6 CLASIFICADOR DE PATRONES – METODO DETERMINISTA. REDES NEURONALES – PERCEPTRON

6.CLASIFICADOR DE PATRONES – METODO DETERMINISTA.
REDES NEURONALES – PERCEPTRON.

6.1. Introducción.

En este capítulo se discutirán algunos de los intentos por modelar la actividad del cerebro biológico usando redes neuronales artificiales.

La razón por la que el cerebro humano puede funcionar tan eficientemente es que usa computación paralela. Miles o aún más millones de células nerviosas llamadas *neuronas* se organizan para trabajar simultáneamente sobre el mismo problema.

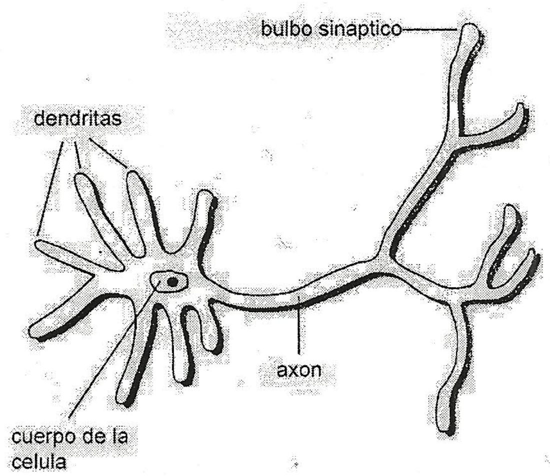

Figura 6. 1. Diagrama de una neurona biológica [16]

1. cuerpo de la célula
2. dendritas (reciben las señales de entrada)
   reciben los datos de los órganos sensoriales tales como ojos, oídos y de los axones de las neuronas
3. axon (envían las señales de salida)
   envían salidas a órganos tales como músculos y a dendritas de otras neuronas.

Gracias a la gran cantidad de conexiones entre neuronas y a las conexiones redundantes entre ellas el funcionamiento del cerebro es relativamente robusto y no se afecta significativamente a pesar de las muchas neuronas que mueren cada día y también a pesar de los traumas.

El cerebro es capaz de rehabilitarse gracias a las neuronas restantes.

Las neuronas en muchas regiones del cerebro biológico se organizan en *capas.*
Una neurona en una capa generalmente recibe sus entradas y sus salidas de otras neuronas en una capa adyacente.

En un intento para formar el modelo matemático de una neurona, contamos con los trabajos de McCulloch & Pitts (1943).

Básicamente su modelo está formado por las siguientes componentes:

- recibe M entradas X_1 , X_2 , ..., X_M
- calcula una suma de ponderaciones

$$S = \sum_{i=1}^{M} \omega_i \; x_i$$

Usando las ponderaciones $\omega_1, \omega_2, ..., \omega_M$

- existen dos umbrales :

