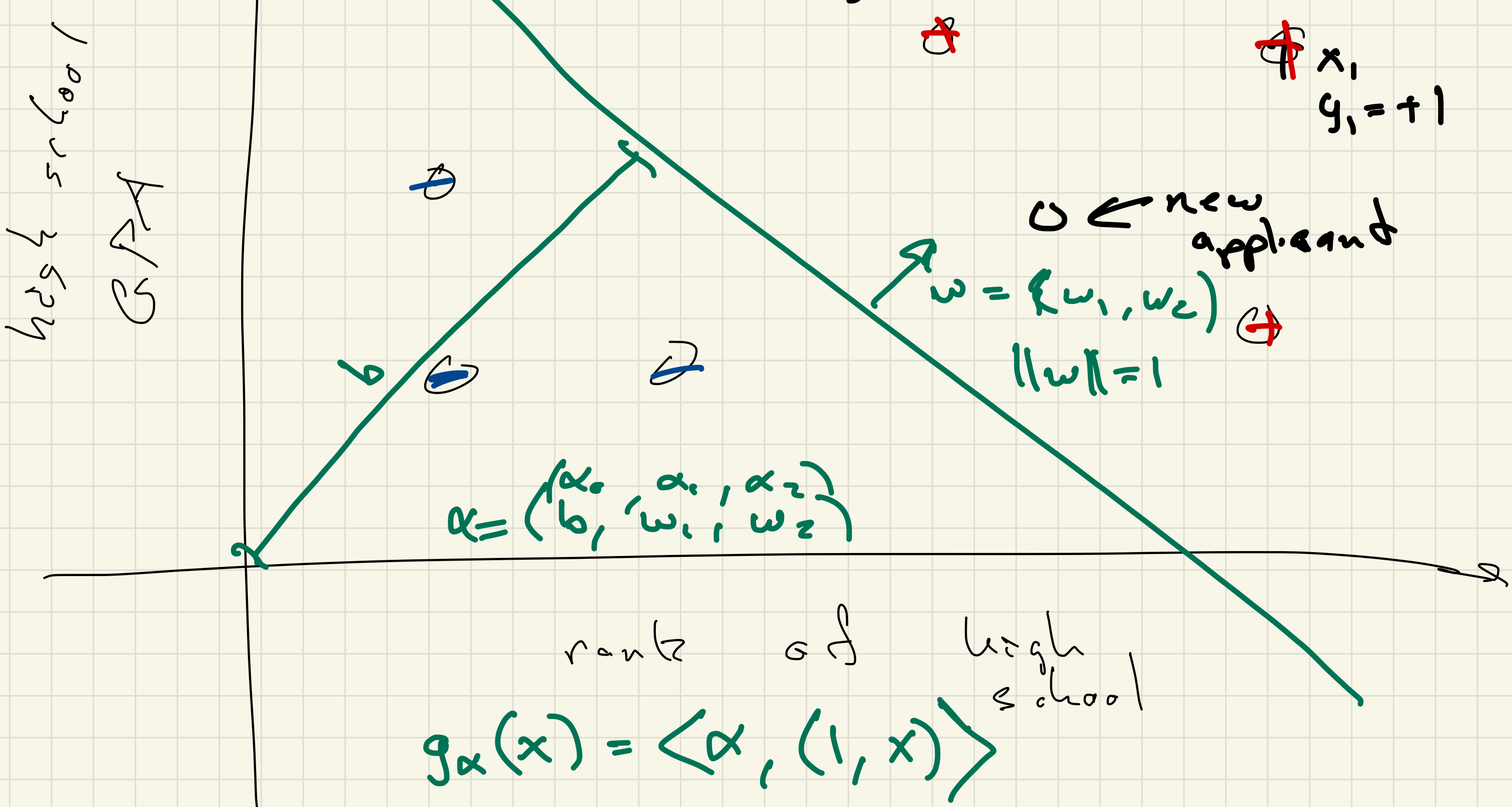
FoDA L26

the Perceptron

Algorithm

Data $X \subset \mathbb{R}^d$

$y \subset \{-1, +1\}$

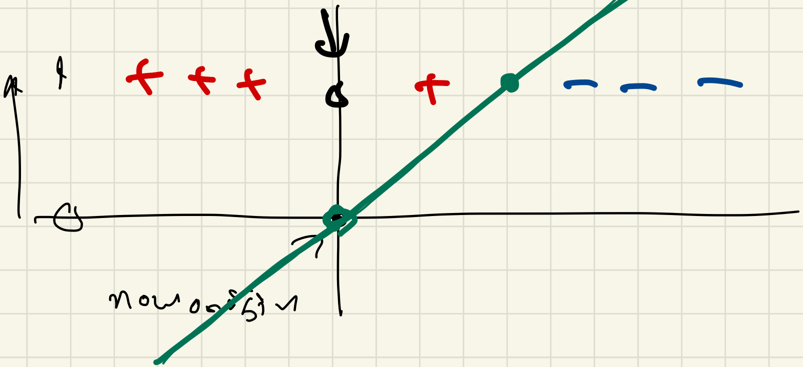


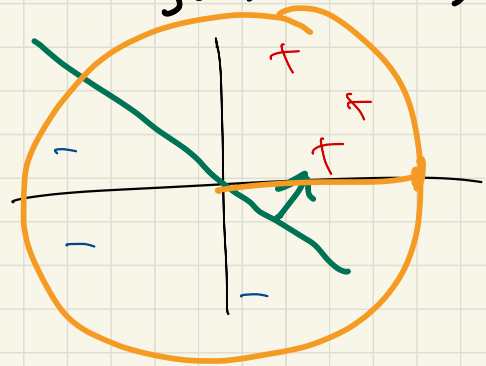
$$g_\alpha(x) = \langle \alpha, (1, x) \rangle$$

Requirements of the Perceptron Algo

1. assume separator goes through origin

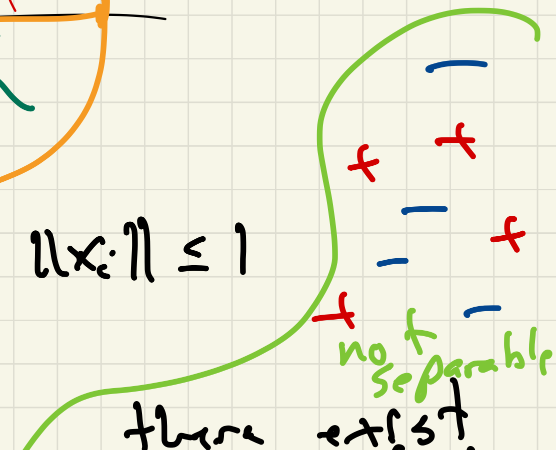
offset $b = 0$





2. assume all $x_i \in X$ have $\|x_i\| \leq 1$

can just $x_i = x_i / \|x_{max}\|$



3. if (X, y) is separable: there exist a perfect separator

