FoDA L26

The Perceptron Algorithm

for linear classification

Nov 29, 2022

Linear Classification

Input $(X, y) \in \mathbb{R}^d \times \{-1, +1\}$

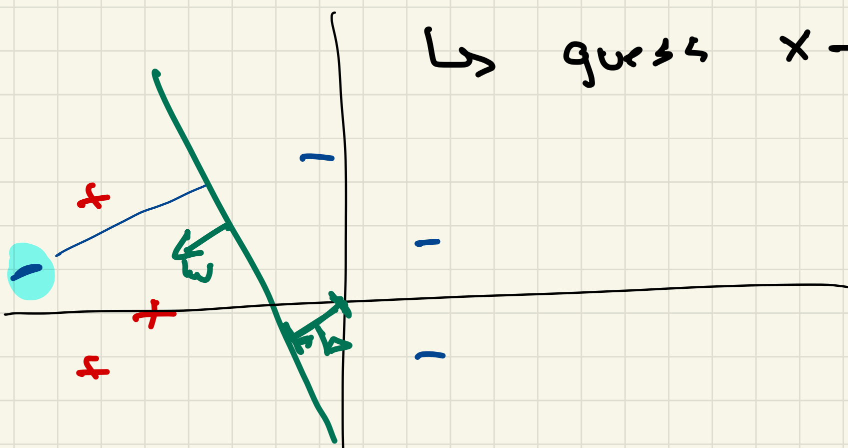
data point (x_i, y_i) $\quad x_i \in \mathbb{R}^d$ $\quad$ label $\{-1, +1\}$

