# Linear Discriminant Function

- Fisher's Linear Discriminant
  ↳ sometimes also referred to as Linear Discriminant Analysis (LDA)

Assumptions Fisher's Linear Discriminant:

① works best if classes are Gaussian-distributed.

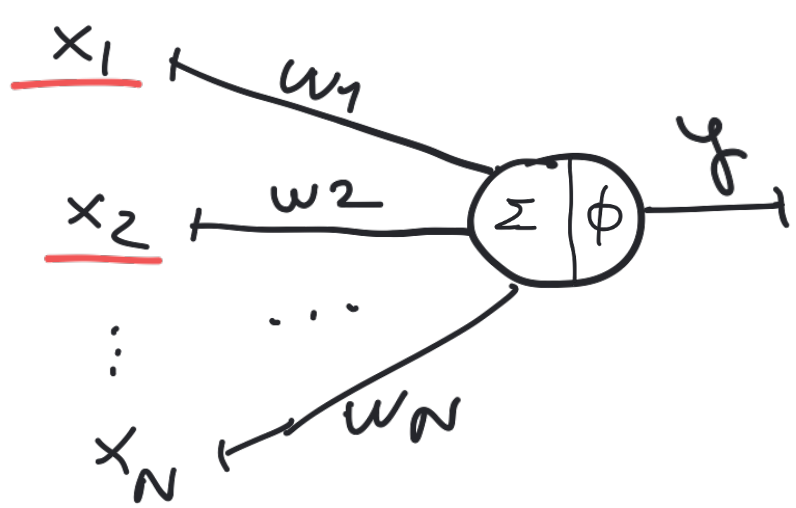
② BUT it still works even if classes are overlapping.

# The Perceptron Algorithm


Excitatory
or
inhibitory
information

NEURON

A cell fires if passes some threshold
↳ we characterize this with an ACTIVATION function $\phi(x)$

