THEOREM:-

Given a set of training example that are linearly separable through origin. Show that initialization of θ doesn't impact perceptron algorithm's ability to eventually converge.

GURUKUL

Assumptions:-

(1) There exists θ^* s.t $\boxed{\dfrac{y^{(i)}(\theta^* \cdot x^{(i)})}{\|\theta^*\|} \geq \gamma}$ for all $i = 1, \dots, n$

and for some $\gamma \geq 0$.

(2) All the examples are bounded $\boxed{\|x^{(i)}\| \leq R}$ $i = 1 \dots n$.

If theta is initialized to 0, we can show by induction that

$$\boxed{\theta^k \cdot \frac{\theta^*}{\|\theta^*\|} \geq k\gamma.} \quad \text{--- (i)}$$

Base class ($k=0$).

$$\theta^0 \cdot \frac{\theta^*}{\|\theta^*\|} = 0.$$

induction class

Assume this holds for some k

$$\theta^k \cdot \frac{\theta^*}{\|\theta^*\|} \geq k\gamma.$$

Now show it holds for $k+1$ given it holds for k.

$$\theta^{(k+1)} \cdot \frac{\theta^*}{\|\theta^*\|} = (\theta^k + \alpha y^{(i)} x^{(i)}) \cdot \frac{\theta^*}{\|\theta^*\|}.$$

$$\geq \frac{\theta^k \cdot \theta^*}{\|\theta^*\|} + \gamma.$$

$$\geq (k+1)\gamma.$$

Now if we initialize θ to general (not necessarily 0) $\theta^{(0)}$ then

$$\boxed{\theta^{(k)} \cdot \frac{\theta^*}{\|\theta^*\|} \geq a + k\gamma} \quad \text{--- (ii)}$$

GURUKUL

Sure find a g c $\boxed{a = \theta^{(0)} \cdot \frac{\theta^*}{\|\theta^*\|}}$

If θ is initialized to 0 we can show by induction that

$$\boxed{\|\theta^{k}\|^2 \le kR^2}.$$

i.e

$$\|\theta^{(k+1)}\|^2 = \|\theta^{(k)} + y^{(i)}x^{(i)}\|^2$$
$$= \|\theta^{(k)}\|^2 + 2\,\theta^{(k)} y^{(i)} x^{(i)} + \|x^{(i)}\|^2$$
$$= \|\theta^{(k)}\|^2 + 2y^{(i)}(\theta^{(k)} \cdot x^{(i)}) + \|x^{(i)}\|^2.$$

we know $\|x^{(i)}\| \le R$. and remove for inequality.
$\because y^{(i)}$

$$\|\theta^{(k+1)}\|^2 \le \|\theta^{(k)}\|^2 + \|x^{(i)}\|^2$$
$$\le \|\theta^{(k)}\|^2 + R^2.$$
$$\le kR^2 + R^2$$
$$= (k+1)R^2$$

which verifies induction.

Now

If we initialize θ to general then :-

$$\boxed{\|\theta^{(k)}\|^2 \le kR^2 + c^2} //$$

Determine c^2 in terms of $\theta^{(0)}$.

$$c^2 = \|\theta^{(0)}\|^2.$$

from above inequality, we can derive inequality.
$\|\theta^{(k)}\| \le c + \sqrt{k}R$ using $\downarrow$.

$$\|\theta^{k}\|^2 \le kR^2 + c^2 \le (\sqrt{k}R + c)^2. \quad \Big| \because a^2 + b^2 \le (a+b)^2.$$

$$\boxed{\therefore \|\theta^{k}\| \le c + \sqrt{k}R}.$$

and using

$$\frac{\theta^k}{\|\theta^k\|} \cdot \frac{\theta^*}{\|\theta^*\|} \le 1.$$

$$\text{or, } \frac{k\gamma}{\sqrt{k}R} \le 1 \qquad \left(\because \frac{\theta^k \cdot \theta^*}{\|\theta^*\|} \ge k\gamma, \ \|\theta^k\|^2 \le kR^2 \right.$$

$$\text{or, } \sqrt{k} \cdot \frac{\gamma}{R} \le 1$$

$$\Rightarrow \boxed{k \le \frac{R^2}{\gamma^2}}$$

In case where we initialized θ to general $\theta^{(0)}$, use inequality for $\frac{\theta^{(k)} \cdot \theta^*}{\|\theta^*\|}$ above and inequality $\|\theta^{(k)}\| \le ...$

to derive a bound on number of iteration k.

using

$$\boxed{\frac{\theta^k}{\|\theta^k\|} \cdot \frac{\theta^*}{\|\theta^*\|} \le 1.} \quad \text{--- (1)}$$

and if we initialized $\theta \to \theta^{(0)}$ then

$$\frac{\theta^k \cdot \theta^*}{\|\theta^*\|} \ge a + k\gamma. \quad \text{--- } (*)$$

$$\|\theta^k\| \le c + \sqrt{k}R \to (**)$$

so the inequality can be deduced from fact that lower bound of $(*)$ should still satisfy inequality (1) and upperbound of $(**)$ should ~~we~~ still make value smaller and as a result inequality satisfied.

Thus

$$\frac{a + k\gamma}{c + R\sqrt{k}} \le 1.$$

$$\text{or, } a + k\gamma \le c + R\sqrt{k}.$$

$$\text{or, } k\gamma - R\sqrt{k} + a - c \le 0$$

or, let $\sqrt{k} = t \Rightarrow t^2 = k$. and it ~~red~~ ~~dud~~ deduces to:-

and, $\gamma t^2 - Rt + a - c \le 0$.

solving for t.

$$t \le \frac{R \pm \sqrt{R^2 - 4\gamma(a-c)}}{2\gamma} \le \frac{R \pm \sqrt{R^2 + 4\gamma(c-a)}}{2\gamma}$$

and $k = t^2$

$$\le \left(\frac{R \pm \sqrt{R^2 + 4\gamma(c-a)}}{2\gamma}\right)^2 = \frac{(R \pm \sqrt{R^2 + 4\gamma(c-a)})^2}{4\gamma^2}$$

and for upperbound taking positive sign.

$$\boxed{\therefore k \le \frac{(R + \sqrt{R^2 + 4\gamma(c-a)})^2}{4\gamma^2}}$$