Perceptron Convergence Theorem

For any finite set of linearly separable labelled examples, the perceptron learning algorithm will halt in a finite number of steps

Let us assume we have a dataset X with m observations & n features & the corresponding targets be Y $\qquad y_i \in \{-1, +1\}$

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} & 1 \\ x_{21} & x_{22} & \cdots & x_{2n} & 1 \\ \vdots & & & & \\ x_{m1} & x_{m2} & \cdots & x_{mn} & 1 \end{pmatrix} \qquad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \qquad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

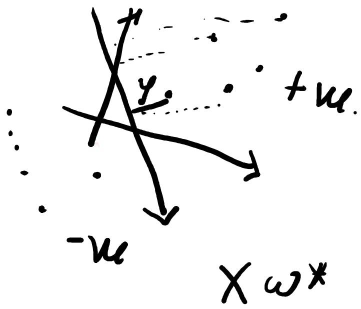
We have unknown weights as below

$$\omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \\ b \end{pmatrix}$$

We assume ω^* such that for any observation $x_i \in X$

$$y_i (x_i \omega^*) > 0$$

We assume $\|\omega^*\| = 1 \Rightarrow \boxed{(\omega^*)^T \omega^* = 1}$



Further, $\boxed{\gamma = \min \{ \underbrace{x_1}_{1 \times (n+1)} \underbrace{\omega^*}_{(n+1) \times 1}, x_2 \omega^*, \cdots x_m \omega^* \}}$

We have previously used the Gradient update rule.

$$\omega_{k+1} = \omega_k + \alpha X^T Y$$

Suppose for ω_k (i.e. weight at k^{th} step); $y_i (x_i \omega_k) < 0$ for some $i \leq m$. We will update ω_k to ω_{k+1}.

$$\omega_{k+1}^T \omega^* = (\omega_k + \alpha X^T Y)^T \omega^*$$

$$= (\omega_k^T + \alpha Y^T X) \omega^*$$

$$\boxed{\omega_{k+1}^T \omega^* = \omega_k^T \omega^* + \alpha Y^T X \omega^*}$$

$$\boxed{w_{k+1}^T w^* = w_k^T w^* + \alpha Y^T X w^*}$$

We know $Xw^* = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} w^* = \begin{pmatrix} x_1 w^* \\ x_2 w^* \\ \vdots \\ x_m w^* \end{pmatrix} > \gamma \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

$$\Rightarrow Xw^* > \gamma \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\Rightarrow Y^T X w^* > \gamma Y^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\Rightarrow Y^T X w^* > \gamma \left(\sum_{i=1}^{m} y_i\right)$$

$$\boxed{\alpha > 0}$$

$$\Rightarrow \alpha Y^T X w^* > \gamma \alpha \left(\sum_{i=1}^{m} y_i\right)$$

$$\Rightarrow \underbrace{w_k^T w^* + \alpha Y^T X w^*} > \gamma \alpha \left(\sum_{i=1}^{m} y_i\right) + w_k^T w^*$$

$$\Rightarrow w_{k+1}^T w^* \geq \underbrace{\gamma \alpha \left(\sum_{i=1}^{m} y_i\right)} + w_k^T w^*$$

for $k = 0, 1, \ldots M$ we get

$$w_1^T w^* \geq \underbrace{\gamma \alpha \left(\sum_{i=1}^{m} y_i\right)} + w_0^T w^* = C + w_0^T w^*$$

$$w_2^T w^* \geq C + w_1^T w^* \quad \text{where } C = \gamma \alpha \left(\sum_1 y_i\right)$$

$$\Rightarrow w_2^T w^* \geq C + \underbrace{C + w_0^T w^*} = w_0^T w^* + 2C$$

$$\Rightarrow w_3^T w^* \geq w_2^T w^* + C = w_0^T w^* + 3C$$

$$\vdots$$

$$w_M^T w^* \geq \boxed{M.C + w_0^T w^*}$$

$$\Rightarrow w_M^T w^* \geq MC$$

$$\Rightarrow \boxed{w_M^T w^* \geq M \alpha \gamma \left(\sum_{i=1}^{m} y_i\right)} \Rightarrow$$

$$(w^*)^T w^* = 1$$

$w_k \to (n+1) \times 1$
$X \to m \times (n+1)$
$Y \to m \times 1$

$$w_{k+1}^T w_{k+1} = \underbrace{(w_k + \alpha X^T Y)}^T (w_k + \alpha X^T Y)$$

$$= w_k^T w_k + \alpha w_k^T X^T Y + \alpha Y^T X w_k + \alpha^2 Y^T X X^T Y$$

$$w_{K+1} w_{K+1} = (w_K + \alpha X^T Y) (w_K$$

$$= \underline{\underline{w_K^T}} w_K + \alpha w_K^T X^T Y + \alpha Y^T X w_K + \alpha^2 Y^T X X^T Y$$

$$\boxed{w_{K+1}^T w_{K+1} = w_K^T w_K + 2\alpha \underline{\underline{Y^T X}} w_K + \alpha^2 \underline{\underline{Y^T X X^T}} Y}$$