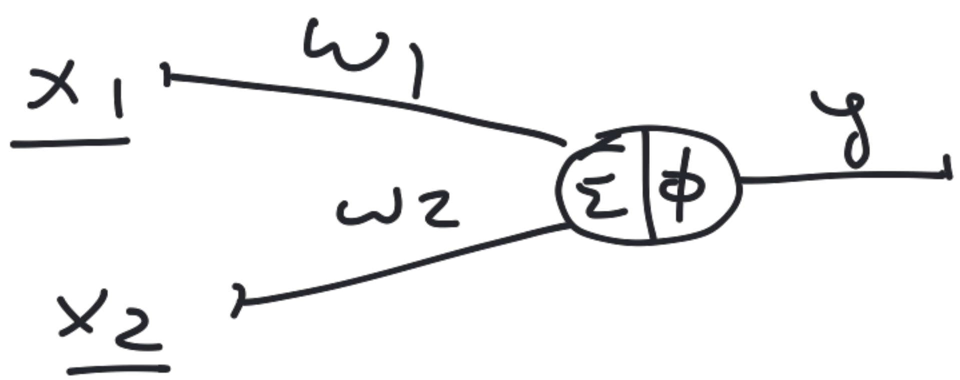



1958

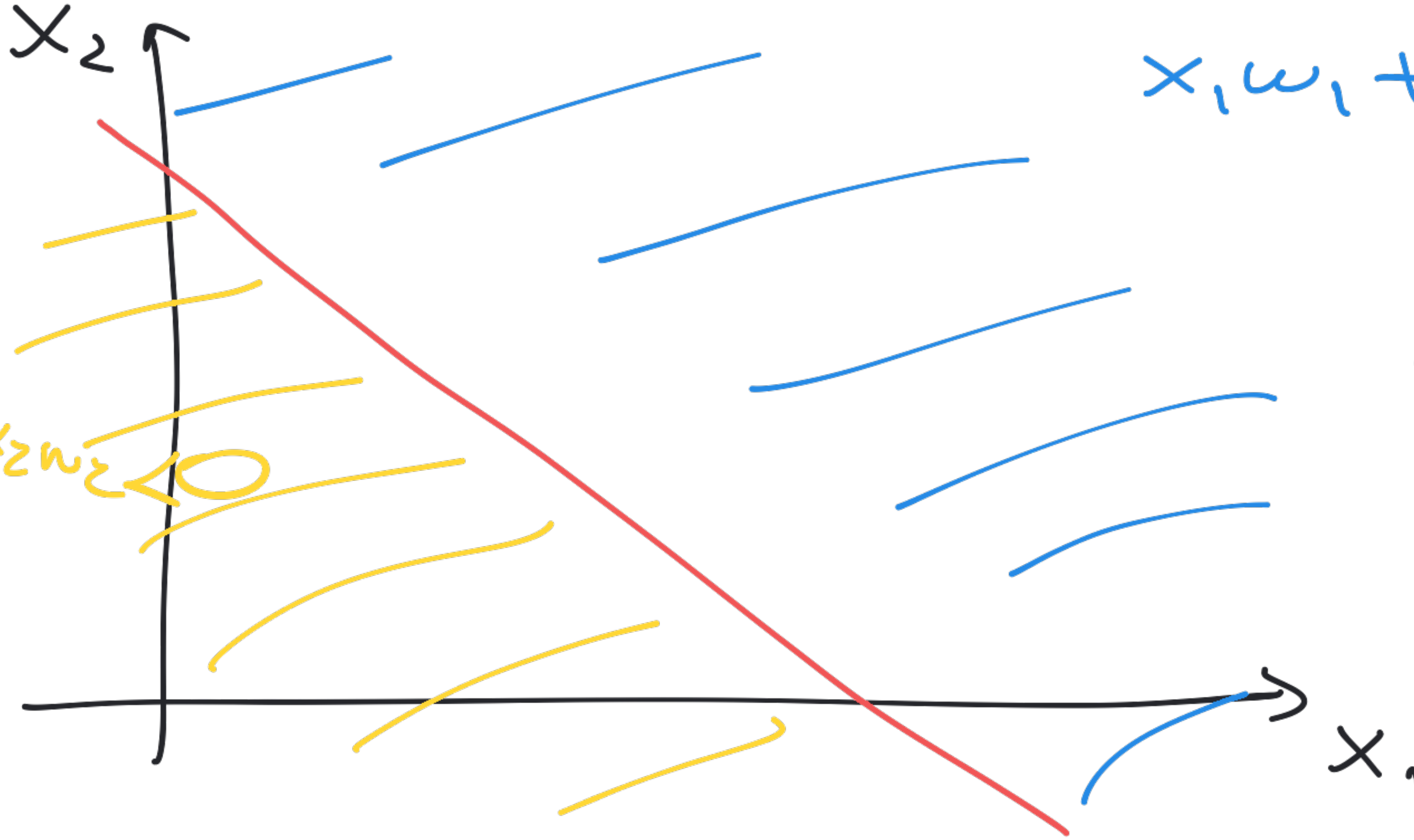
$\boxed{\text{ADALINE}}$

$$y = \phi\left(\sum_{i=1}^{N} w_i \cdot x_i\right)$$

$$\phi(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$




$$y = \phi(x_1 w_1 + x_2 w_2)$$




$x_1 w_1 + x_2 \cdot w_2 \geq 0$

$\downarrow$

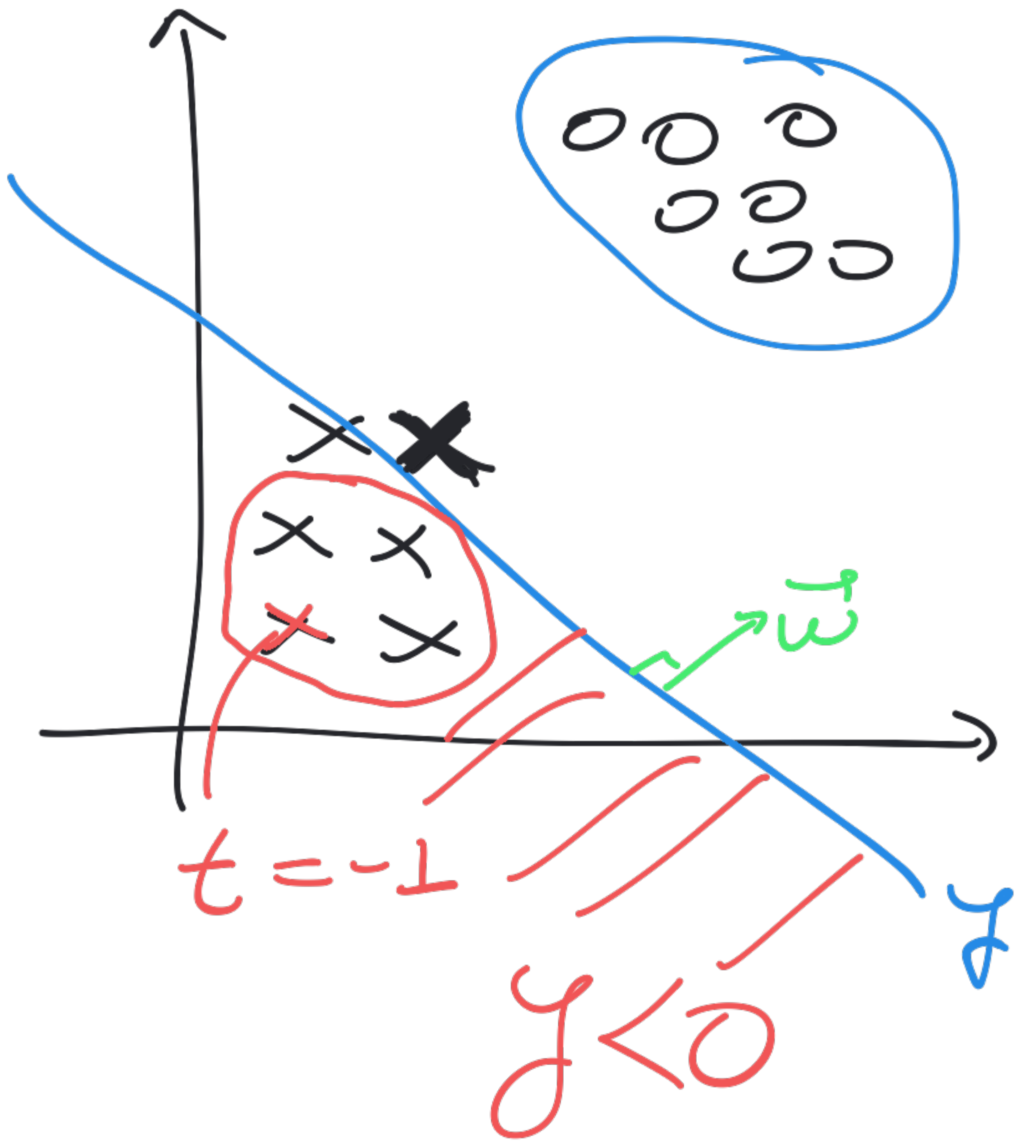
