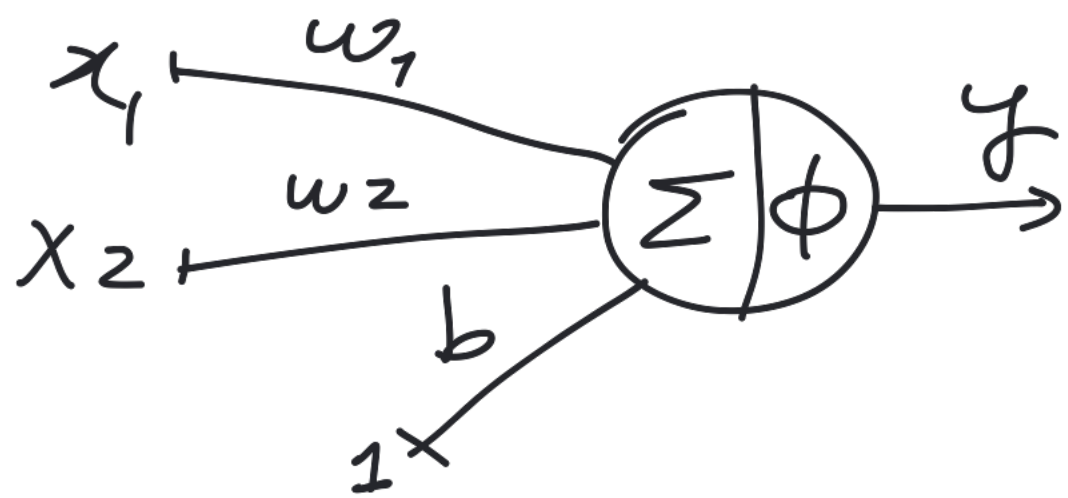
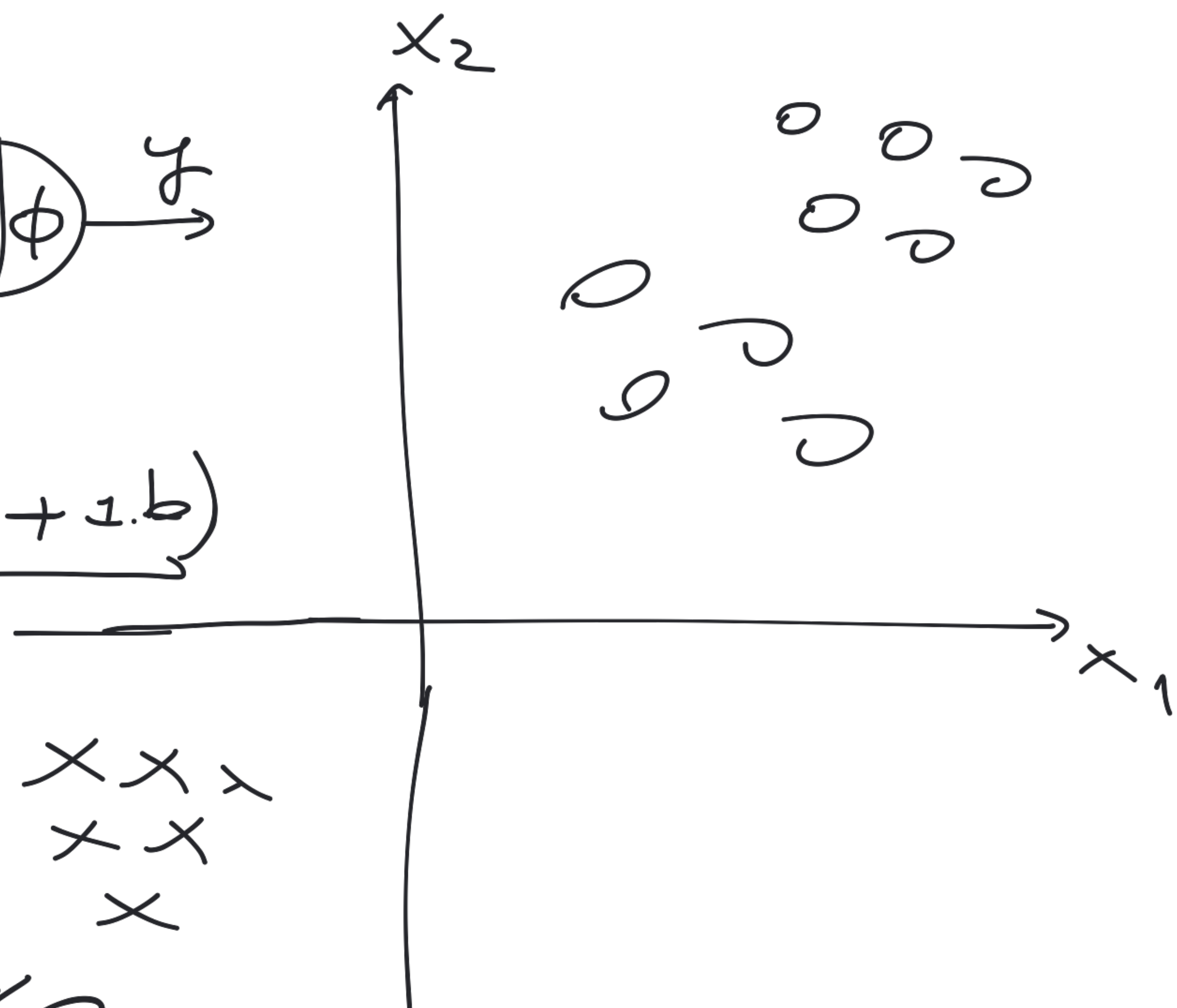
1 Linearly classes C_1 and C_2