For w^*, $Y^T X w^*$ is maximum $\Rightarrow$ $2\alpha Y^T X w_K \le 2\alpha Y^T X w^*$

$$\Rightarrow \boxed{2\alpha Y^T X w_K \le 2\alpha K_1}$$

where $K_1 = Y^T X w^*$

We assume $\underline{\underline{K_2}} = \underline{\underline{Y^T X X^T Y}}$

[Y^T is of order $1 \times m$
X is of order $m \times (n+1)$
X^T is of order $(n+1) \times m$
Y is of order $m \times 1$]

We get

$$w_{K+1}^T w_{K+1} \le w_K^T w_K + 2\alpha K_1 + \alpha^2 K_2$$

We can use recursively for $K = 0, 1, \dots M$

$$w_1^T w_1 \le \underline{\underline{w_0^T w_0}} + K \quad \text{where } K = 2\alpha K_1 + \alpha^2 K_2$$

$$w_2^T w_2 \le \underline{\underline{w_1^T w_1}} + K \le w_0^T w_0 + 2K$$

$$w_3^T w_3 \le w_2^T w_2 + K \le w_0^T w_0 + 3K$$

$w_0^T w_0 \le M w_0^T w_0$

$\vdots$

$$w_M^T w_M \le \underline{\underline{w_0^T w_0}} + MK. \Rightarrow \boxed{w_M^T w_M \le M w_0^T w_0 + M\alpha(2K_1 + \alpha K_2)}$$

$$[\because w_0^T w_0 + MK \le M w_0^T w_0 + MK]$$

We have.

1) $\omega_M^T \omega^* \geq M\alpha\gamma\left(\sum_{i=1}^{m} y_i\right)$ ✓

2) $\omega_M^T \omega_M \leq M\omega_0^T\omega_0 + M\alpha(2K_1 + \alpha K_2)$ ✓

Now $\omega_M^T \omega^* = \|\omega_M^T\| \, \|\omega^*\| \cos\theta$ [definition of dot product]

$\leq \|\omega_M^T\|$ $\left[\because \underbrace{\|\omega^*\| = 1}\ \&\ 0 \leq \cos\theta \leq 1\right]$

$\Rightarrow \|\omega_M^T\| \, \|\omega^*\| \cos\theta = \|\omega_M^T\| \cos\theta$

$\Rightarrow 0 \leq \|\omega_M^T\| \cos\theta \leq \|\omega_M^T\|$

$$\Rightarrow \|\omega_M^T\| \geq M\alpha\gamma\left(\sum_{i=1}^{m} y_i\right) ✓$$

$$\Rightarrow \sqrt{\omega_M^T \omega_M} \geq M\alpha\gamma\left(\sum_i y_i\right)$$

$$\Rightarrow \omega_M^T \omega_M \geq M^2\left[\alpha\gamma\left(\sum_i y_i\right)\right]^2$$

$$\Rightarrow M\omega_0^T\omega_0 + M\alpha(2K_1 + \alpha K_2) \geq M^2\left[\alpha\gamma\left(\sum_i y_i\right)\right]^2$$

$$\Rightarrow \frac{\omega_0^T\omega_0 + \alpha(2K_1 + \alpha K_2)}{\left[\alpha\gamma\left(\sum_i y_i\right)\right]^2} \geq M.$$

$f(\omega) = Y^T X \omega \quad \nabla f = X^T Y$

$\nabla f^T \nabla f$

$$\Rightarrow M \leq \left[\frac{\omega_0^T\omega_0 + \alpha\left(2Y^T X \omega^* + \alpha Y^T X X^T Y\right)}{\alpha^2 \gamma^2 \left(\sum_{i=1}^{m} y_i\right)^2}\right]$$

$$\Rightarrow M \leq \frac{\omega_0^T\omega_0 + \alpha\left(2Y^T X \omega^* + \alpha Y^T X X^T Y\right)}{\alpha^2 \gamma^2 (Y^T Y)}$$