Perceptron Algo

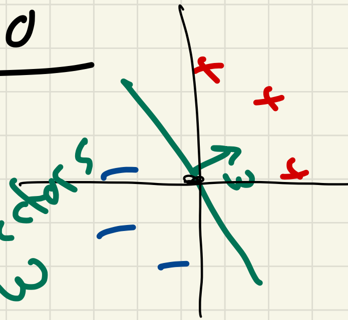
Input (X, y) $\quad y_i \in \{-1, +1\}$

$\|x_i\| \le 1$

Output w s.t. $y_i \langle x_i, w \rangle > 0$

$\|w\| = 1$

duality between pts x_i and vectors w



0. Init $\boxed{w = y_i X_i}$ for any $(x_i, y_i) \in (X, y)$

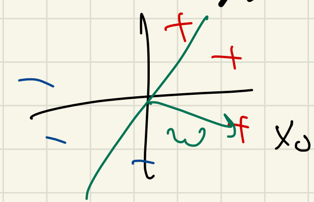
1. repeat

   for any $(x_i, y_i) \in (X, y)$

   s.t. $y_i \langle w, x_i \rangle < 0$ $\quad$ update $w = w + y_i x_i$

   misclassified

   until (no misclassified pts or $\underbrace{T \text{ steps}}_{T \text{ updates}}$)

