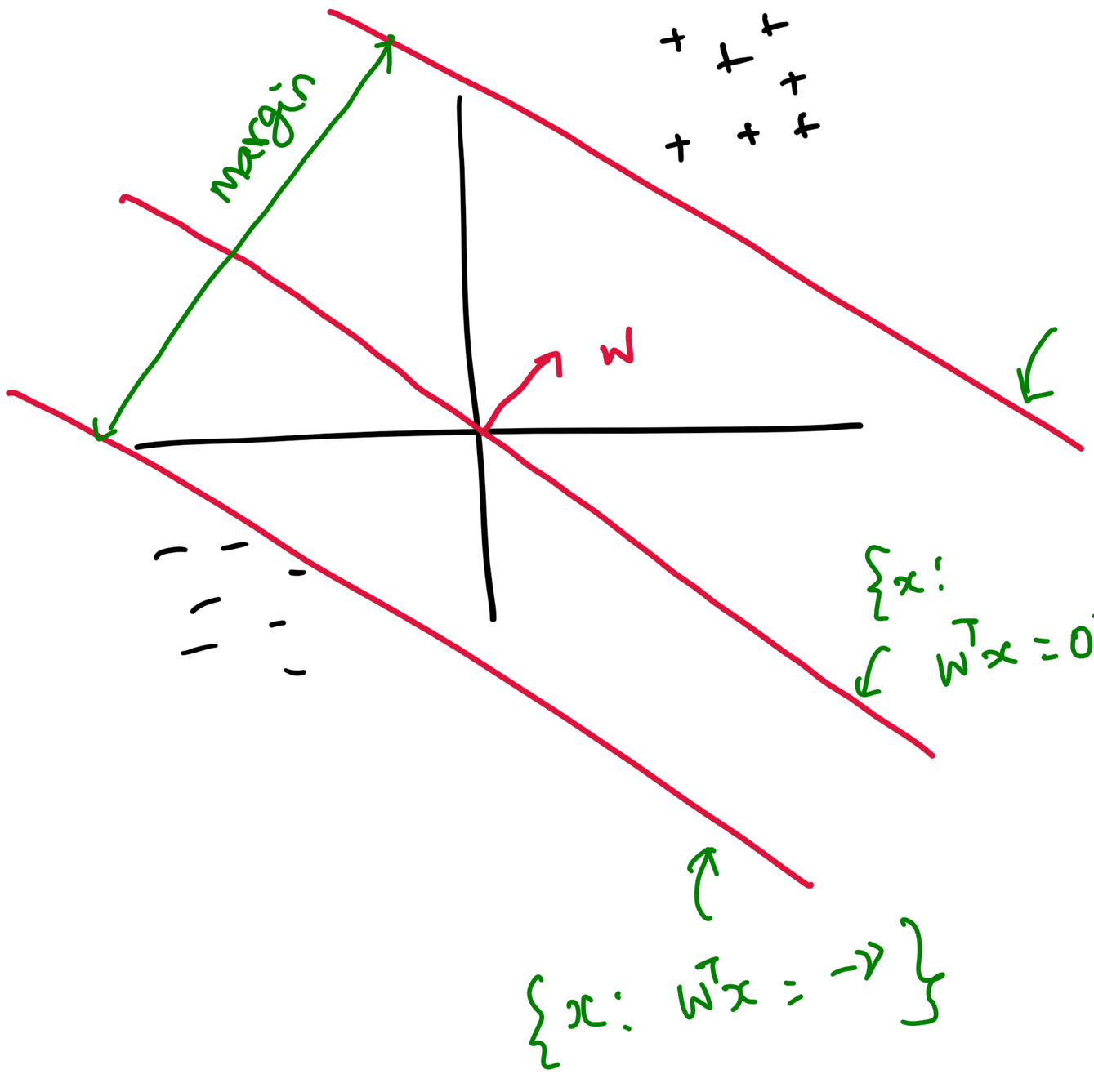
GOAL: Show Perceptron converges!

