Perceptron Algorithm Proof

Let's show that the perceptron algorithm converges in a finite number of updates.

Theorem

Let $x_1, \dots, x_n$ be a set of n training examples with the assumptions that

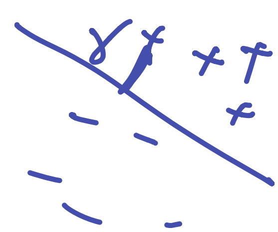
(1) $\|x_i\| \leq R \quad \forall i = 1, \dots, n.$

that is all training examples have their Euclidean norm bounded. $\left(\|x_i\| = \sqrt{\sum_{j=1}^{d} x_{ij}^2}\right)$

(2) there exist a linear classifier that correctly classifies all the training examples. In other words there exist a margin $\gamma > 0$ such that $\forall i = 1, \dots, n$

$$y_i \cdot x_i w^* \geq \gamma \qquad \|w^*\| = 1$$

γ ensures that each example is classified correctly with a finite margin.



then the perceptron algorithm makes at most

$$\frac{R^2}{\gamma^2} \text{ errors.}$$

in other words the algorithm converges in a finite number of updates or steps $K \leq \frac{R^2}{\gamma^2}$.

Proof:

We initialise the weight vector with zero.

that is $\vec{w}^1 = 0$

step 1: At an iteration $k+1$,
we can check how far is $\vec{w}^{k+1}$ from $\vec{w}^*$ by checking the dot product:

$\vec{w}^{k+1} \cdot \vec{w}^* = (\vec{w}^k + y_i \vec{x}_i) \cdot \vec{w}^*$

we want to show that it increases with each iteration

when we make a mistake at example $\vec{x}_i$

$\vec{w}^{k+1} \cdot \vec{w}^* = \vec{w}^k \vec{w}^* + y_i \vec{x}_i \vec{w}^*$

$$\boxed{\vec{w}^{k+1} \cdot \vec{w}^* \geq \vec{w}^k \vec{w}^* + \underline{\gamma} \quad \text{from } (1)}$$

digression

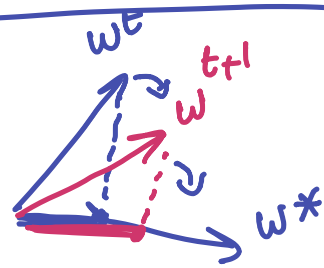
The inner (dot) product btw two vectors u & v is based on the projection of one into the other one.

$u \cdot v > 0$ similar direction.

$u \cdot v = 0$

$u \cdot v < 0$

Indeed



the inner (dot) product btw w^{k+1} and w^* increased.

By induction:

$\vec{w}^1 = 0$

$\vec{w}^2 \cdot \vec{w}^* \geq \vec{w}^1 w^* + \gamma$

$\vec{w}^3 \vec{w}^* \geq w^2 \cdot w^* + \gamma \geq w^1 w^* + \gamma + 2\gamma$

$\vdots$

$\vec{w}^{k+1} \vec{w}^* \geq k \cdot \gamma$

$\vec{w}^{k+1} \cdot \vec{w}^* \geq k \cdot \gamma$.

$\underbrace{\|w^{k+1}\|}_{\text{length of } \vec{w}^{k+1}} \times \underbrace{\|w^*\|}_{\text{length of } \vec{w}^*} \geq \vec{w}^{k+1} \times \vec{w}^*$

$\|w^*\| = 1$

$$\boxed{\|w^{k+1}\| \geq k \times \gamma}$$

(I)

step2: Now let's show that

$$\|w^{k+1}\|^2 = \|w^k + y_i \vec{x}_i\|^2 \quad \text{update of } \omega.$$

$$= \|w^k\|^2 + 2 y_i \vec{x}_i w^k + \|y_i \vec{x}_i\|^2$$

$$= \|w^k\|^2 + \underbrace{2 y_i \vec{x}_i \cdot \vec{w}^k}_{<0} + \underbrace{y_i^2}_{1} \underbrace{\|x_i\|^2}_{\leq R^2}$$

i was missclass in iteration k.

from (I)

$$\boxed{\|w^{k+1}\|^2 \leq \|w^k\|^2 + R^2}$$

upper bound for $\|w^{k+1}\|^2$!

by induction:

$$\|w^1\|^2 = 0$$

$$\|w^2\|^2 \leq 0 + R^2$$

$$\|w^3\|^2 \leq \|w^2\|^2 + R^2 \leq 2R^2$$

$$\vdots$$

$$\boxed{\|w^{k+1}\|^2 \leq kR^2} \quad (\text{II})$$

(I) $\wedge$ (II) $\|w^{k+1}\|^2 \leq kR^2$

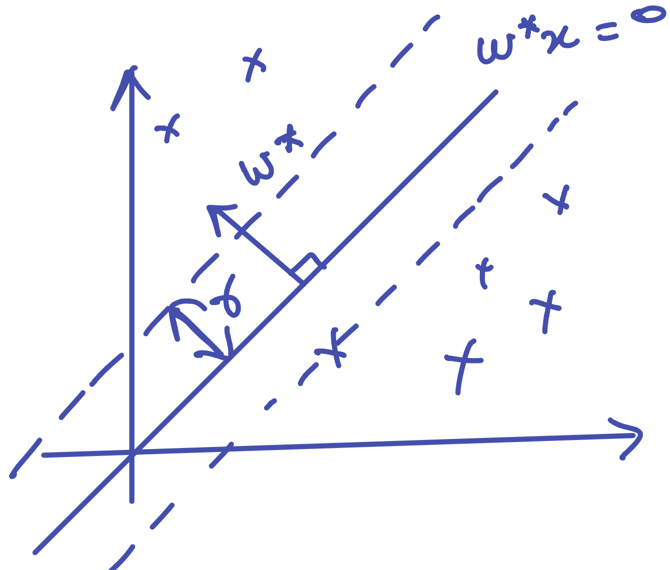
$$\|w^{k+1}\| \geq k\,\theta$$

therefore:

$$k^2\gamma^2 \leq \|w^{k+1}\|^2 \leq kR^2$$

$$\Rightarrow \quad k^2\gamma^2 \leq kR^2 \Rightarrow \quad k\gamma^2 \leq R^2$$

$$k \leq \frac{R^2}{\gamma^2}$$ ■



⚠ observe this bound is

1) - independent of the data.

2) - depends on the normalized margin.

3) - tight bound that is the number of updates can be $\frac{R^2}{\gamma^2}$ indeed in some cases.