1.1





$$y = \phi(\underbrace{w_1x_1 + w_2x_2 + 1.b})$$

$$w_1x_1 + w_2x_2 + b > 0$$

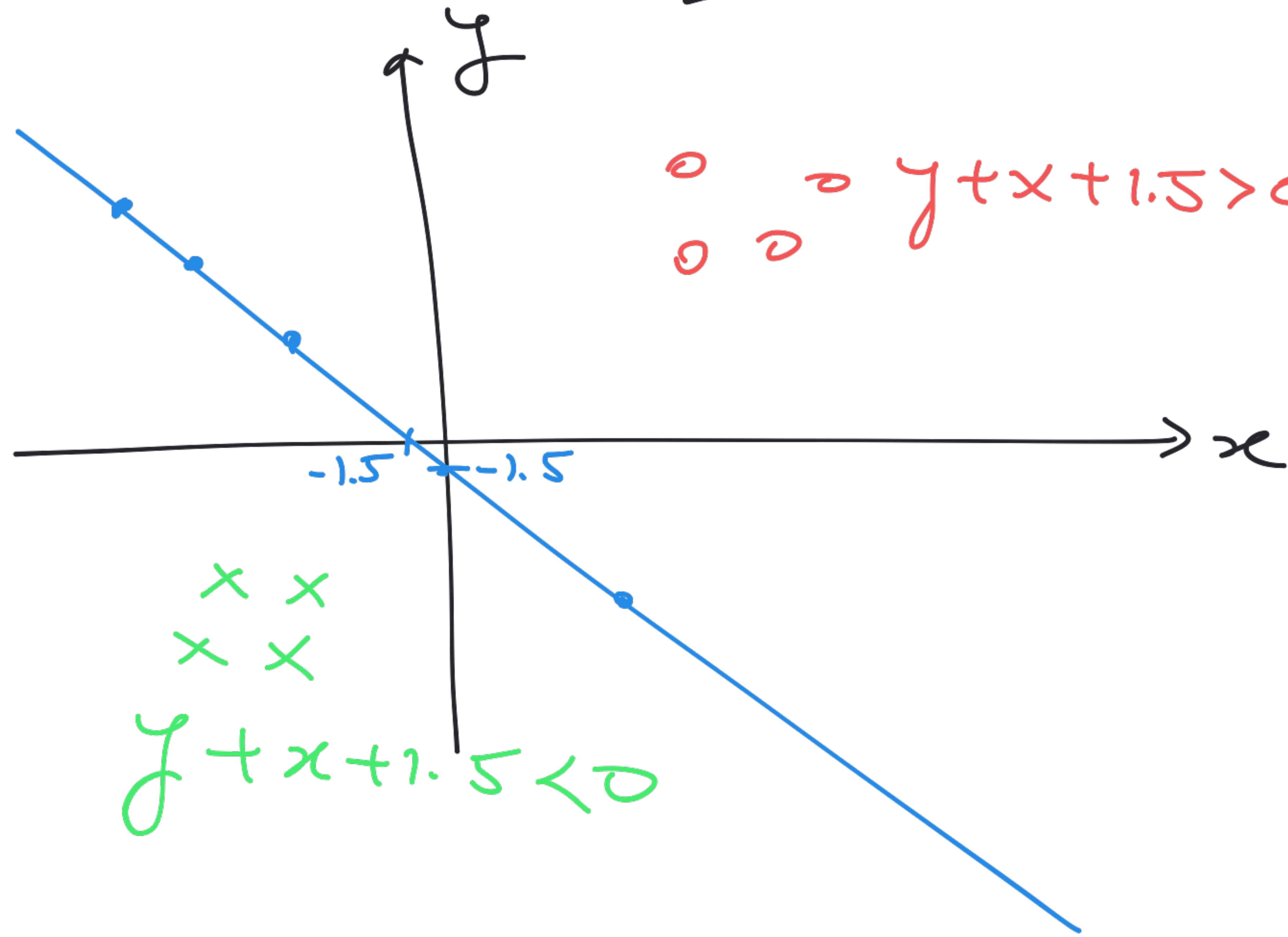
$$\Rightarrow (x_1, x_2) \in C_o$$

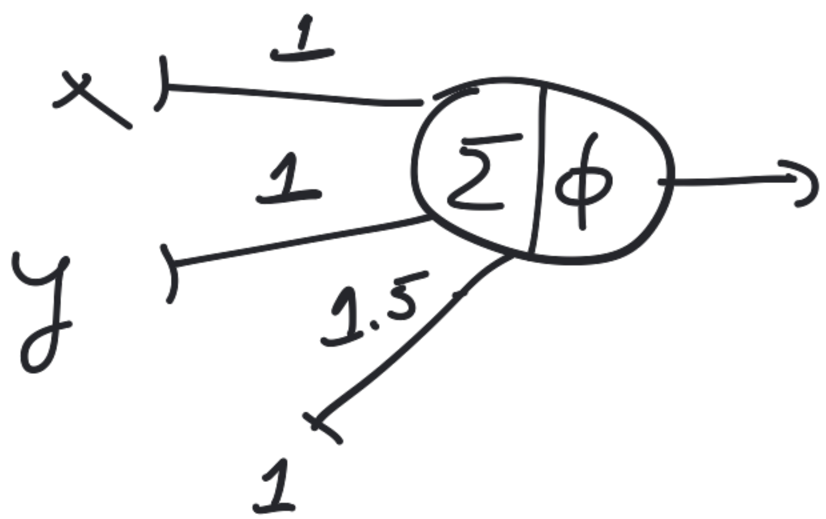
if $w_1x_1 + w_2x_2 + b < 0$

$$\Rightarrow (x_1, x_2) \in C_x$$

1.2

$$y = -x - 1.5 \iff \boxed{y + x + 1.5 = 0}$$





Perceptron Limitations:

1. Can only separate linearly separable classes

2. The solution is not unique
   - ↳ May converge to local optima
   - ↳ Because it depends on initial value of parameters.

3. User-defined LEARNING RATE (OR STEP SIZE) η.
   - ↳ η controls the speed of convergence
   - ↳ $0 < \eta < 1$.

Support Vector Machine (SVM)

↳ Introduces Margin

↳ It finds linear decision boundary that maximizes margin

↳ It uses kernel space to find a linear boundary

$$\boxed{y(x) = w^T \cdot \underbrace{\phi(x)} + b}$$

$\phi(x)$: high-dimensional features

$\phi(x)$ is kernel space and can be characterized with different kernel functions, e.g :