class 1

$x_1 w_1 + x_2 w_2 < 0$

$\downarrow$

class -1

$\{x_i, t_i\}_{i=1}^{N}$ , $t_i \in \{-1, 1\}$
class labels




o - class 1
x - class -1

$y = w^T x_i + b$

$$\overbrace{(w^T x_i + b) \cdot t_i}^{y} > 0$$

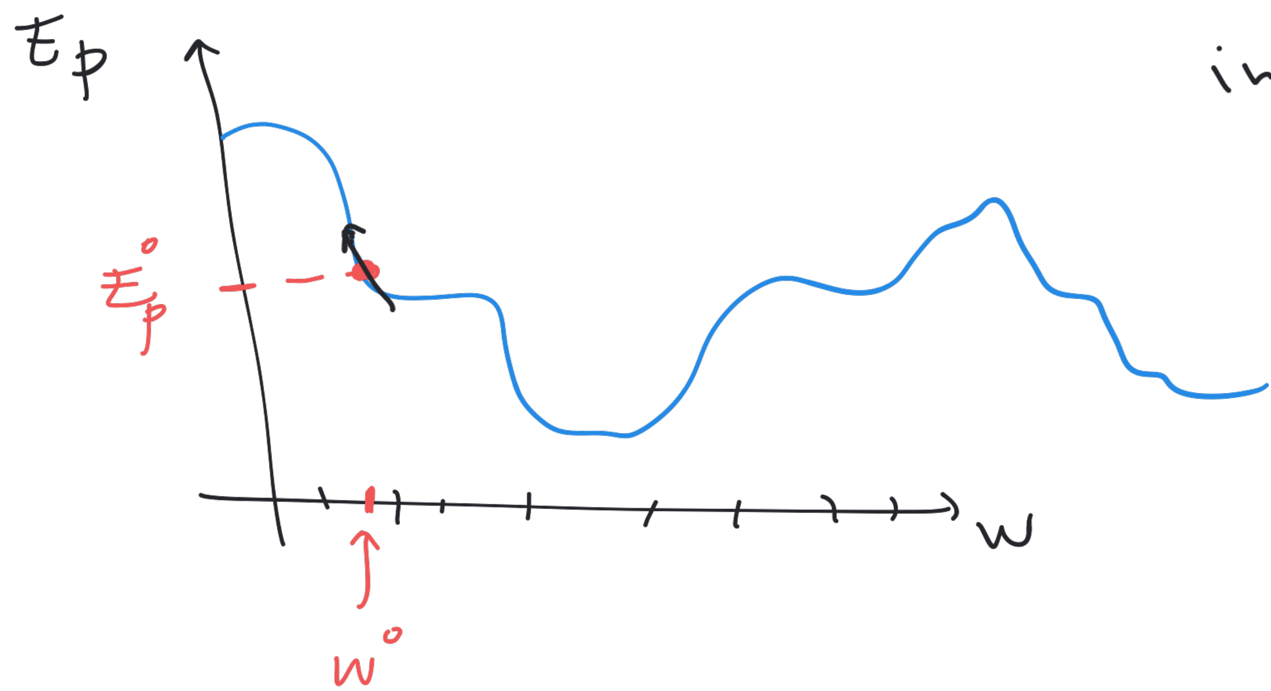
↳ this holds for all $x_i$ that are correctly classified (on the right side of the boundary)