$$\begin{cases} 0 & si \quad S < T \\ 1 & si \quad S > T \end{cases}$$

en donde T es el umbral

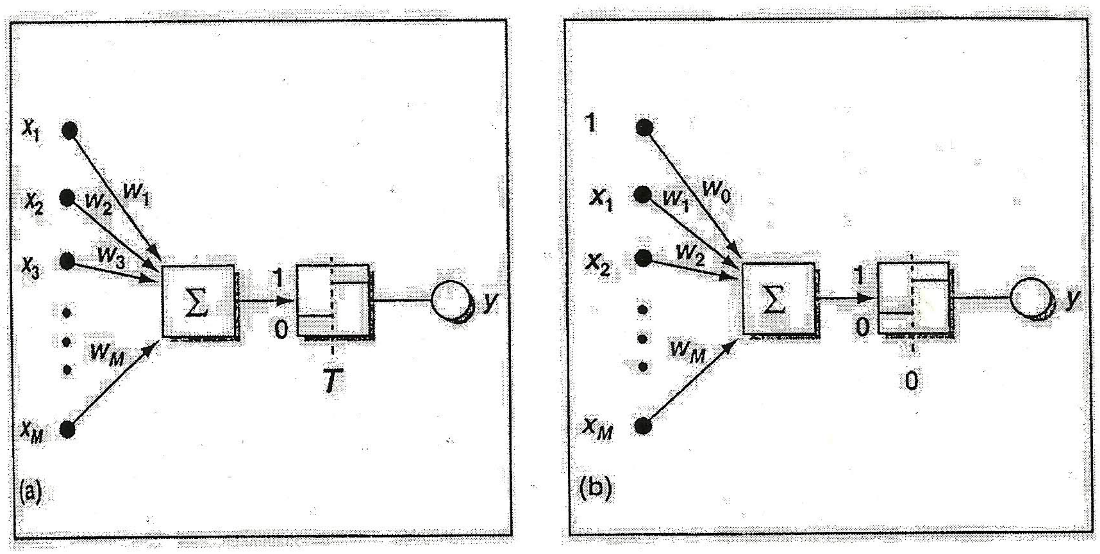

Figura 6.2: (a) modelo de la neurona de Mc-Culloch-Pitts. (b) modelo de una neurona con una ponderación de sesgo

La combinación ($(\sum, T)$ se llama *nodo* en una red neuronal.

La acción del modelo neurona es *output* 1 si:

$$\begin{cases} 1 \quad si \quad \sum_{i=1}^{M} \omega_i \; x_i > T \\ 0 \quad de\ otra\ manera \end{cases}$$

Por conveniencia, para describir algoritmos de aprendizaje se re-escribe la suma

$$D = \sum_{i=1}^{M} \omega_i \; x_i$$

En donde $\omega_0 \;=\; -T$ y $x_0 \;=\; 1$

$$\text{Salida} = \begin{cases} 0 \quad si \quad D > 0 \\ 1 \quad si \quad D \leq 0 \end{cases}$$

La ponderación $\bar{\omega}_0$ se llama *ponderación sesgo.*

Para producir redes neuronales electrónicas es necesario programar el hardware y conectarlo con computación paralela como lo hacen las neuronas para realizar cálculos simultáneamente.

Una dificultad de usar redes neuronales es el *aprendizaje* , esto es, lograr encontrar $\omega_i\,'s$ aceptables.

Una vez obtenidas las ponderaciones de las $\omega_i\,'s$ el problema de clasificar muestras es sencillo.

Construir una red neuronal es difícil y caro, entonces se recurre a la simulación sobre un solo procesador.

6.2. REDES SIN CAPAS OCULTAS: PERCEPTRON.

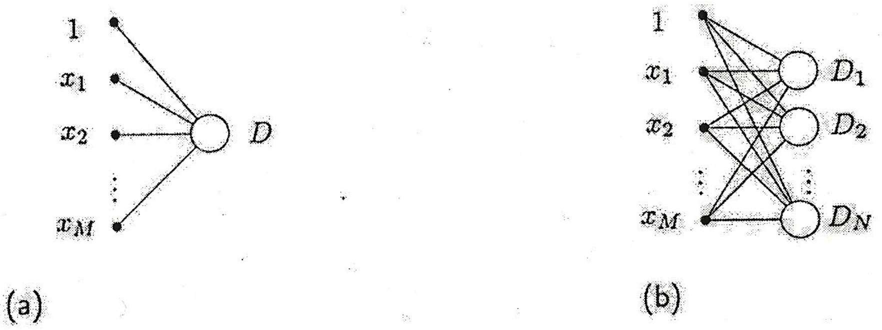

Figura 6.4 (a) Red neuronal 2 capas con un nodo de salida. La ponderación conecta la entrada x_i con la salida D y es w_i. b) Rec neuronal 2 capas con múltiples nodos de salida. La ponderación conecta entradas x_i con salida d_j, es w_j

la red neuronal de la figura 6.3 se llama de dos capas ya que tienen una capa de nodos de entradas directamente condectados con nodos de dalida. Pueden existir capas de nodos entre entradas y salidas estas capas se llaman capas ocultas. En la figura 6.3 se calcula $\sum_{i=1}^{M} \omega_i \, x_i$ y se compara con un umbral T. Rosenblat [32] usó la neurona de la figura 6.2 b para construir un clasificador de aprendizaje que llamó *perceptrón.*