Output linear classifier $(w, b) \in \mathbb{R}^d \times \mathbb{R}$

$$g_{w,b}(x) = b + \langle w, x \rangle$$

↳ guess $x \to +1$ if $g_{w,b}(x) > 0$

$ \to -1$ if $g_{w,b}(x) < 0$

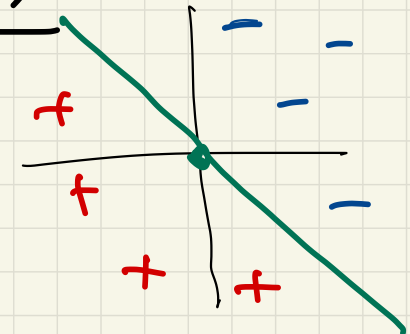


Simplification for Perceptron

① Assume decision boundary passes through origin

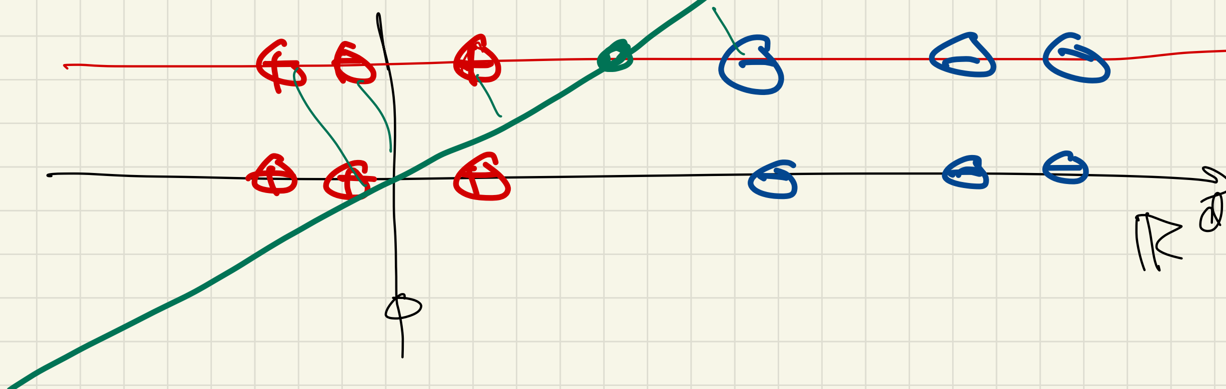
$b = 0$

$g_w(x) = \langle x, w \rangle$



convert $x \to (1; x) \in \mathbb{R}^{d+1}$

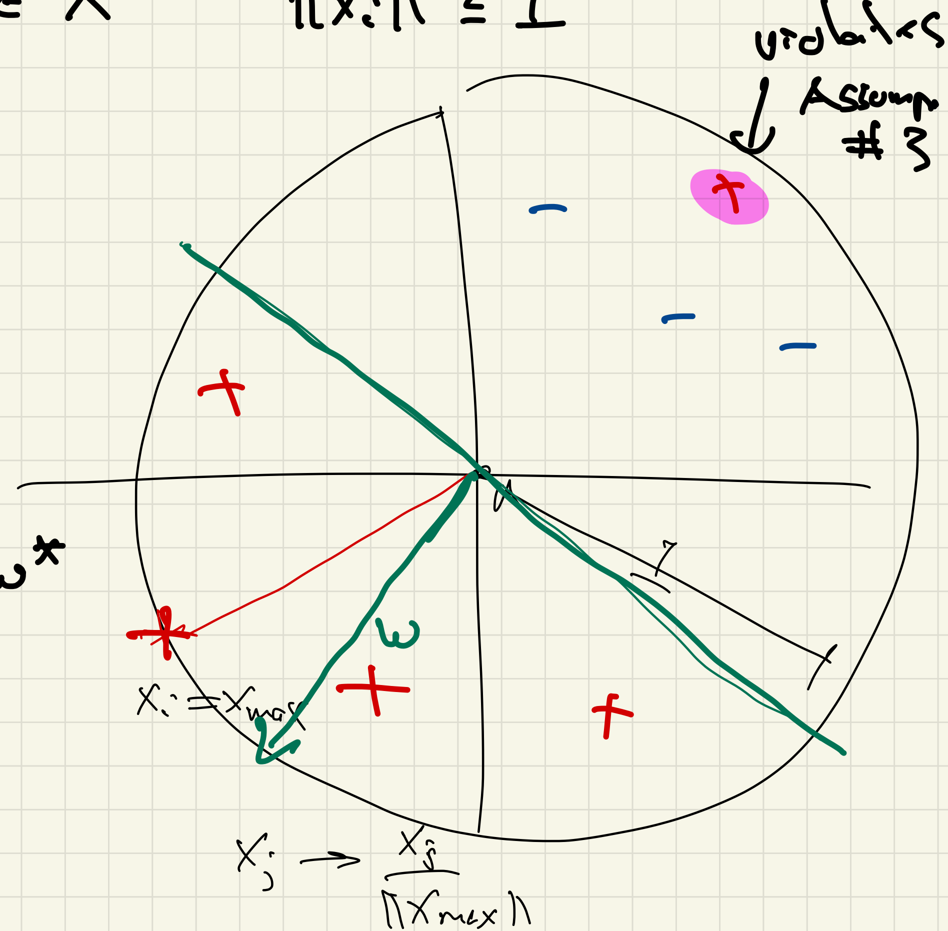
$(b, w) = (\alpha_0, \alpha_1 \ldots \alpha_d) = \alpha \in \mathbb{R}^{d+1}$



② Assume $\forall x_i \in X \quad \|x_i\| \le 1$

③ Assume the exist a "perfect" classifier w^* s.t. that $\forall (x_i, y_i) \in (X, y)$

$$\text{sign}(g_{w^*}(x_i)) = y_i$$



Perceptron Algorithm

0. Initialize : set $\underline{w = y_i x_i} \in \mathbb{R}^d$ for any $(x_i, y_i) \in (X, y)$

1. <u>repeat</u>

   for any $(x_i, y_i) \in (X, y)$ s.t. $\underbrace{\overbrace{y_i \langle x_i, w \rangle}^{g_w(x_i)} < 0}_{\text{misclassified}}$