$$E_p(w,b) = -\sum_{n \in M} (w^T x_n + b) \cdot t_n$$

$M \equiv$ set of all misclassified points.

$$\boxed{\arg\min_{w,b} E_p(w,b)}$$

1. First randomly select parameters $w, b$ for the boundary.
2. Compute the $E_p(w,b)$
3. Update $w^{t+1}$ and $b^{t+1}$ in the direction that minimizes $E_p$.




Gradient points in the direction of maximum change.

We want to move in the opposite direction of gradient.

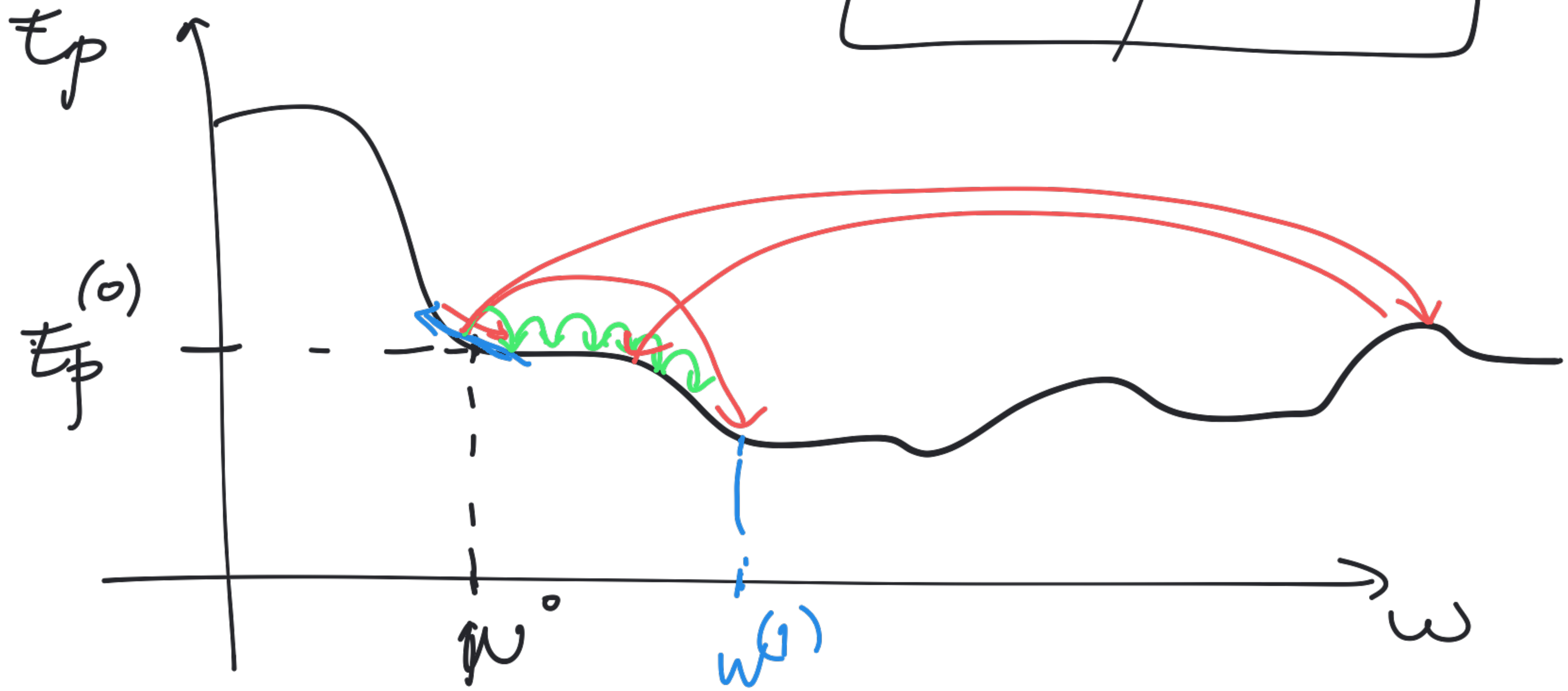
$$w^{(t+1)} \longleftarrow w^{(t)} - \eta \frac{\partial E_p(w,b)}{\partial w}$$