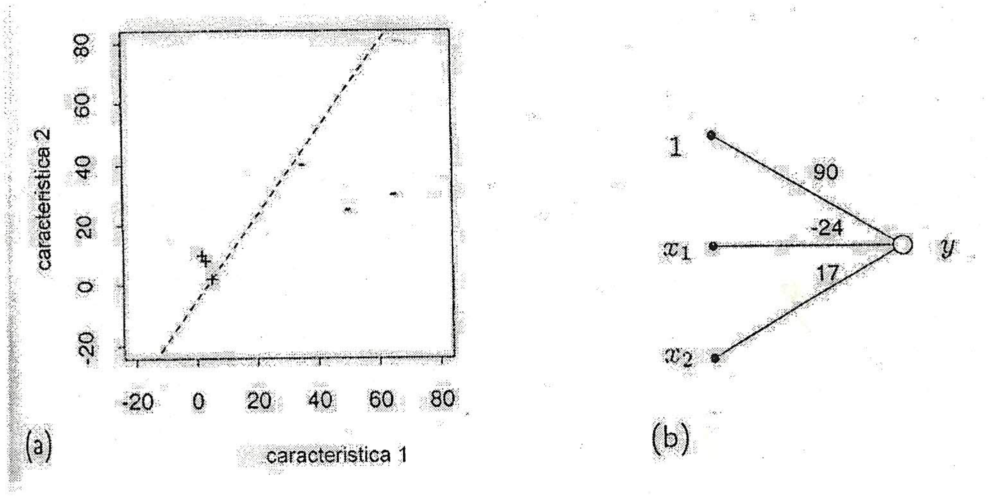

Figura 6.5: (a) La frontera de decisión final calculada por la técnica de perceptrón. (b) Red neuronal con sus ponderaciones finales [12]

aplicando el algoritmo de perteptrón se encontró la siguiente recta que se muestra en la figura 6.4

$$D = \omega_0 + \omega_1 \, x_1 + \omega_2 \, x_2 = 90 - 24 \, x_1 + 17 \, x_2 \qquad (6\text{-}1)$$

Y el clasificador trabaja así: por ejemplo, si queremos clasificar el punto:

a) **x** = (2,3) entonces evaluando la ecuación (1) $D=93>0$ entonces (2,3) clasifica en la clase w_1

b) **x** = (20,0) dá como resultado para $D<0$ entonces clasificamos (20,0) en la clase w_2.

Concepto de recompensa-castigo: método de perceptrón.

