PERCEPTRON LEARNING CONVERGENCE THEOREM

Recall: Perceptron learning updates weights ONLY if:

$$y = f(\bar{x} \cdot \bar{w}) \neq t$$

if so, the update is: $\bar{w}_{(NEW)} = \bar{w}_{(OLD)} + t\bar{x}$

2

Perceptron Learning Convergence theorem

"If there is a weight vector $\bar{w}^*$, such that $f(\bar{x}(p) \cdot \bar{w}^*) = t(p)$ for all p, then,
for any starting vector $\bar{w}$, the perceptron learning rule will converge to a weight vector (not necessarily $\bar{w}^*$) that gives the correct response for all training patterns, and will do so in a FINITE number of steps."

Originally, the training set is composed of:

$$F^+ = \{ x(p) \,/\, t(p) = +1 \}$$
$$\cup$$
$$F^- = \{ x(p) \,/\, t(p) = -1 \}$$

NOTE THAT FINDING A WEIGHT VECTOR FOR THIS TRAINING SET (SOLUTION) IS THE SAME AS FINDING ONE FOR:

$$\boxed{F^+ \cup -F^-} = F$$

where: $-F^- = \{ -x(p) \,/\, t(p) = -1 \}$

BUT NOW ALL THE TARGETS WILL BE +1

4

So, let's consider the process of perceptron training for such modified training set:

Let's consider ONLY pattern presentations that REQUIRE a weight upgrade, in a sequence, starting with $\bar{x}(0)$, which fails to satisfy the target with weights $w(0)$

$$\bar{x}(0) \cdot \bar{w}(0) \leq 0 \rightarrow y = -1 \Rightarrow y \neq 1 = t$$

MODIFICATION:

$$\bar{w}(1) = \bar{w}(0) + \bar{x}(0)$$

NOTE: We will assume $\theta = 0$, $\alpha = 1$ for simplicity then: this set of weights eventually fails:

$$\bar{X}(1) \cdot \bar{W}(1) \leq 0$$

MODIFICATION:

$$\bar{W}(2) = \bar{W}(1) + \bar{X}(1)$$

and so on.

IN GENERAL: $\bar{X}(k-1) \cdot \bar{W}(k-1) \leq 0$

causes the modification:

$$\bar{W}(k) = \bar{W}(k-1) + \bar{X}(k-1)$$

the whole CHAIN OF CHANGES, FROM $W(0)$ IS:

$$\bar{W}(k) = \underbrace{\underbrace{\bar{W}(0) + \bar{X}(0)}_{W(1)} + \bar{X}(1)}_{W(2)\ \ldots} + \ldots + \bar{X}(k-1) \quad \text{ⓐ}$$

WE WANT TO SHOW THAT K CANNOT BE ARBITRARILY LARGE

- ASSUME $\bar{W}^*$ EXISTS, such that $\bar{X} \cdot W^* > 0 \quad \forall X \in F$
- Now find the minimum:

$$m = \min\{\bar{X} \cdot \bar{W}^*\}$$ (pattern $\bar{X}$ that has the smallest projection on $\bar{W}^*$)

from ⓐ:

$$\bar{W}(k) \cdot \bar{W}^* = [\bar{W}(0) + \bar{X}(0) + \ldots + \bar{X}(k-1)] \cdot \bar{W}^*$$

$$\Rightarrow \bar{W}(k) \cdot \bar{W}^* = \bar{W}(0) \cdot \bar{W}^* + \underbrace{\bar{X}(0)\bar{W}^* + \ldots + \bar{X}(k-1)\bar{W}^*}_{k \text{ terms}}$$

then:

$$\bar{W}(k) \cdot \bar{W}^* \geq \bar{W}(0) \cdot \bar{W}^* + k\,m \quad \text{ⓑ}$$

because m is smaller or equal to any of the k terms in the sum.

6

Recall the

Cauchy - Schwarz Inequality:

$$(\bar{a} \cdot \bar{b})^2 \leq |\bar{a}|^2\,|\bar{b}|^2$$

or

$$|\bar{a}|^2 \geq \frac{(\bar{a} \cdot \bar{b})^2}{|\bar{b}|^2} \quad \left(\text{if } |\bar{b}|^2 \neq 0\right)$$

In this context:

$$|W(k)|^2 \geq \frac{(W(k) \cdot W^*)^2}{|W^*|^2} \quad \text{if } |W^*|^2 \neq 0 \ \text{(ok)}$$

substituting, from ⓑ:

$$|W(k)|^2 \geq \frac{(W(0)W^* + km)^2}{|W^*|^2} \quad \text{Ⓐ}$$

8

Now, consider the general UPDATE RULE:

$$\bar{W}(k) = \bar{W}(k-1) + \bar{X}(k-1) \quad \text{— ⓒ}$$

when $\bar{X}(k-1) \cdot \bar{W}(k-1) \leq 0$ ⓓ

Performing the inner product of each side of ⓒ with itself, and preserving the equality:

$$|\bar{W}(k)|^2 = |\bar{W}(k-1)|^2 + 2\,\bar{X}(k-1) \cdot \bar{W}(k-1) + |\bar{X}(k-1)|^2$$

then, because of ⓓ:

$$|W(k)|^2 \leq |\bar{W}(k-1)|^2 + |\bar{X}(k-1)|^2 \quad \text{ⓔ}$$

Let's assign: $M = \max\{|\bar{x}|^2\}$.
then, in:

$$|\bar{w}(k)|^2 \le |\bar{w}(k-1)|^2 + |\bar{x}(k-1)|^2$$
$$\le \underbrace{|\bar{w}(k-2)|^2 + |\bar{x}(k-2)|^2}_{C} + |\bar{x}(k-1)|^2$$
$$\le |\bar{w}(k-3)|^2 + |\bar{x}(k-3)|^2 + |\bar{x}(k-2)|^2 + |\bar{x}(k-1)|^2$$
$$\vdots$$
$$|\bar{w}(k)|^2 \le |\bar{w}(0)|^2 + \underbrace{|\bar{x}(0)|^2 + |\bar{x}(1)|^2 + \ldots + |x(k-1)|^2}_{k \text{ terms}}$$

$$\Rightarrow \boxed{|w(k)|^2 \le |\bar{w}(0)|^2 + kM} \quad \text{because } M \ge |\bar{x}(i)|^2 \text{ for any } i$$

B

Now: Combining A and B

$$\frac{(w(0)\cdot w^* + km)^2}{|w^*|} \le |w(k)|^2 \le |w(0)|^2 + kM$$

> Assume for simplicity that $\bar{w}(0) = \bar{0}$:

$$\Rightarrow \frac{(km)^2}{|w^*|^2} \le kM$$

$$\Rightarrow \boxed{k \le \frac{M\,|w^*|^2}{m^2}}$$

k has an upper bound!
k is not infinite!

If we assume that the (hypothetic) Solution vector w^* is of unit norm:

$$|w^*|^2 = 1$$

then: $$k \le \frac{M}{m^2}$$

then the maximum number of effective updates is M/m^2

Since w^* (and m) are not known, the above result cannot be used to predict the lenght of training.