$$b^{(t+1)} \longleftarrow b^{(t)} - \eta \frac{\partial E_p(w,b)}{\partial b}$$

$\eta \equiv$ learning rate

controls the speed or the "step" size in the negative direction of gradient

$$\boxed{0 < \eta < 1}$$



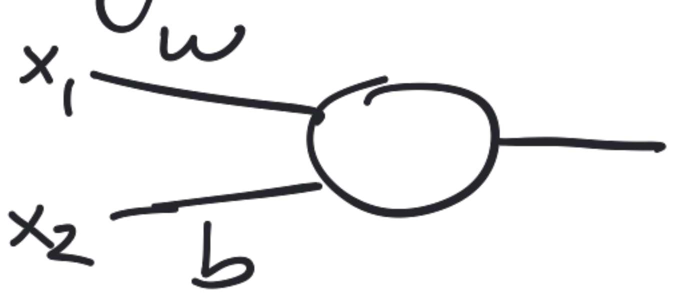

- if $\eta$ is large : it's faster but chaotic behavior.
- if $\eta$ is small : it takes longer to find a solution.

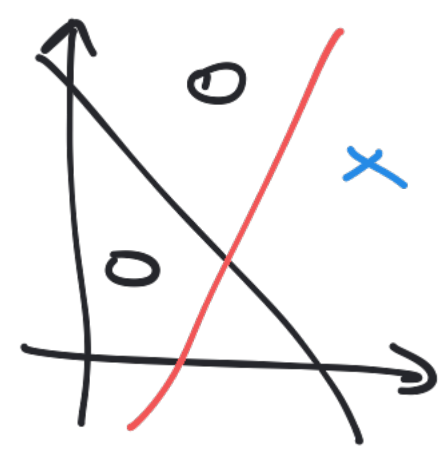
$$\boxed{E_p(w,b) = -\sum_{n \in M} (w^T \underline{x_n} + b) \cdot \underline{t_n}}$$

$$w^{(t+1)} \leftarrow w^{(t)} - \eta \cdot \frac{\partial E_p(w,b)}{\partial w} = w^{(t)} + \eta \cdot x_n \cdot t_n$$

$$b^{(t+1)} \leftarrow b^{(t)} - \eta \cdot \frac{\partial E_p(w,b)}{\partial b} = b^{(t)} + \eta \cdot t_n$$