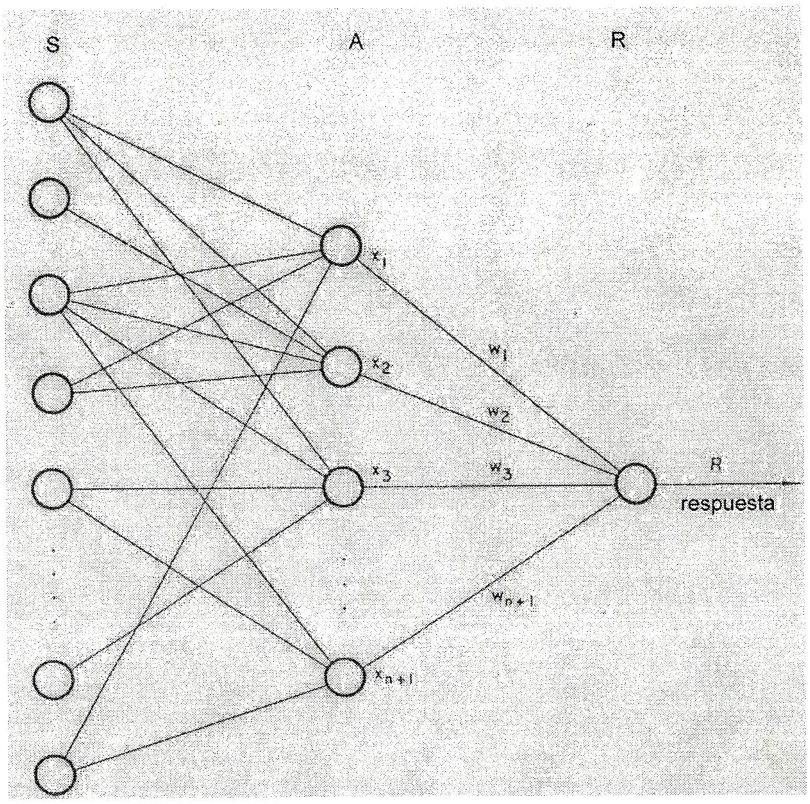

Figura 6.5 Modelo básico de perceptrón [36]

El método de aprendizaje de una máquina de perceptrón de la figura 6.5 es un esquema simple para la determinación iterativa del vector de ponderaciones **w.** Este esquema de llama *algoritmo de perceptrón* y puede describirse como sigue:

algoritmo

Dados dos conjuntos de valores perteneciente a dos clases w_1 y w_2, respectivamente, **w** (1) el vector de ponderaciones inicial, el cual se escoge arbitrariamente.

Si estamos en el paso $k \Rightarrow$

- si $\mathbf{x}(k) \in \omega_1$ y $\mathbf{w}'\mathbf{x}(k) \leq 0$ reemplazamos $\mathbf{w}(k)$ por :
  $\mathbf{w}(k+1) = \mathbf{w}(k) + c\,\mathbf{x}(k)$
  en donde c es una corrección de incremento.

- si $\mathbf{x}(k) \in \omega_2$ y $\mathbf{w}'(k)\,\mathbf{x}(k) \geq 0$ reemplazamos $\mathbf{w}(k)$ por :
  $\mathbf{w}(k+1) = \mathbf{w}(k) - c\,\mathbf{x}(k)$
  en donde c es una corrección de incremento

- de otra manera no cambia quiere decir que clasificó bien, si $\mathbf{w}(k+1) = \mathbf{w}(k)$

Este es un procedimiento recompensa-castigo. La convergencia ocurre cuando el vector **w** no cambia

Ejemplo 6.1.

Consideremos los patrones de la figura 6.4 (a). Se desea aplicar el algoritmo de perceptrón a estos patrones y tratar de encontrar una solución para el vector de ponderaciones. Por inspección vemos que estos patrones son separables linealmente, y que el algoritmo puede tener éxito.

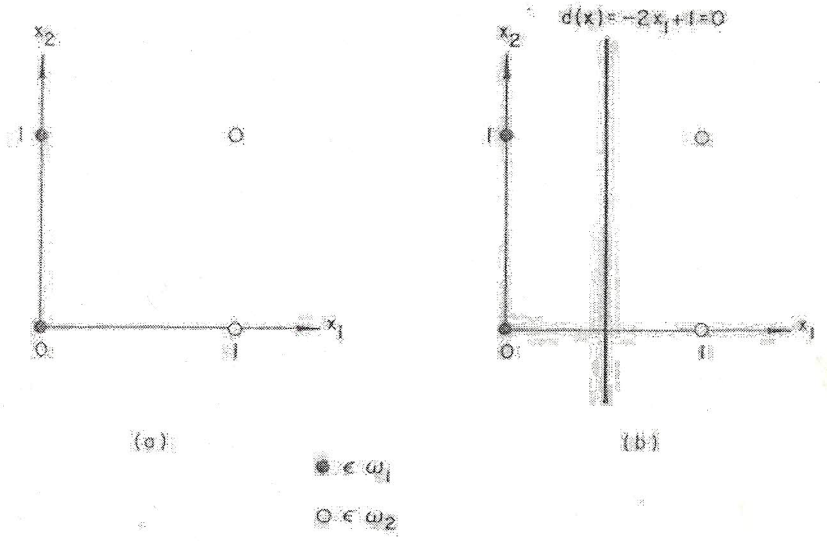

Figura 6.4 Ilustración de algoritmo de perceptrón. A) patrones pertenecientes a dos clases b) frontera de decisión determinada por aprendizaje [36]

