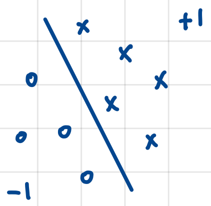
Perceptron Convergence

Recall

Given a linearly seperpable dataset, the perceptron algorithm always converge to a hyperplane that divides the two categories

Hyperplane $\mathcal{H}$ $\{x : w^T x = 0\}$



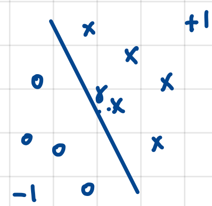
Algorithm...

When classified correctly... $\begin{cases} y = +1 \Rightarrow w^T x > 0 \\ y = -1 \Rightarrow w^T x < 0 \end{cases}$ $\quad y w^T x > 0$

When classified incorrectly... $\begin{cases} y = +1 \Rightarrow w \leftarrow w + x \\ y = -1 \Rightarrow w \leftarrow w - x \end{cases}$ $\quad w \leftarrow w + yx$

Margin : Min distance to hyperplane

$$\gamma = \min_{x,y \in D} |x^T w^*| > 0$$



Convergence Proof.

1. Data linearly seperable

   $\exists\, w^*$ s.t. $\forall (x,y) \in D$. $\quad y w^{*T} x > 0$

2. Assume, WLOG

   $\|w^*\| = 1$

   $\|x_i\| \leq 1 \quad \forall i$ Shrink Covariates so that max value ≤ 1

Claim: w approaches w^* with every update (i.e. $w^T w^*$ decreasing when adjusted for $\|w\|$)

update: $w \leftarrow w + yx$

update performed when $y w^T x \leq 0$

$$w'^T w^* = (w^T + y x^T) w^*$$

$$= w^T w^* + y x^T w^*$$

We know not only $y x^T w^* > 0$ (def of w^*)

but also $y x^T w^* \geq \gamma$

$\geq w^T w^* + \gamma \qquad (*)$

$w^T w$ increases by atleast γ by every update

$$\gamma = \min_{x,y \in D} |x^T w^*| > 0$$

After m updates

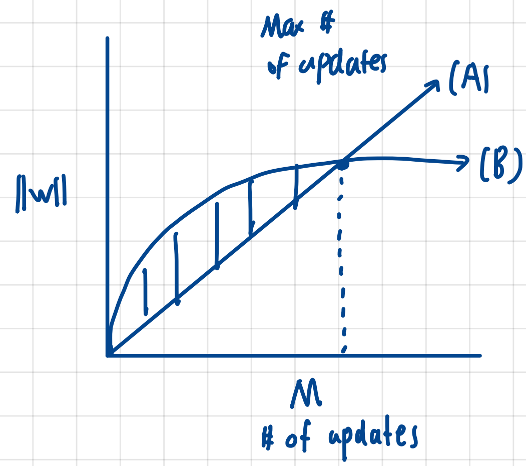
After 0 update, $w^T w^* = 0$

$w'^T w^* \geq w^T w^* + \gamma \qquad (*)$

$\therefore\ w^T w^* > m\gamma \Rightarrow \|w^T \cdot w^*\| \geq m\gamma$

$\Rightarrow \|w^T\| \underbrace{\|w^*\|}_{=1} \geq m\gamma$ (Cauchy-Schwartz)

$\Rightarrow \|w\| \geq m\gamma \qquad (A)$



$w'^T w' = (w + yx)^T (w + yx)$

$= w^T w + \underbrace{2yw^T x}_{<0} + \underbrace{y^2 x^T x}_{\leq 1}$

$yw^T x < 0$. If update is made, classification must be wrong

$y^2 = 1 \quad y^2 x^T x \leq 1$ because $\|x\| \leq 1$.

$\leq w^T w + 1 \qquad (**)$

By (**), $w^T \cdot w$ increases by at most 1

$\Rightarrow \|w\| \leq \sqrt{m} \qquad (B)$

$\therefore \sqrt{m} \geq m\gamma$

$\frac{1}{\sqrt{m}} \geq \gamma$

$\sqrt{m}\,\gamma \leq 1$

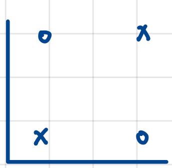
$\boxed{m \leq \frac{1}{\gamma^2}}$

You cannot make that many mistakes

Algorithm cvgs in finite step.

Dataset with larger γ converge faster

Perceptron cannot solve XOR problem



No hyperplane that fits thru this