Exercise

Prove that, if the Perceptron algorithm is initialised with the null vector, then the coefficient η does not affect learning.

Solution

Given a training set S = {(x, t)}, $x \in \mathbb{R}^{n+1}$, $t \in \{-1, +1\}$, $\eta \geq 0$ (learning rate)

First, we initialise the value of the weights as a null vector.

Then repeat the following:

(a) Select (randomly) one of the training examples (x,t)

(b) If $o = \text{sign}(w \cdot x) \neq t$, then $\quad w \leftarrow w + \eta(t - o)x$

In the first iteration, we have w_0 as a null vector, so $o_1 = \text{sign}(0) = 0 \neq \overset{\in\{-1,1\}}{t_0}$, then

$$w_1 \leftarrow w_0 + \eta(t_1 - o_1)x_1 = \eta(t_1 - 0)x_1 = \eta \cdot t_1 \cdot x_1$$

In the subsequence iterations, the weights are not updated if $o = \text{sign}(w \cdot x) = t$. So, for the second iteration where $o = \text{sign}(w \cdot x) \neq t$, we have

$$o_j = \text{sgn}(w_1 \cdot x_j) \neq t_j$$
$$w_j \leftarrow w_1 + \eta(t_j - o_j)x_j = \eta[t_1 x_1 + x_j(t_j - o_j)]$$

where j > 1. Similarly, for the next iteration where $o = \text{sign}(w \cdot x) \neq t$ again, we have k > j, and

$$o_k = \text{sgn}(w_j \cdot x_k) \neq t_k$$
$$w_k \leftarrow w_j + \eta(t_k - o_k) \cdot x_k = \eta[t_1 x_1 + x_j(t_j - o_j) + x_k(t_k - o_k)]$$

Hence, we can expand this to the general form

$$w_n = \eta\left[t_1 x_1 + \sum_{i=2}^{n} x_i(t_i - o_i)\right]$$

where $o_i = \text{sgn}(w_{i-1} \cdot x_i) \neq t_i \quad \forall i \in \{1, 2, \dots, n\}$

Note that the coefficient η does not exist within the square brackets, so it only scales the learning but not the dire it does not affect learning.