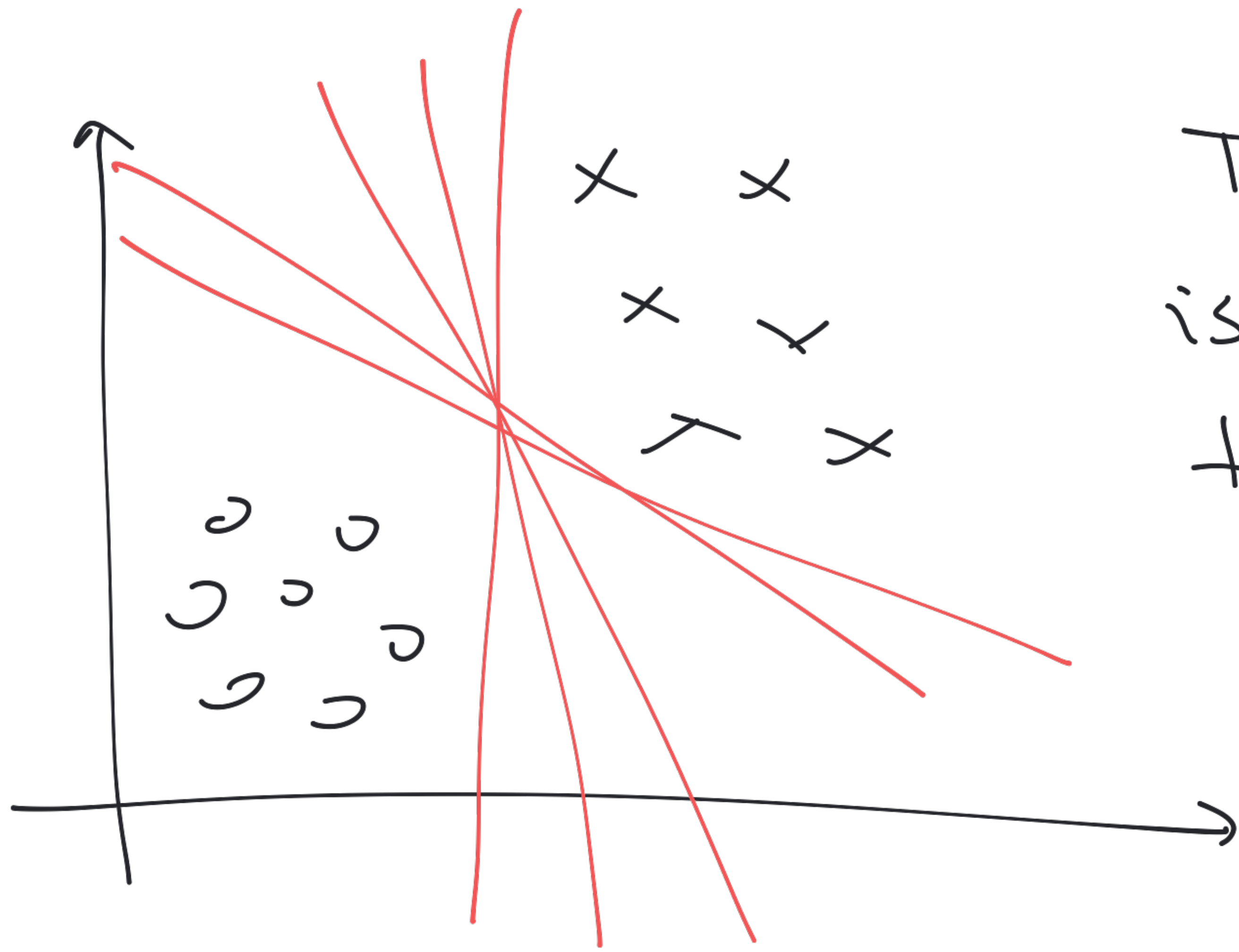
The perceptron updates the boundary using one sample at a time.