Assumptions on dataset

① RADIUS

$\forall i, \quad \|x_i\|_2 \leq R$

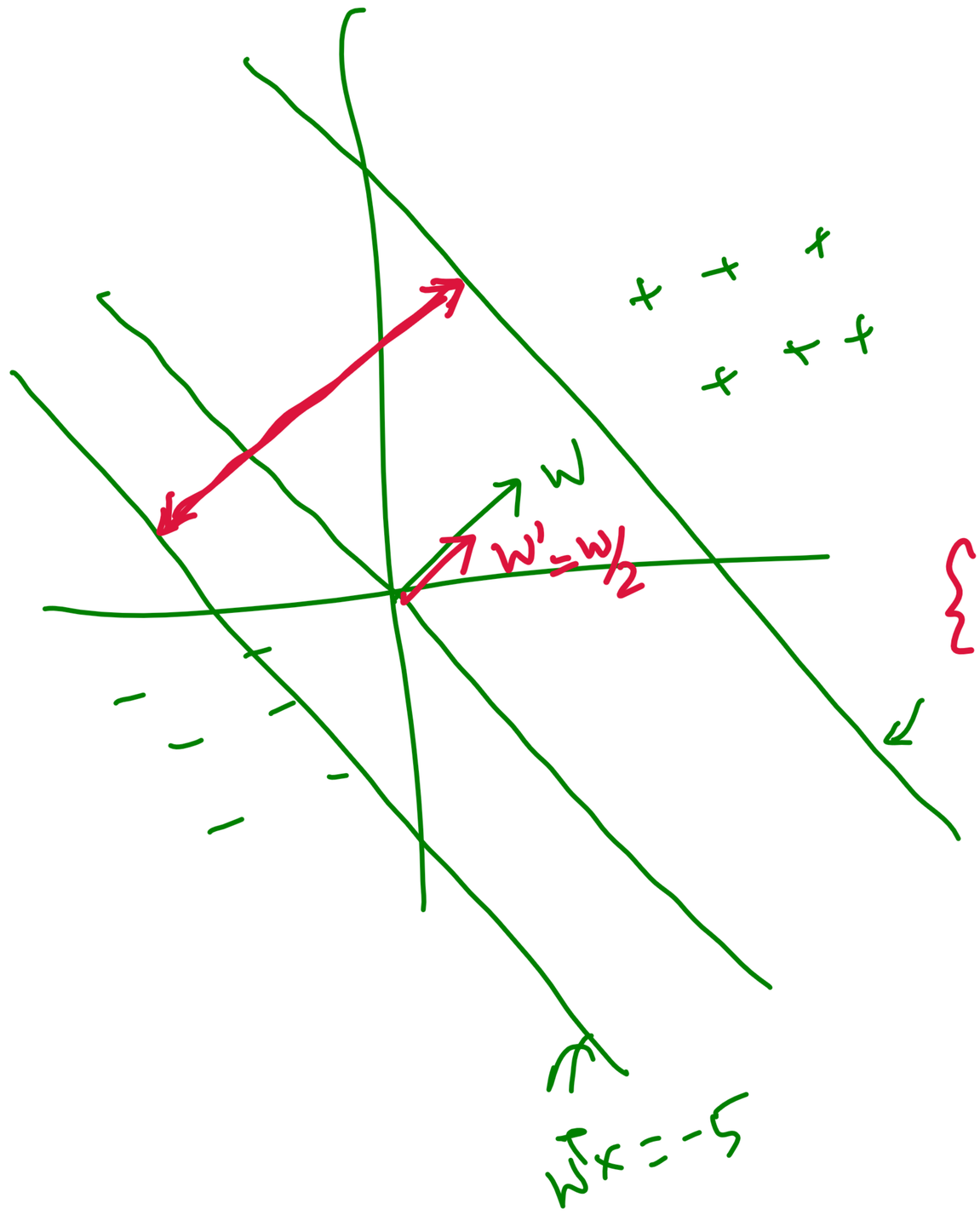
② Linear seperability with γ margin.



Given $D = \{(x_1, y_1) \cdots (x_n, y_n)\}$, D is L.S with margin γ if $\exists\, w^*$ s.t

$$\boxed{(w^{*T} x_i)\, y_i > \gamma} \quad \forall i \quad [\text{for some } \gamma > 0]$$

