Proof of PLA Convergence

Fábio Colaço (fabioiuri@live.com)

1 Introduction

This proof will be centred in calculating a lower and upper bound on the number of iterations needed to get an optimal set of weights that linearly divides the input set and classifies it correctly. If we succeed, then the PLA converges.

2 Assumptions

- **(A_1)** The input set X is linearly separable in a way that all classifications are correctly divided;
- **(A_2)** The PLA weight vector at iteration 0 is the zero vector: $w(0) = \vec{0}$.
- **(A_3)** The classification set is binary and assumes the following values: $y(t) \in \{-1, 1\}$.

  Lets also denote the set of optimal weights by $\boldsymbol{w^*}$, that exists by (A_1).

3 Upper Bound

Lets pick a friendly analytic formula to derive the upper bound for the proof:

$$\|w(t+1)\|^2 = \|w(t) + y(t)X(t)\|^2 \tag{1}$$

Since $w(t)$ missclassifies $X(t)$, an update on the weights is needed: $w(t+1)$. Using simple mathematics on the PLA weight vector update formula, we can apply the Euclidean norm and square it, obtaining the same value.

$$\begin{aligned} \|w(t+1)\|^2 &= \|w(t)\|^2 + \|y(t)X(t)\|^2 + 2\,\|w(t)\|\,\|y(t)X(t)\| \qquad (2) \\ &\leq \|w(t)\|^2 + y(t)^2\,\|X(t)\|^2 \qquad (3) \end{aligned}$$

The result obtained in (3), comes from the analysis of the reduction of $2\,\|w(t)\|\,\|y(t)X(t)\|$ presented in Corollary 1. Lets rewrite:

$$\begin{aligned} 2\,\|w(t)\|\,\|y(t)X(t)\| &= 2y(t)\,\|w(t)\|\,\|X(t)\| \\ &= \zeta(t) \end{aligned}$$

Corollary 1. Given that the product between 2 values with different signs will always be negative and that the product between 2 values in which one of them is zero, will also be zero: $2\|w(t)\|\,\|y(t)X(t)\|$ can never exceed 0.

Proof. By inspection, it's known that $w(t)$ missclassifies $X(t)$, and given that $y(t) \in \{-1, 1\}$ (by A_3), we can do a proof by exhaustion:

$$\begin{aligned} y(t) = -1 &\implies \|w(t)\|\,\|X(t)\| \geq 0 \implies \zeta(t) \leq 0 \\ y(t) = +1 &\implies \|w(t)\|\,\|X(t)\| \leq 0 \implies \zeta(t) \leq 0 \end{aligned}$$

□

Due to $y(t)^2$ being always 1, we also can remove it from the rightmost term:

$$\begin{aligned} \|w(t+1)\|^2 &\leq \|w(t)\|^2 + \|X(t)\|^2 \\ &\leq tR^2 \end{aligned} \tag{4}$$

Where $R = max_{1 \leq n \leq N} \|X(n)\|$.

The t multiplier comes by induction, since we only have 2 terms: the R^2 (will appear on all expansions), and the $\|w(t)\|^2$, which can be expanded and will lead to t subsequent recursive calls, stopping on $w(0)$. Because we are only worried about defining an upper bound, using R as the max will allow us to do it easily, without being too loose on the calculated bound.

Lemma 1. tR^2 is an upper bound of $\|w(t+1)\|^2$.

4 Lower Bound

Now let us try to get a lower bound:

$$w^* w^T(t+1) = w^*[w^T(t) + y(t)X^T(t)] \tag{1}$$
$$= w^* w^T(t) + y(t)[w^{*T} X(t)] \tag{2}$$
$$\geq w^* w^T(t) + \rho \tag{3}$$

Where $\rho = \min_{1 \leq n \leq N} y(n)[w^{*T} X(n)]$.

Corollary 2. $\forall n \in [1, N] : \rho \leq y(n)[w^{*T} X(n)] \wedge \rho \neq 0$. Therefore, the step (3) is perfectly valid.

Proof. Looking at the definition of ρ, we can see that it is defined as the minimum of the set $y(n)[w^{*T} X(n)]$. With that, we can trivially infer that ρ will always be less or equal than any member of the same set. Also, ρ can never be 0 because y(n) is never 0 by A_3, nor $w^{*T} X(n)$ is 0 because w^* is the optimal set, therefore must classify $X(n)$ correctly - $Sign(X(n)) = -$ or $Sign(X(n)) = +$. □

$$w^* w^T(t+1) \geq t\rho \tag{4}$$

Such as applied on the last step of the upper bound case, we use induction to add the t multiplier.

By applying the Cauchy-Schwarz inequality:

$$\begin{aligned}
\|w(t+1)\|^2 \|w^*\|^2 &\geq [w^* w^T (t+1)]^2 \\
\|w(t+1)\|^2 \|w^*\|^2 &\geq (t\rho)^2 \\
\|w(t+1)\|^2 \|w^*\|^2 &\geq t^2\rho^2 \\
\|w(t+1)\|^2 &\geq \frac{t^2\rho^2}{\|w^*\|^2} \qquad (5)
\end{aligned}$$

Lemma 2. $\frac{t^2\rho^2}{\|w^*\|^2}$ is a lower bound of $\|w(t+1)\|^2$.

5 Concluding

Theorem 1. Given the assumptions we worked with, the Perceptron Learning Algorithm always converges, due to the number of iterations t needed to get from $w(0)$ to w^*, being bounded by a finite and real value $\frac{R^2\|w^*\|^2}{\rho^2}$.

Proof. By Lemma 1 and Lemma 2, we know that:

$$\frac{t^2\rho^2}{\|w^*\|^2} \leq \|w(t+1)\|^2 \leq tR^2 \qquad (1)$$

$$\frac{t\rho^2}{\|w^*\|^2} \leq \frac{\|w(t+1)\|^2}{t} \leq R^2 \qquad (2)$$

$$t\rho^2 \leq \frac{\|w(t+1)\|^2 \|w^*\|^2}{t} \leq R^2 \|w^*\|^2 \qquad (3)$$

$$t \leq \frac{\|w(t+1)\|^2 \|w^*\|^2}{t\rho^2} \leq \frac{R^2 \|w^*\|^2}{\rho^2} \qquad (4)$$

$$t \leq \frac{R^2 \|w^*\|^2}{\rho^2} \qquad (5)$$

Since all R, $\|w^*\|$ and ρ terms are finite, $\frac{R^2\|w^*\|^2}{\rho^2}$ is finite and real (ρ is always $\neq 0$, by Cor. 2). $\square$