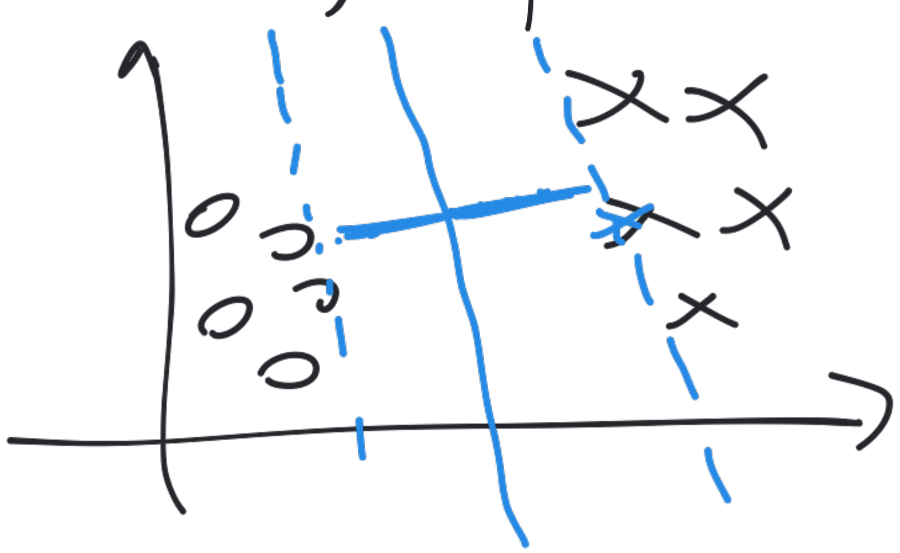
**Online Update**

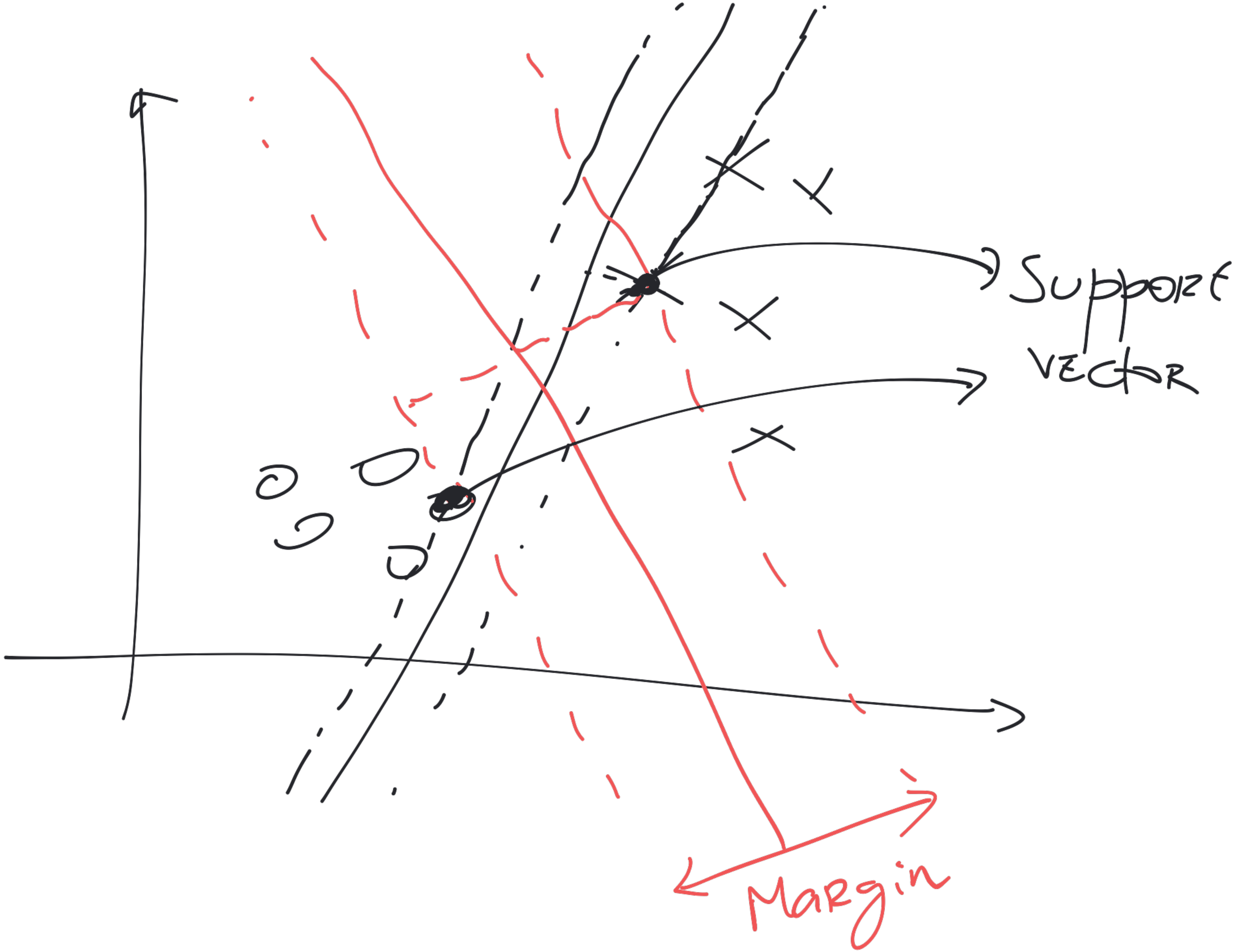
If data is not linearly separable, the perceptron will <u>not</u> converge.



The perceptron is guaranteed to find a solution in a finite # steps if the classes linearly separable.

In 1963, Vapnik:




Support vector
Margin