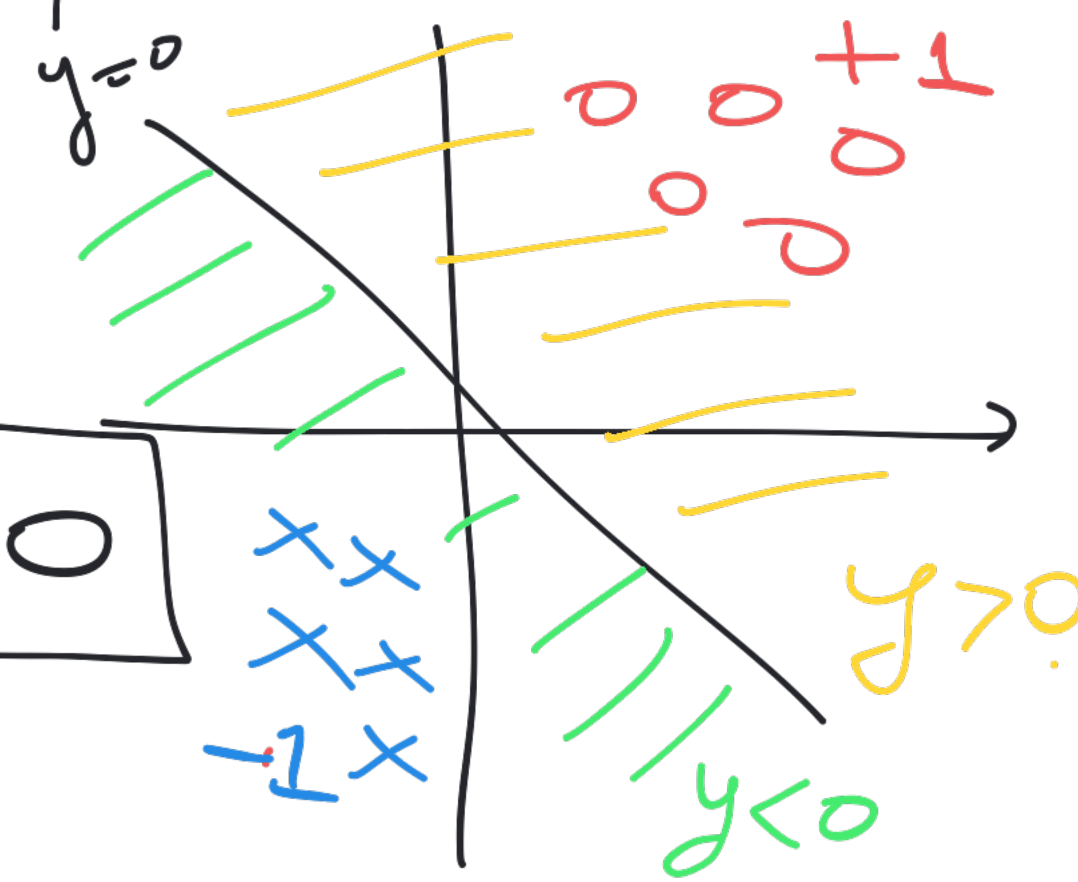
- Radial basis fct (RBF)
- polynomial
- Gaussian kernel

