## ⊙ Proof of Perceptron

→ The way we are gonna prove that this algorithm converges is by following idea,

| |
|---|
| we say that perceptron algorithm cannot make more than a certain mistakes. |

↳ Idea of convergence.

→ If we think about it, it makes sense that if the mistakes are gonna be infinite, then it will never converge. So, if we can prove that the mistakes are gonna be finite, then we can say it converges.

### Analysis of mistakes of Perceptron

→ we know that, an update in the perceptron happens only when a mistake happens.

→ Let's say $w^l$ is the current guess, and then a mistake happens wrt $(x, y)$.

Then, $w^{l+1} = w^l + x \cdot y$

Let's understand the growth or length of this update.

$$\|w^{l+1}\|^2 = \|w^l + x \cdot y\|^2 = (w^l + xy)^T (w^l + xy)$$

$$= \|w^l\|^2 + \underbrace{2 \cdot (w^{l^T} x) y}_{\leq 0 \text{ (because mistake)}} + \underbrace{\|x\|^2 \cdot \underbrace{y^2}_{+1 \text{ assumed } \in \mathbb{R}}}_{\leq R^2}$$

→ If we observe, why we said 2nd term $(2\cdot(w^{lT}x)y)\cdot \leq 0$, it is because that, $w^{lT}x$ is the prediction (−1/+1), and $y$ is actual (−1/+1).

And considering update happens only on mistake, these both have to be opposite values in case, and the total value will turn out to be negative.

Hence $2(w^{lT}x)y \leq 0$, ⟶ ignored

So we can upper bound whole thing as,

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

→ This means that the new weight length will be utmost previous length + $R^2$,

→ same applies for previous length.

$$\|w^{l+1}\|^2 \leq (\|w^{l-1}\|^2 + R^2) + R^2$$

..... If we continue, we get (for $l$ mistakes)

$$\|w^{l+1}\|^2 \leq \underbrace{\|w^0\|^2}_{=0} + lR^2$$

$$\therefore \boxed{\|w^{l+1}\|^2 \leq lR^2} \text{ ——— } (1)$$

→ No. of mistakes

→ So we say that the max length it can grow is '$l$' times $R^2$, and $l$ is the no. of updates/mistakes.

→ From the other side, to understand about ($w^{l+1}$) quantity. let's use $w^*$.

we have,

$$w^{l+1} = w^l + x \cdot y$$

$$(w^{l+1})^T \cdot w^* = (w^l + x \cdot y)^T \cdot w^*$$

$$= w^{l\,T} w^* + \underbrace{(w^{*T} x) \cdot y}_{\geq \gamma}$$

(based on $w^*$ assumption)

$$(w^{l+1})^T \cdot w^* \geq w^{l\,T} w^* + \gamma$$

→ we can take same argument for $w^l$,

$$(w^{l+1})^T \cdot w^* \geq (w^{l-1\,T} w^* + \gamma) + \gamma$$

..... ~~after~~ $l$ mistakes.

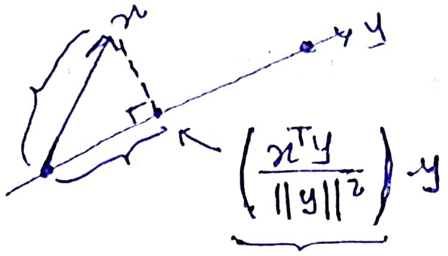
$$(w^{l+1})^T \cdot w^* \geq \underbrace{w^{0\,T} w^*}_{0} + l\gamma$$

$0$ (As $w^0 = 0$)

$$\boxed{(w^{l+1})^T \cdot w^* \geq l\gamma} \rightarrow (2)$$

→ It says that as we make mistakes, the dot product with $w^*$ is increasing.

Now, If we try to make use of these two eqn.
First we can recall the Cauchy-Schwartz inequality,
for any $x, y$



$$\left\| \left(\frac{x^T y}{\|y\|^2}\right)\cdot y \right\|^2 \le \|x\|^2$$

Pythagorous theorem...

$$(x^T y)^2 \cdot \frac{\|y\|^2}{\|y\|^4} \le \|x\|^2$$

$$\Rightarrow \boxed{(x^T y)^2 \le \|x\|^2 \cdot \|y\|^2}$$

From eq (2),

$$l\gamma \le (w^{l+1})^T \cdot w^*$$ , Apply square.

$$l^2\gamma^2 \le \left[(w^{l+1})^T \cdot w^*\right]^2 \le \|w^{l+1}\|^2 \cdot \underbrace{\|w^*\|^2}_{1}$$ From CS

$$l^2\gamma^2 \le \|w^{l+1}\|^2$$

$$\therefore \boxed{\|w^{l+1}\|^2 \ge l^2\gamma^2}$$ → From (3)

$$l^2\gamma^2 \underset{\text{From (3)}}{\le} \|w^{l+1}\|^2 \underset{\text{From (1)}}{\le} lR^2$$

$$l^2\gamma^2 \le lR^2$$

$$\boxed{l \le R^2/\gamma^2}$$

→ Radius Margin Bound.
→ $l$ — no. of mistakes.

→ So, by observing $\boxed{l \leq R^2/\gamma^2}$,

This tells us that $l$ has to be

utmost of $R^2/\gamma^2$,

and at the first assumption, we said $\gamma > 0$.

And so we clearly say '$l$' is gonna be

finite.

And $l$ is the number of mistakes, so we can finally say that the no. of mistakes is a finite quantity & Hence the "Perceptron will be converged."

→ If we just take linearly seperable assumption without considering $\gamma$, i.e, $\gamma = 0$.

Then $l \leq R^2/0 \leq \infty$ (Infinity).

In this case it is not certain $\varepsilon$. hence it can't be converged.

$\boxed{\text{Perceptron}}$

no. of mistakes $(l) \leq \frac{R^2}{\gamma^2}$

- under assumptions,

→ $\|x_i\|^2 \leq R^2$

→ Dataset — linearly seperated with margin '$\gamma$'

$$(w^{*T} x_i)\, y_i \geq \gamma \quad \forall\, i, \quad \gamma > 0$$