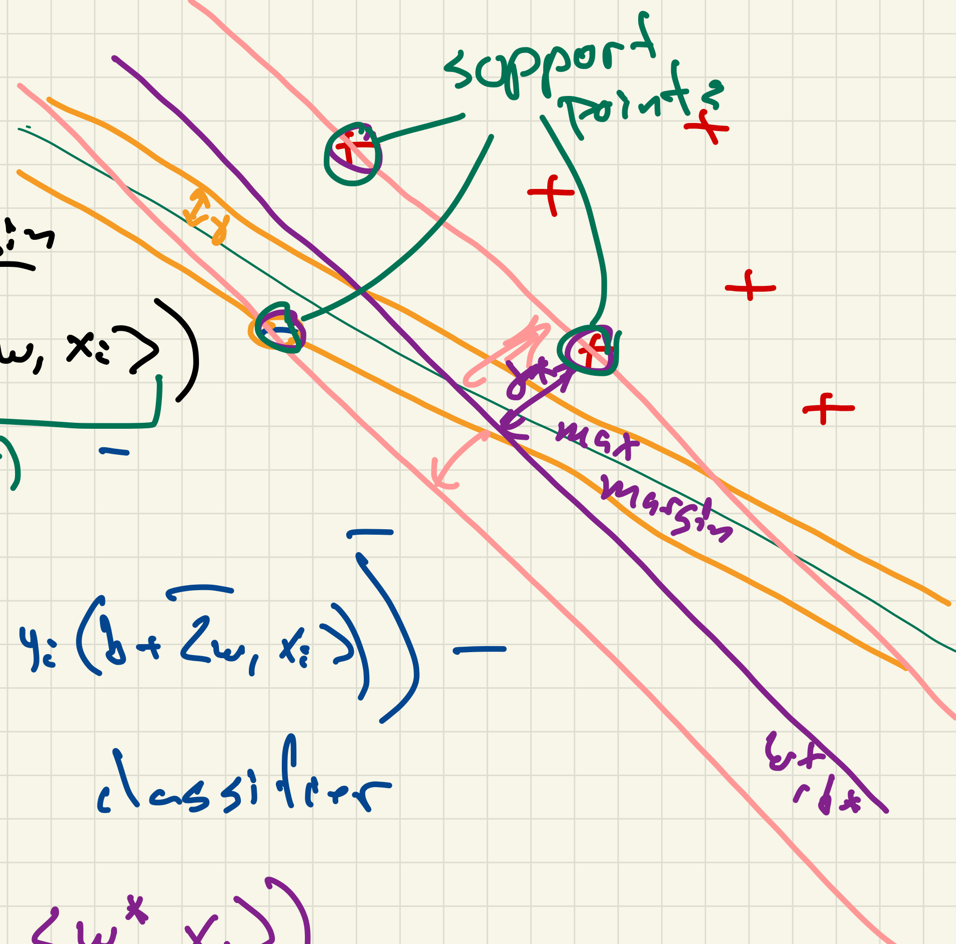


2. return $w / \|w\|$

The Margin

given (w,b) the margin

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

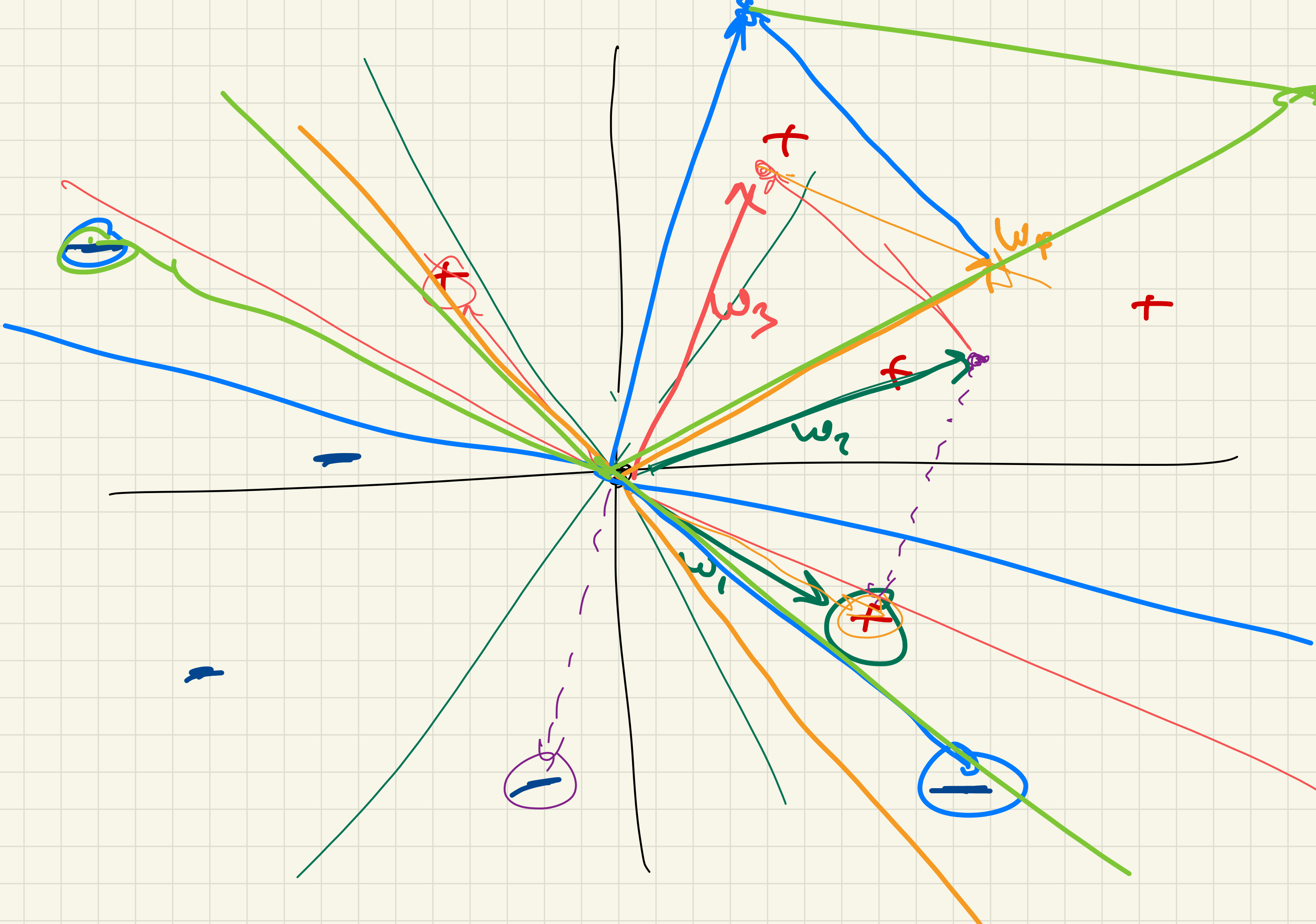


$$(w^*, b^*) = \arg\max_{(w,b)} \left(\min_{(x_i,y_i)} y_i \left(b + \langle w, x_i\rangle\right)\right)$$

↳ max margin classifier

$$\gamma^* = \min_{(x_i,y_i)} y_i \left(b^* + \langle w^*, x_i\rangle\right)$$

max margin



w_3
w_4
w_2
w_1
x
+
−

<u>Claim</u> Perceptron finds a perfect classifier in $\left(\frac{1}{\gamma^*}\right)^2$ updates.

t steps $\quad \|w\|^* = 1$

---

2 quantities $\langle w, w^* \rangle \geq t\gamma^*$

$$\boxed{\|w\|^2 \leq t}$$

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

$$t \leq \left(\frac{1}{\gamma^*}\right)^2$$

update w/ mistakes x_i, y_i

$$\langle w + y_i x_i, w + y_i x_i \rangle = \langle w, w \rangle + (y_i)^2 \langle x_i, x_i \rangle + 2y_i \langle w, x_i \rangle$$

$$\leq \langle w, w \rangle + 1 + 0$$

after t steps $\langle w, w \rangle \leq t$

$$\langle w + y_i x_i, w^* \rangle = \langle w, w^* \rangle + (y_i) \langle x_i, w^* \rangle \geq \langle w, w^* \rangle + \gamma^*$$

$\geq t\gamma^*$