We say $(x^{(1)}, y_1), \cdots, (x^{(n)}, y_n)$ are <u>linearly separable with const $\gamma > 0$</u> if $\exists$ a unit vector $w^* \in \mathbb{R}^p$, $\|w^*\| = 1$ such that

$$y_i \cdot (w^* \cdot x^{(i)}) > \gamma > 0 \quad i = 1, \cdots, n.$$

Perceptron Learning Algorithm

Input: sample pts with labels

$$(x^{(1)}, y_1), \cdots, (x^{(n)}, y_n)$$

Output: weight $w = (w_1, \cdots, w_p)^T$

Step 1: Set $k = 0$, $w^{(0)} = 0$

Step 2: While $\exists (x^{(i)}, y_i)$ such that $y_i(w^{(k)} \cdot x^{(i)}) < 0$, update

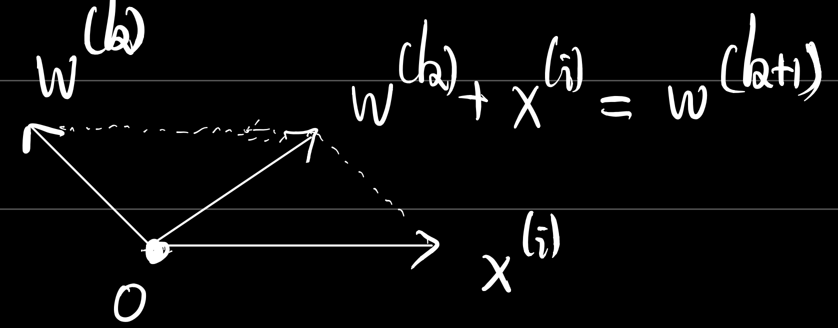
$$w^{(k+1)} = w^{(k)} + y_i x^{(i)}$$

$$k \leftarrow k+1$$

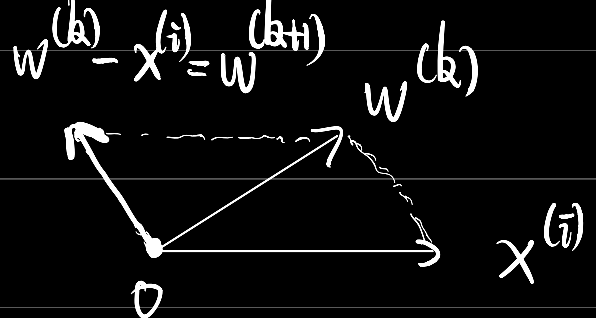
Remark: Here is a heuristic argument why $w^{(k+1)}$ is better than $w^{(k)}$ for classification.

Suppose $(x^{(i)}, y_i)$ is misclassified by $\operatorname{sign}(w^{(k)} \cdot x^{(i)})$.

$y_i = 1$



$y_i = -1$



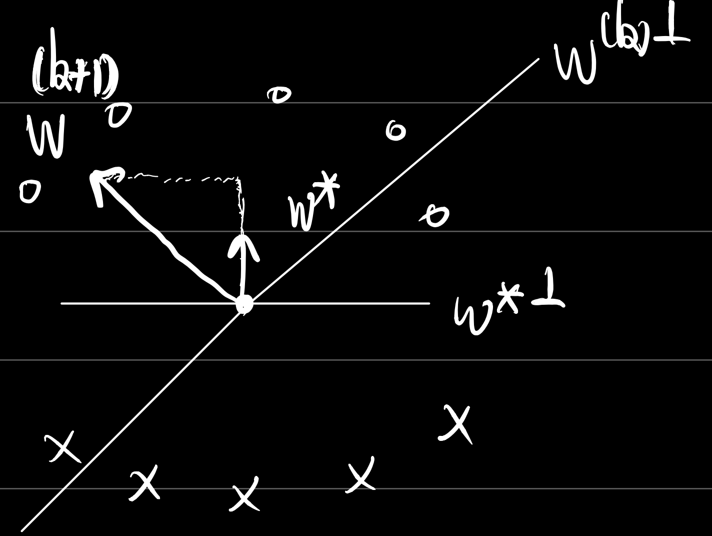
Theorem: Suppose

(1) the sample pts with labels $(x^{(1)}, y_1), \dots, (x^{(n)}, y_n)$ are linearly separable with const $\gamma > 0$.

(2) $x^{(1)}, \cdots, x^{(n)}$ are bounded, i.e. $\exists R > 0$, such that $\|x^{(i)}\| \leq R$

Then the perceptron learning algorithm makes at most $\frac{R^2}{\gamma^2}$ updates before returning a separating hyperplane.

Proof:



Let w^* be the unit vector in the definition of linear separability, then

$$y_j \cdot (w^* \cdot x^{(j)}) > \gamma > 0 \qquad j = 1, \cdots, n. \text{ ①}$$

Compute the projection of $w^{(k+1)}$ on w^*:

if $(x^{(i)}, y_i)$ is misclassified,

$$w^{(k+1)} \cdot w^* = \left(w^{(k)} + y_i x^{(i)}\right) \cdot w^*$$

$$= w^{(k)} \cdot w^* + y_i \left(x^{(i)} \cdot w^*\right)$$

$$\overset{①}{>} w^{(k)} \cdot w^* + \gamma$$

This shows: $w^{(k+1)} \cdot w^* > w^{(k)} \cdot w^* + \gamma$.

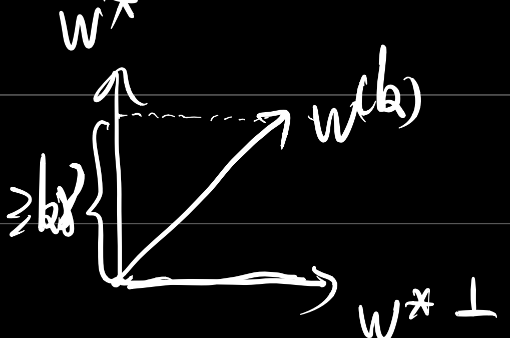
Similarly:

$$\left.\begin{aligned} w^{(k)} \cdot w^* &> \cancel{w^{(k-1)} \cdot w^*} + \gamma \\ \cancel{w^{(k-1)} \cdot w^*} &> \cancel{w^{(k-2)} \cdot w^*} + \gamma \\ &\vdots \\ \cancel{w^{(1)} \cdot w^*} &> \overset{=0}{w^{(0)} \cdot w^*} + \gamma = \gamma \end{aligned}\right\} \text{add}$$

Add the inequalities to get

$$w^{(k)} \cdot w^* > k\gamma > 0$$

Thus $\quad \|w^{(k)}\| \geq |w^{(k)} \cdot w^*| = w^{(k)} \cdot w^* > k\gamma > 0$



(to be continued)