   update $w \leftarrow w + y_i x_i$

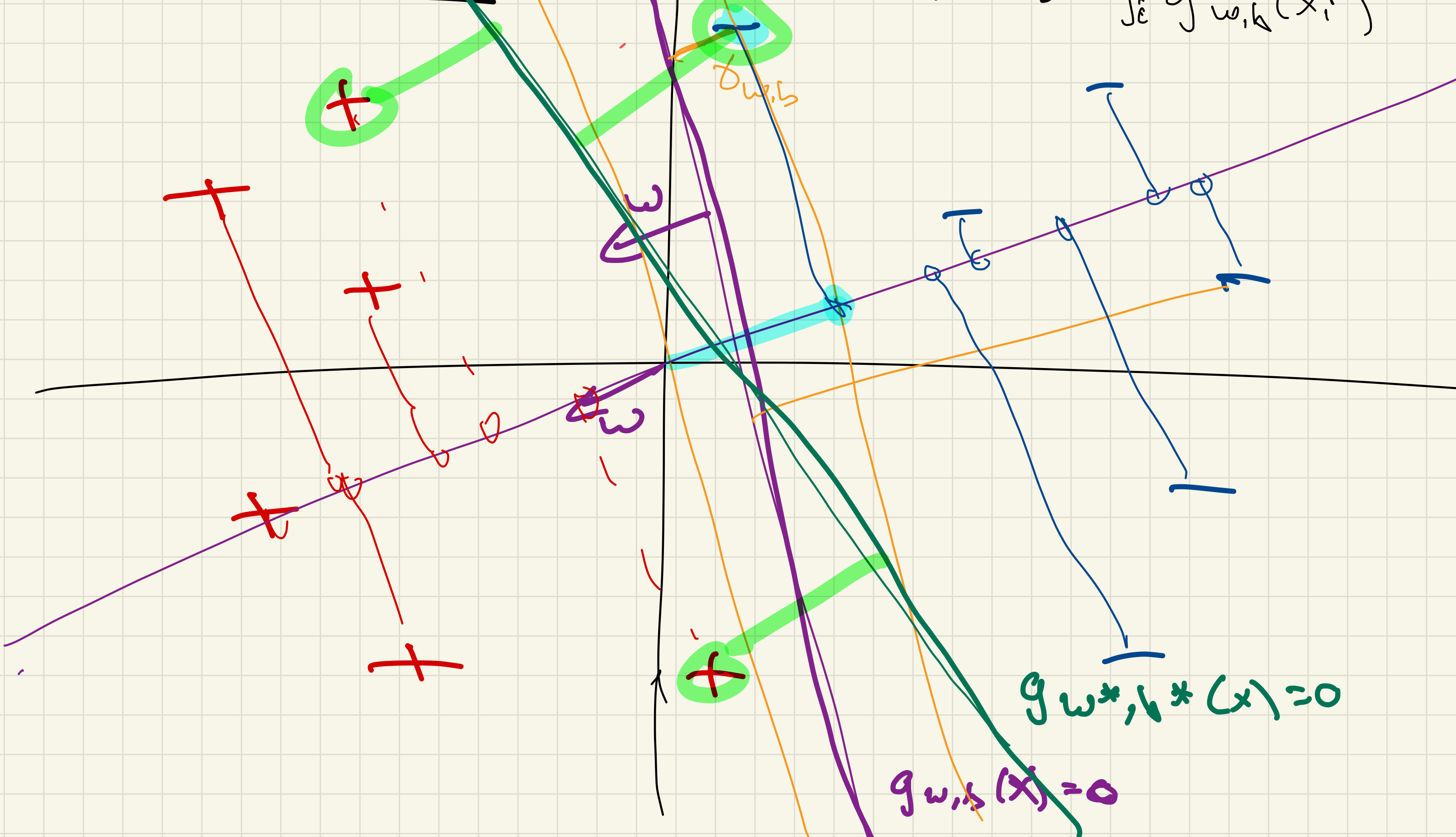
   <u>until</u> (no point misclassified or T steps)