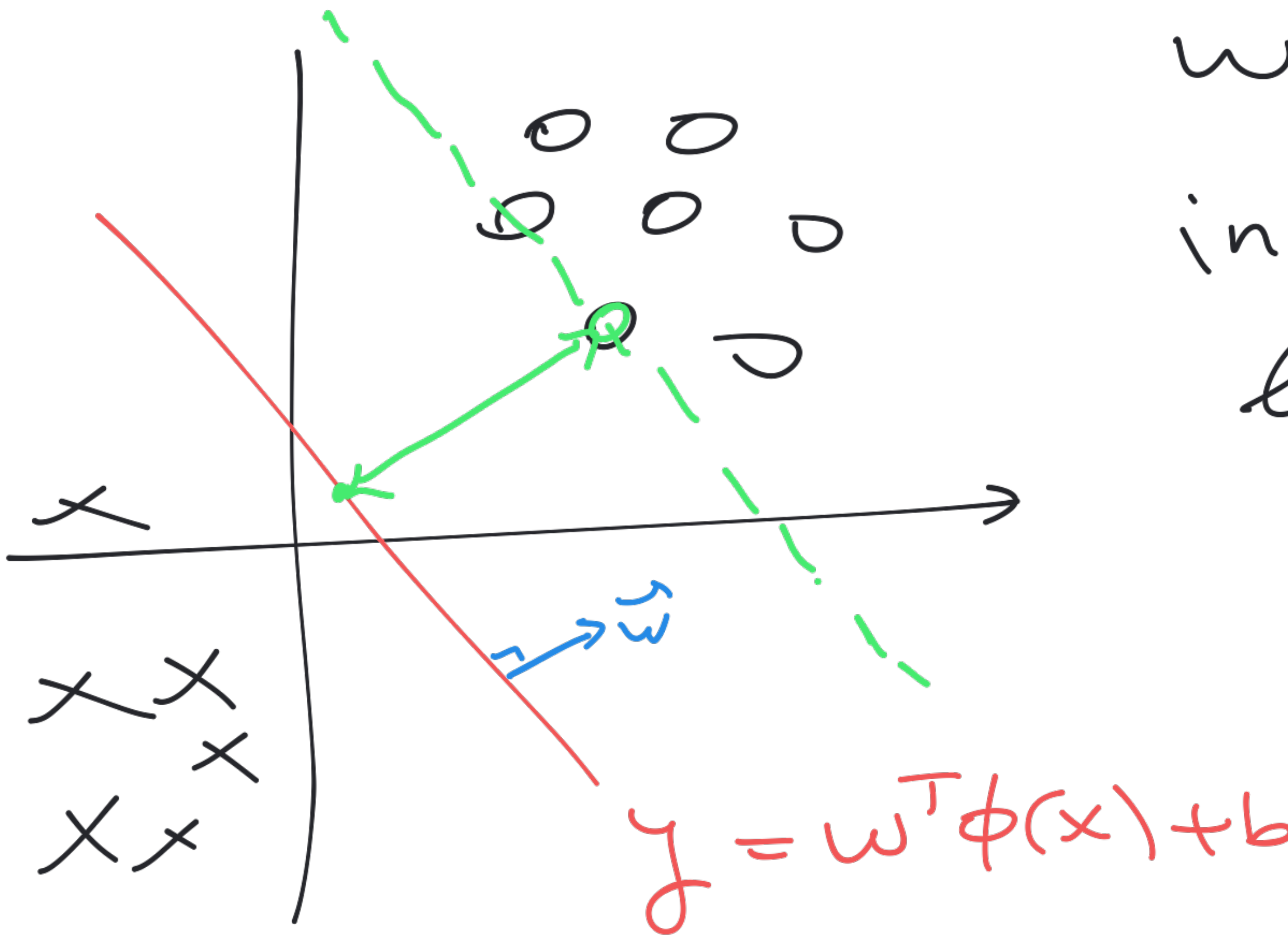
$\{x_i, t_i\}_{i=1}^{N}$, $t_i \in \{-1, 1\}$

- If $y(x_n) > 0$ for points with $t_n = 1$

  and $y(x_n) < 0$ for points with $t_n = -1$

$$\boxed{t_n \cdot y(x_n) > 0}$$





we are interested in finding a linear hyperplane that correctly classifies all points:

$$\boxed{t_n \cdot y(x_n) > 0}$$

- Distance of any point x to the boundary $y(x) = \dfrac{|y(x)|}{||w||}$

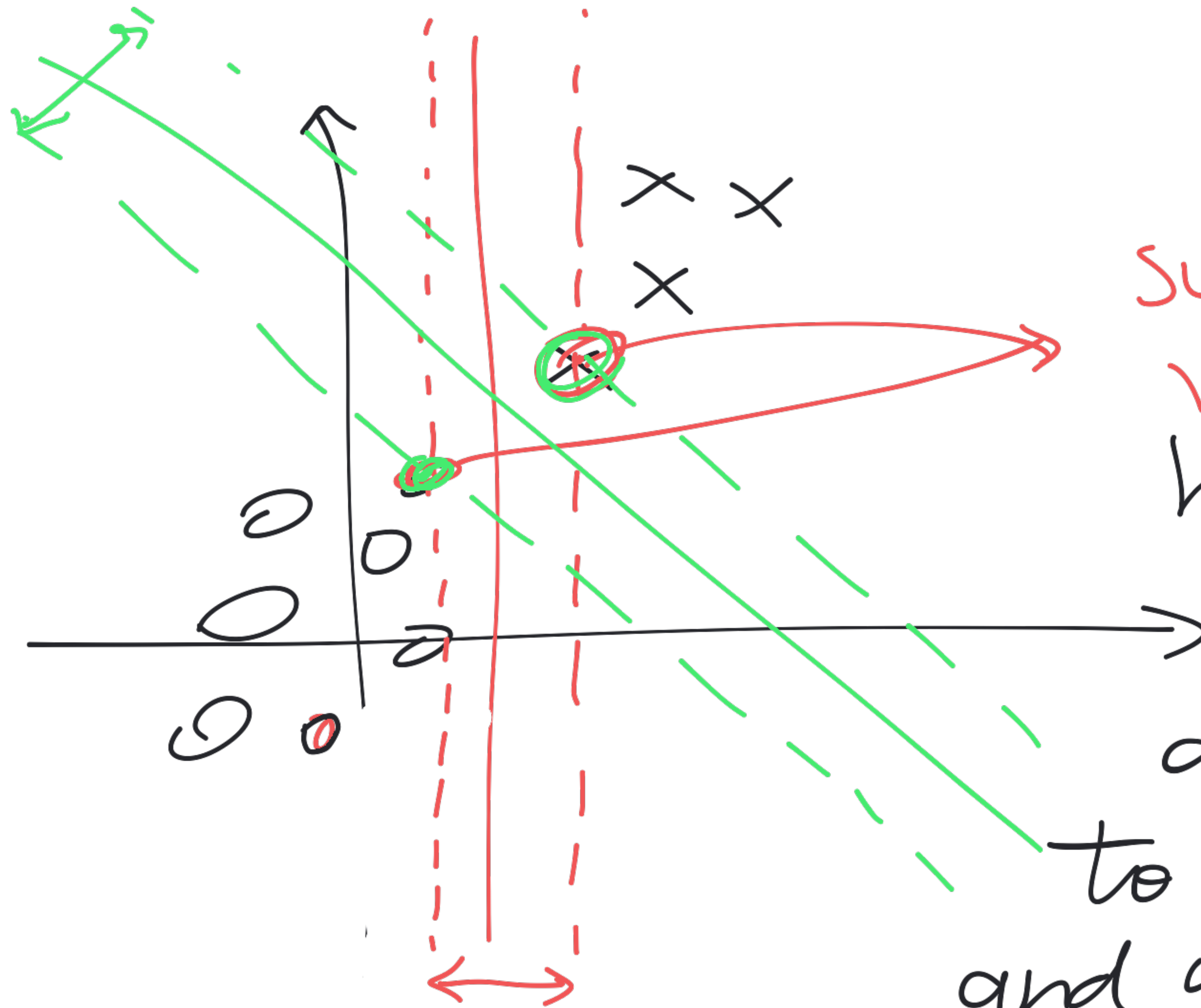
Then, the margin is in $\perp$ direction of w:

$$\frac{t_n \cdot y(x_n)}{\|w\|} = \frac{t_n(w^T\phi(x_n)+b)}{\|w\|}$$

The margin is perpendicular to this distance. Then:

$$\arg\max_{w,b} \left\{ \frac{1}{\|w\|} \cdot \min_n \left[t_n(w^T\phi(x_n)+b) \right] \right\}$$

support vectors



have the smallest distance to boundary and yet are still correctly classified.

→ Margin only uses support vectors
that is, fully characterized by S.V.s.

↳ We will not use the entire data to find w, b !!

↳ This decreases memory needs

$$\arg\max_{w,b} \left\{ \frac{1}{\|w\|} \min_n \left[t_n(w^T\phi(x_n)+b) \right] \right\}$$

$$\begin{aligned} w &\longrightarrow K.w \\ b &\longrightarrow K.b \end{aligned}$$

The distance from any point ~~x~~ x_n is unchanged

$$\frac{t_n . y(x_n)}{\|w\|}$$

Knowing this, we can consider:

$$t_n(w^T\phi(x_n)+b) = 1$$

The points where $t_n(w^T\phi(x_n)+b) = 1$ are the support vectors.

For all x_n:

$$\boxed{t_n(w^T\phi(x_n)+b) \geq 1}$$

We are then interested to maximize margin subject to $t_n(w^T\phi(x_n)+b) \geq 1$

maximizing $\frac{1}{\|w\|}$ is the same as minimizing $\|w\|^2$

$$\arg\min_{w,b} \|w\|^2$$

Subject to $t_n(w^T\phi(x_n)+b) \geq 1$

Read APPENDIX E from Bishop textbook