Assignment 1 - Problem 1

→ **Statement of the Theorem** -

The weights $v_1, v_2, \ldots v_n$ of a perceptron are adjusted on a iteration by iteration basis. This final weight vector is obtained in a finite number of steps and was proved by Rosenblatt.

→ **Theory** -

We define the following -

Input vector → $w(n) = [1, w_1(n), w_2(n) \ldots, w_m(n)]^T$

Weight vector → $v(n) = [b, v_1(n), \ldots \ldots, v_m(n)]^T$

Linear combiner o/p $f(n) = v^T(n) \cdot w(n)$

$$\textcircled{1} \begin{cases} v^T w_i > 0 \rightarrow \text{for every i/p vector } w_i \text{ belonging to class } C_1 \\ v^T w_i \leq 0 \rightarrow \text{for every i/p vector } w_i \text{ belonging to class } C_2 \end{cases}$$

Below is the brief overview regarding various variables which would be used -

$w_1 \ldots w_N$ → Activity of associators → $(w_{i_k} * l_k)$

w_{i_k} ↓ I/p, l_k ↓ class.

y → This is the final weight vector

Eqn ② ⇒ $(w_i, y) > \theta > 0 \quad i = 1, \ldots N.$

θ → A threshold 'θ' has been considered for neuron activation.

$v_0, v_1, v_2, \ldots v_n$ → Weight vectors.

$$V_n = \begin{cases} V_{n-1} & \text{if } (\omega_{i_n}, V_{n-1}) > \theta \\ V_{n-1} + \omega_{i_n} & \text{if } (\omega_{i_n}, V_{n-1}) \leq \theta \end{cases} \rightarrow (3)$$

Eqn (3) is similar to eqn (1) with just an additional threshold factor considered.

→ The theorem can be restated as - "The goal of the perception is to accurately classify the signals into one of the two classes C_1 (o/p = +1) & to C_2 (o/p = -1) based on a set of weights applied to the network.

→ The adaption of weights could be done by following the error-correction rule, also known as Perceptron convergence algorithm [finite number of corrections].

→ A signal is said to be classified appropriately when $(\omega_i, V) > \theta$ (threshold)

→ A signal is not classified appropriately i.e., not assigned to the correct class when $(\omega_i, V) \leq \theta$.

→ Algorithm formulation/Weight updation -

$$(4) \begin{cases} V(n) = V(n-1) & \text{if } V^T \omega_i > 0 \text{ \& } \omega_i \text{ belongs to class 1} \\ V(n) = V(n-1) & \text{if } V^T \omega_i \leq 0 \text{ \& } \omega_i \text{ belongs to class 2} \end{cases}$$

Otherwise,

$$(5) \begin{cases} V(n) = V(n-1) + \eta\,\omega_i & \text{if } V^T \omega_i > 0 \text{ \& } \omega_i \text{ belongs to class 2} \\ V(n) = V(n-1) + \eta\,\omega_i & \text{if } V^T \omega_i \leq 0 \text{ \& } \omega_i \text{ belongs to class 1} \end{cases}$$

[η → learning parameter]

→ Proof -

Initial condition $V_0 = 0$, and we assume C_1 & C_2 are linearly seperable. Consider ω_{i_n} belongs to class 1 and it was wrongly classified i.e., $\omega_{i_n} \cdot V_{n-1} \leq \theta$ for $n = 1, ----$

then from eqn (5), we write as -

(6) ⇒ $V_n = V_{n-1} + \omega_{i_n}$ → ω_{i_n} belongs to class 1 & considering $\eta = 1$.

$n=1 \Rightarrow V_1 = V_0 + \omega_1 = \omega_1$ ($\because V_0 = 0$, considered)

$n=2 \Rightarrow V_2 = V_1 + \omega_2 = \omega_1 + \omega_2$

$n=3 \Rightarrow V_3 = V_2 + \omega_3 = \omega_1 + \omega_2 + \omega_3$.

(7) ⇒ hence $V_n = \omega_1 + \omega_2 + ---- + \omega_n$.

For a fixed solution v^*, we may define a +ve number α as -

$$\alpha = \min_{\omega_i \in C_1} v^{*T} \omega_i$$

multiply (7) with v^{*T}

$$v^{*T} \cdot V_n = v^{*T} \omega_1 + v^{*T} \omega_2 + ----- + v^{*T} \omega_n.$$

hence $v^{*T} V_n \geq n\alpha$

(8) ⇒ $[v^{*T} V_n]^2 \geq n^2 \alpha^2$

Using Cauchy- Schwartz inequality.

$$\| v^* \|^2 \| V_n \|^2 \geq n^2 \alpha^2$$

$$\Rightarrow \| V_n \|^2 \geq \frac{n^2 \alpha^2}{\| v^* \|^2} \rightarrow (9)$$

in eqn (9) $\because$ α & $\|v^*\|$ are constants, so let $c = \frac{\alpha^2}{\|v^*\|^2}$

then eqn (9) would change as -

$$\boxed{\|v_n\|^2 \geq c n^2} \rightarrow (10)$$

We next follow another development -

$v_k = v_{k-1} + w_{i_k}$ for $k = 1, \ldots n$ & $w_{i_k} \in C_1$

taking squared euclidean norm

$$\|v_k\|^2 = \|v_{k-1}\|^2 + \|w_{i_k}\|^2 + 2\, v_{k-1}^T \cdot w_{i_k} \rightarrow (11)$$

we know, $v_{k-1}^T \cdot w_{i_k} \leq 0$,

$$\& \; M = \max_{i=1,\ldots N} \|w_i\|^2$$

then eqn (11) $\Rightarrow \|v_k\|^2 - \|v_{k-1}\|^2 = 2(v_{k-1}, w_{i_k}) + \|w_{i_k}\|^2 \leq 2\theta + M$

$\Rightarrow$ ~~$\|v_k\|^2 \leq$~~ adding the above inequalities for

$k = 1, 2, \ldots n$, then -

$$\boxed{\|v_k\|^2 \leq \|v_0\|^2 + (2\theta + M)\, n} \rightarrow (12)$$

with our initial assumption of $v_0 = 0$

then (12) $\Rightarrow \|v_k\|^2 \leq nM$. $\rightarrow$ (13)

Eqn (13) is clearly conflicting with eqn (10) for large values of n. So we can state that n cannot be larger than some n_{max} i.e., $\frac{n_{max}^2 \alpha^2}{\|v^*\|^2} = n_{max} M \Rightarrow \boxed{n_{max} = \frac{M \|v^*\|^2}{\alpha^2}}$

$\rightarrow$ Hence, for $v_0 = 0$, $\eta = 1$, a solution vector v^* exists and the perceptron must terminate after atmost n_{max} iterations [finite].