2. <u>return</u> $w \leftarrow \frac{w}{\|w\|}$

the Margin

$$\gamma = \min_{(x_i, y_i) \in (X, y)} y_i(\langle w, x_i \rangle + b)$$

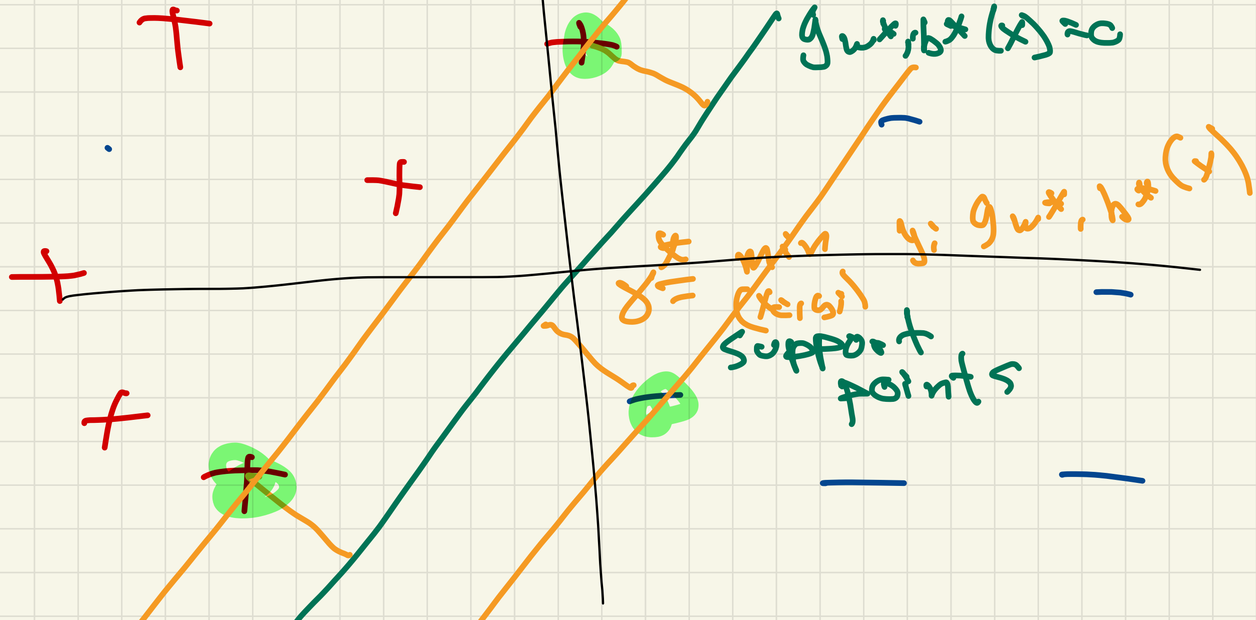
$$y_i \, g_{w,b}(x_i)$$

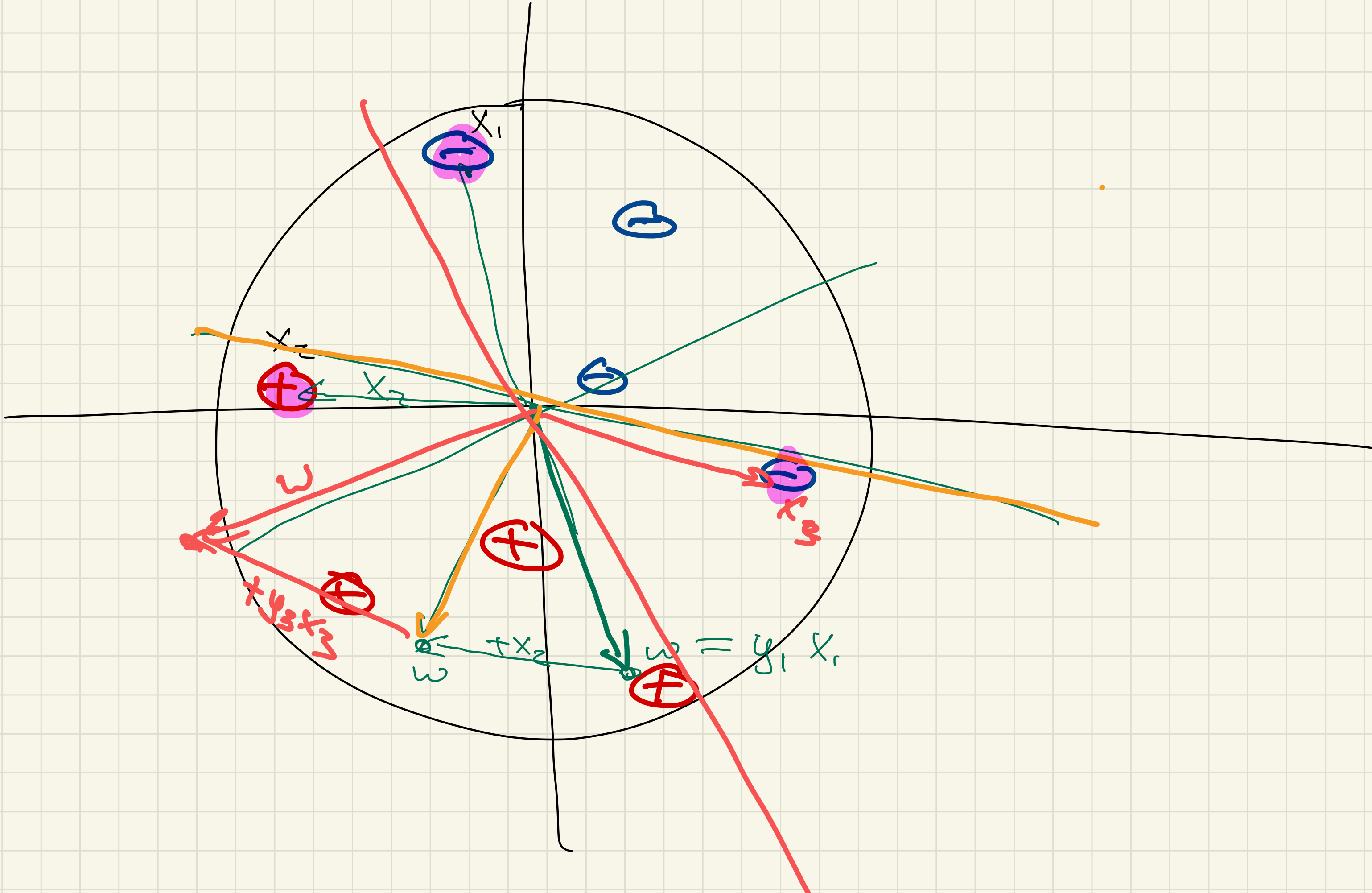


Max-Margin Separator

$$w^*, b^* = \underset{(w,b)\in\mathbb{R}^d\times\mathbb{R}}{\arg\max} \; \boxed{\min_{(x_i,y_i)\in(X,Y)} y_i \, g_{w,b}(x_i)}$$

margin





x1
x2
x2
w
+y3x3
x3
w
+x2
w = y1 x1

Why does Perceptron work?

uses at most $T = (1/\gamma^*)^2$ steps

max-margin

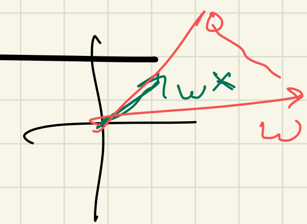
① $\langle w, w^* \rangle$ ② $\|w\|^2 \le t \Rightarrow \|w\| \le \sqrt{t}$

$$\langle w + y_i x_i, w + y_i x_i \rangle = \langle w, w \rangle + \overbrace{(\cancel{y_i})^2}^{1} \overbrace{\langle x_i, x_i \rangle}^{\le 1} + \overbrace{2 y_i \langle w, x_i \rangle}^{\le 0}$$

$$\le \langle w, w \rangle + 1$$

$\|w\|^2 \le t$ after t step

$$\boxed{\langle w, w^* \rangle \ge t\gamma^*}$$



$$\langle w + y_i x_i, w^* \rangle = \langle w, w^* \rangle + \overbrace{y_i \langle x_i, w^* \rangle}^{\ge \gamma^*}$$

$$t\gamma^* \le \langle w, w^* \rangle \le \langle w, \frac{w}{\|w\|} \rangle = \|w\| \le \sqrt{t}$$

$t \le (1/\gamma^*)^2 \Longrightarrow$ Must terminate after $(1/\gamma^*)^2$ steps