≥ 0 ✗



$(w^T x > 5)$

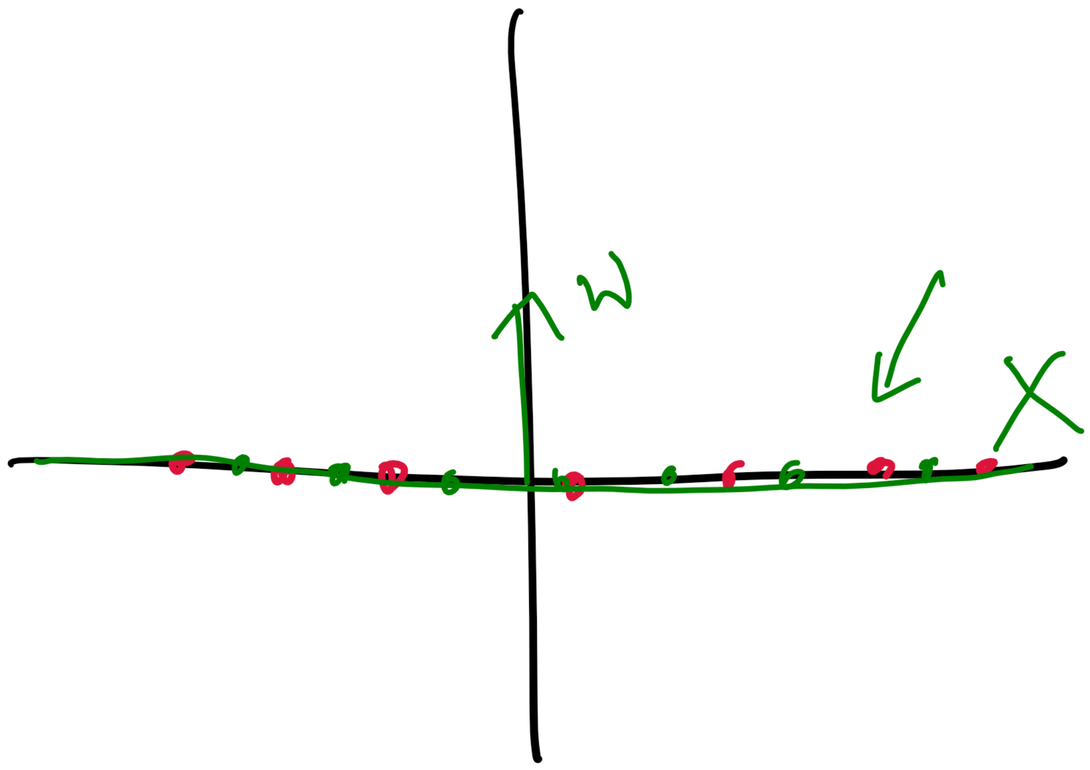
$\{x: w^T x = 5\}$

$= \{x: (\frac{w}{2})^T x = 2.5\}$

$\rightarrow w'^T x = 2.5$

(3) Without loss of generality

Can assume $\|w^*\| = 1$



$(\vec{w}\cdot\vec{x})\,y \geq 0$.

≥ 0

✓

Analysis of "mistakes" of Perceptron.

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

- Say w_k is the current guess and a mistake occurs w.r.t (x, y)

↑ # mistakes.

$$w_{t+1} = w_t + x \cdot y \quad \{\pm 1\} \quad \in \mathbb{R}^d$$

$$\|x\|_2^2 = x^T x = \sum_{i=1}^{d} x_i^2$$

$$\|w_{t+1}\|_2^2 = \|w_t + x \cdot y\|_2^2$$

$$= (w_t + x \cdot y)^T (w_t + xy)$$

$$= w_t^T w_t + (w_t^T x) y + w_t^T (x \cdot y) + (x \cdot y)^T (x \cdot y)$$

$$= \|w_t\|^2 + 2\underbrace{(w_t^T x) y}_{<0} + \|x\|^2 \quad [\text{because of mistake}]$$

