For a misclassified example:-

$$\text{sign}(y(t)) \neq \text{sign}(w^T(t)\, X(t))$$

Expected output Actual output

Remember that $y(t) \to -1$ or 1 only

$\therefore y(t) . (w^T(t) X(t)) < 0$ -ve

// we will update the weights using this equation

$$\vec{w}_n = \vec{w}_o + yX$$

// After update @ iteration $t+1$

$$y(t) . w^T(t+1) . X(t) = y(t) [w^T(t) + y(t) X(t)]^T . X(t)$$

$$= y(t) [w(t) X(t) + y^T(t)\, X^T(t) X(t)] \rightarrow y^T(t) = y(t)$$

Note ↑

$$= y(t) w(t) X(t) + y^2(t)\; X^T(t) X(t)$$

$$\| X(t) \|^2$$

// Notes

$y^2(t) = 1$ always $\because (-1)^2 = (1)^2 = 1$

$X^T(t) X(t) = \| X \|^2 \rightarrow$ magnitude

$$= y(t) w(t) X(t) + (1) \| X(t) \|^2 = y(t) w^T(t+1) X(t)$$

$$\therefore \underbrace{y(t) w^T(t+1) X(t)}_{\text{new update}} \geq \underbrace{y(t) w(t) X(t)}_{\text{old value}} + \geq 0$$

which indicates the new update will move always in the correct direction #

Now there is an interesting Question →

Does the perceptron always converge ?

lets prove this #

Perceptron Convergence Proof:-

[1] we assume that the data is linearly separable ✓✓

$\therefore \exists w^*$ that separates the training data.

$\therefore$ we need to prove that the algorithm will find a linear separator (w) after finite # of iterations. $t \leq \square$ ✓✓

→ this can be done if we found an upper bound for the # of iterations.

1. let $\rho = \min_n (y_n \cdot w^{T*} x_n)$ → ρ هي اقل قيمة الجبرية الـ ه بتفصل الـ data

$\rho > 0$ Because we assume that w^* can separate the data so there will be no missclassified data.

2. $w^T(t) w^* = w^T(t-1) w^* + y(t-1) x^T(t-1) w^*$

we got this equation from the update equation $w_n = w_{old} + yX$

$\because y(t-1) x^T(t-1) w^*$ is one sample

$\because \rho = \min (y_n w^{T*} x_n)$

$\therefore \rho \leq y(t-1) x^T(t-1) w^*$

$\therefore w^T(t) w^* \geq w^T(t-1) w^* + \rho$ → (1)

$w_n = w_0 + yX$ $w(0) = 0$

Assume $w(0) = 0$, if we keep recursive calls

$\because w^T(t-1) w^* = w^T(t-2) w^* + \rho$

$\cancel{w^T(0) w^*} + \rho$

$\therefore w^T(t) w^* = (w^T(t-2) w^* + \rho) + \rho = w^T(t-2) w^* + 2\rho$

→ for t iterations

[2] $w^T(t) w^* \geq t\rho$ + 0 → w(0) (2) $t\rho$

$\because w(t) = w(t-1) + x(t-1)\,y(t-1)$ $\quad -1 \quad +1 \quad 1$

$\therefore \|w(t)\|^2 = \|w(t-1) + x(t-1)\,y(t-1)\|^2$

$= \|w(t-1)\|^2 + y^2(t-1)\|x^2(t-1)\| + \underbrace{2\,w^T(t-1)\,x(t-1)\,y(t-1)}_{\text{less than } 0}$

$y^2(t-1) = 1$ $\quad \|x(t-1)\|^2$

• Because we did an update step, which mean that the sign of $y(t-1) \neq \text{sign}(w^T(t-1)\,x(t-1))$

$w(0) = 0$ $\quad \|w(t-1)\|^2 \leq$

$\therefore$ ~~w(t)~~

$\therefore \|w(t)\|^2 \leq \|w(t-1)\|^2 + \|x(t-1)\|^2$

$\|w(t-1)\|^2 + \|x(t-1)\|^2$

// Again using recursive calls

$\|w(t-1)\|^2 \leq \|w(t-2)\|^2 + \|x(t-2)\|^2$

$\therefore \|w(t)\|^2 \leq \|w(t-2)\|^2 + \|x(t-2)\|^2 + \|x(t-1)\|^2$

we know that $w(0) = 0$ $\quad R^2 + R^2$ $\quad 0 + \sum^{t} \|x(t-1)\|^2$

$\therefore$ let $R = \max_n \|x_n\|$ $\quad i = 9, 5, 4, \ldots t$

$\therefore \|w(t)\|^2 \leq t R^2$ مظبوط ✓

$\because w^T(t)\,w^* = \|w(t)\|\,\|w^*\| \cos\theta$ $\quad -1, 1$

$\therefore \|w(t)\|\,\|w^*\| \geq w^T(t)\,w^*$

تعالى أشرحلك:

الـ $\cos\theta$ رقم بين الـ 1 والـ -1، عشان نخلي رقم بعدما يتضرب $G \cdot \cos\theta$ = رقم تاني

ده معناه إن الرقم التاني أصغر عشان لما بنضرب في 0 وهنا إحنا اكيد بنصغر

$\because w^T w^* \geq t\rho$

$\therefore \cancel{\|w(t)\|}\|w^*\| \geq t\rho$

$\because \|w(t)\|^2 \leq tR^2$

$\therefore \|w(t)\| \leq \sqrt{t}\,R$ ✓

$\therefore \cancel{\sqrt{t}}\,R\,\|w^*\| \geq \rho t$

$\therefore \sqrt{t} \leq \dfrac{\cancel{\sqrt{t}}\,R\|w^*\|}{\rho}$

$\therefore \sqrt{t} \leq \dfrac{R\|w^*\|}{\rho}$

$$\boxed{\therefore t \leq \frac{R^2\|w^*\|^2}{\rho^2}}$$

which is an upper bound to t

→ we do not know exactly its value

But we know that the algorithm will converge #

→ we don't say that the algorithm converge @ w^*

But we say that it will converge @ some $w(t)$

7 > 5

a 7 7

a 7 5

θ

t

Converge

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$