Perceptron Convergence Theorem

Group Members
- Ahwin Oghenenikevwe
- Uche Ekekwa
- Akshay Pala

<u>Theorem</u>

Let $w_1, \cdots w_N$ be a set of vectors in a Euclidean space of fixed finite dimension, satisfying the hypothesis that there exists a vector y such that

$$(w_i, y) \geq \theta > 0 \quad i = 1, \cdots, N \qquad \text{—(1)}$$

Consider the infinite sequence $w_{i_1}, w_{i_2}, w_{i_3}, \cdots,\ 1 \leq i_K \leq N$ for every K, such that each vector $w_1, \cdots w_N$ occurs infinitely often. Recursively construct a sequence of vectors $v_0, v_1, \cdots v_n$ as follows:

$$v_0 \text{ is arbitrary}$$

$$v_n = \begin{cases} v_{n-1} & \text{if } (w_{i_n}, v_{n-1}) > \theta \\ v_{n-1} + w_{i_n} & \text{if } (w_{i_n}, v_{n-1}) \leq \theta \end{cases} \qquad \text{—(2)}$$

The sequence (v_n) is convergent. For some index m, $v_m = v_{m+1} = v_{m+2} = \cdots = \tilde{v}$.

w_i represents the activity of the associators.

y represents a satisfactory assignment of associator weights.

w_{i_n} represents the training sequence.

v_n represents the error-correction procedure.

θ represents threshold to be exceeded for perceptron to be correct.

From (2) as n varies, the sequence v_n changes by the addition of one or another of the set $w_1, \dots, w_N$. Hence convergence implies convergence in a finite number of steps. The term w_{i_n} is inessential hence the new training sequence is such that correction takes place at every step.

$$\therefore \quad v_n = v_{n-1} + w_{i_n} \text{ and } (w_{i_n}, v_{n-1}) \leq \theta \text{ for each } n. \qquad -(3)$$

$$v_n = v_0 + w_{i_1} + \dots + w_{i_n} \quad \text{from (3)} \qquad -(4)$$

Premultiplying (4) by v_n

$$\therefore \|v_n\|^2 = v_n (v_0 + w_{i_1} + \dots) \qquad -(5)$$

If n is sufficiently large and we select a minimum positive constant, c, that is less than each term in (5)

we then have:
$$\|v_n\|^2 > cn^2 \qquad -(6)$$

From (1) above, $(w_i, y) > \theta > 0$

we have: $(v_n, y) > (v_0, y) + n\theta > 0$

From (6), we can see that (v_n, y) from (5) can be replaced with cn^2.

Hence using Cauchy - Schwartz inequality,

$$\|v_n\|^2 \geq \frac{(v_n \cdot y)^2}{\|y\|^2} > \frac{[(v_0, y) + n\theta]^2}{\|y\|^2}$$

$$\frac{[(v_0, y) + n\theta]^2}{\|y\|^2} = \frac{(v_0, y)^2 + n^2\theta^2 + 2(v_0, y)n\theta}{\|y\|^2}$$

can be factorized to $\dfrac{\theta^2}{\|y\|^2}\left[n + \dfrac{(v_0, y)}{\theta}\right]^2$

From (6)

$$\frac{\theta^2}{\|y\|^2}\left[n + \frac{(v_0, y)}{\theta}\right]^2 = Cn^2$$

$\therefore$ If $(w_0, y) \geq 0$ then C, a constant can be represented with $\dfrac{\theta^2}{\|y\|^2}$ ...

From (3), we can deduce that :

$$\|v_n\|^2 \leq \|v_0\|^2 + (2\theta + M)n$$

where $M = \max \|w_i\|^2$ ...

the integer-argument function from ③ replacing
n iterations with K yields:

$$\|v_K\|^2 - \|v_{K-1}\|^2 = 2(v_{K-1}, w_{i_K}) + \|w_{i_K}\|^2 \leq 2\theta + m$$

$$\|v_n\|^2 > Cn^2 \quad \text{and} \quad \|v_n\|^2 \leq \|v_0\|^2 + (2\theta + M)n$$

if n is sufficiently large both equations cannot hold true hence we can say that n or K used interchangeably is capped.