$$< \|w_t\|^2 + \|x\|^2$$

$$\|w_{t+1}\|^2 < \|w_t\|^2 + R^2 \quad [\text{by Radius assumption}]$$

$$< (\|w_{t-1}\|^2 + R^2) + R^2 \quad \rightarrow \text{using same argument}$$

$$< \|w_0\|^2 + t R^2$$

$$= t R^2 \quad [\because w_0 = 0 \in \mathbb{R}^d]$$

$$\boxed{\|w_{t+1}\|^2 < t R^2} \quad \text{— (A).}$$

$$w_{t+1}^T w^* = (w_t + xy)^T w^*$$

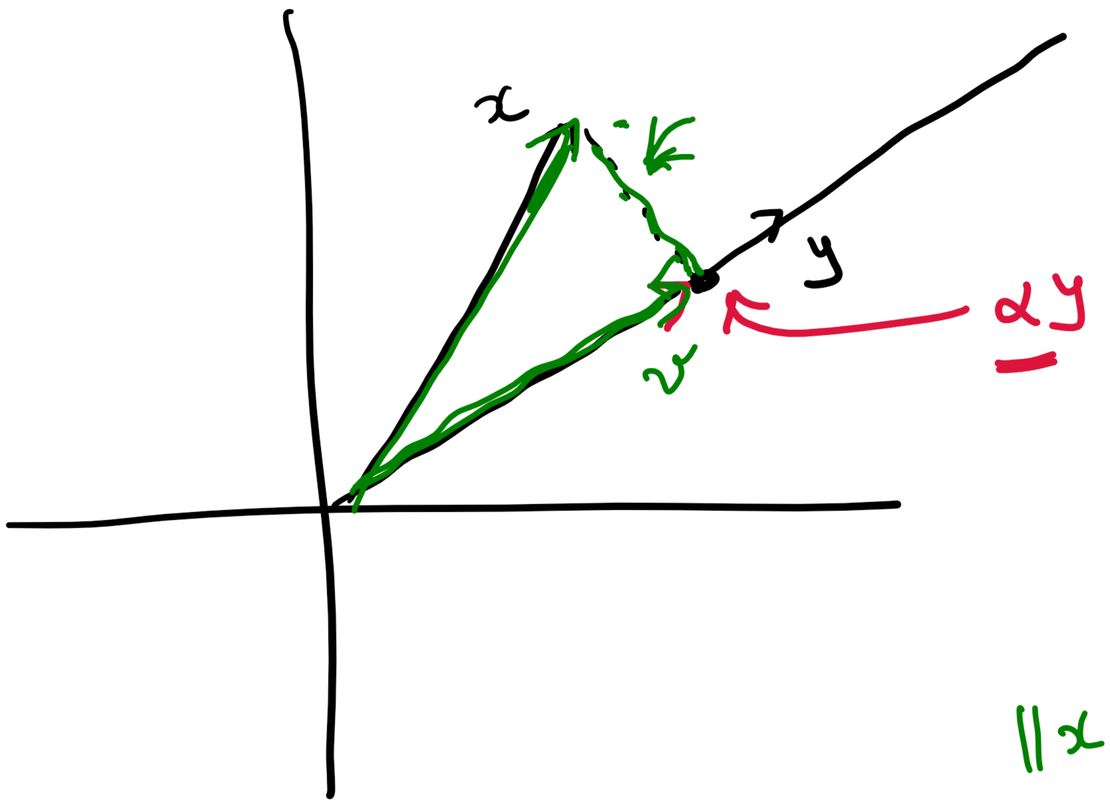
$$= w_t^T w^* + (w^{*T} x) y$$

$$> w_t^T w^* + \gamma \quad [\text{Margin assumption}]$$

$\leftarrow$

$$w_{k+1}^T w^* > k\gamma \quad \text{— (B)}$$

$$\|w_{k+1}\|^2 < kR^2 \quad \text{— (A)}$$



$$\min_{\alpha} \|x - \alpha y\|_2^2$$

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

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

$$(x^T y)^2 \leq \|x\|^2 \|y\|^2$$

$$x^T y \leq \|x\| \|y\|$$

[Cauchy Schwartz inequality]

$$\|w_{t+1}\|\,\|w^*\| \overset{\text{C.S}}{\geq} w_{t+1}^T w^* \overset{(B)}{>} t\gamma$$

$$\Rightarrow \quad \|w_{t+1}\| > t\gamma \longrightarrow (C).$$

Combining (A) & (C)

$$\|w_{t+1}\|^2 < tR^2$$

$$\|w_{t+1}\|^2 > t^2\gamma^2$$

$$\Rightarrow \quad t^2\gamma^2 < tR^2$$

$$\boxed{t < \frac{R^2}{\gamma^2}}$$

#mistakes !

$\Rightarrow$ Perceptron converges!

Important

- Perceptron may not find w^* !

Shortcomings

① there may exist a w^* with a large margin but perceptron does not "necessarily" find it.

② works only for linearly separable datasets.

Adv :- "online" algorithm.