Antes de aplicar el algoritmo todos los patrones deben ser aumentados. Las clases son las siguientes:

$$\omega_1 = \{ (0, 0, 1)', (0, 0, 1)' \}$$

$$\omega_2 = \{ (1, 0, 1)', (1, 1, 1)' \}$$

Haciendo c=1, la secuencia de pasos para encontrar las ponderaciones de la recta que separa estas clases es la siguiente:

Iteración 1

$$\mathbf{w}(1) = (0, 0, 0)$$
$$\mathbf{x}(1) = (0, 0, 1)$$
$$\mathbf{x}(1) = \in w_1 \quad \text{y}$$

$t = 1 \qquad \mathbf{x}_1 = (0, 0, 1) \qquad\qquad \mathrm{w}(2) = \mathrm{w}(1) \pm 1 \cdot \mathrm{x}(1)$

$$\mathbf{w}'(1)\,\mathbf{x}(1) = (0,0,0)\begin{pmatrix}0\\0\\1\end{pmatrix} = 0, \quad \Rightarrow \qquad \mathbf{w}(2) = \mathbf{w}(1) + \mathbf{x}(1) = \begin{pmatrix}0\\0\\1\end{pmatrix}$$

$t = 2 \qquad \mathbf{x}_2 = (0, 0, 1)' \in w_1$

$$\mathbf{w}'(2)\,\mathbf{x}(2) = (0,0,1)\begin{pmatrix}0\\0\\1\end{pmatrix} = 1, \quad > 0 \Rightarrow \qquad \mathbf{w}(3) = \mathbf{w}(2) = \begin{pmatrix}0\\0\\1\end{pmatrix}$$

$t = 3 \qquad \mathbf{x}_3 = (1, 0, 1)' \in w_2$

$$\mathbf{w}'(3)\,\mathbf{x}(3) = (0,0,1)\begin{pmatrix}1\\0\\1\end{pmatrix} = 1, \quad > 0 \Rightarrow \qquad \mathbf{w}(4) = \mathbf{w}(3) + \mathbf{x}(3) = \begin{pmatrix}-1\\0\\0\end{pmatrix}$$

$t = 4 \qquad \mathbf{x}_4 = (1, 1, 1)' \in w_2$

$$\mathbf{w}'(4)\,\mathbf{x}(4) = (-1,0,0)\begin{pmatrix}1\\1\\1\end{pmatrix} = -1, \quad < 0 \Rightarrow \qquad \mathbf{w}(5) = \mathbf{w}(4) = \begin{pmatrix}-1\\0\\0\end{pmatrix}$$

hubo correcciones en paso 1 y paso 3

Interación 2

La máquina continúa: repite la pasada de los puntos

$t = 5 \qquad \mathbf{x}_5 = (0, 0, 1)' \in w_1$ y

$$\mathbf{w}'(5)\mathbf{x}(5) = 0, \quad \Rightarrow \quad \mathbf{w}(6) = \mathbf{w}(5) + \mathbf{x}(5) = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$t = 6 \qquad \mathbf{x}_6 = (0, 1, 1)' \in w_1$ y

$$\mathbf{w}'(6)\mathbf{x}(6) = 1, \quad > 0 \Rightarrow \quad \mathbf{w}(7) = \mathbf{w}(6) = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$t = 7 \qquad \mathbf{x}_7 = (1, 0, 1)' \in w_2$ y

$$\mathbf{w}'(7)\mathbf{x}(7) = 0, \quad \Rightarrow \quad \mathbf{w}(8) = \mathbf{w}(7) - \mathbf{x}(7) = \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}$$

$t = 8 \qquad \mathbf{x}_8 = (1, 1, 1)' \in w_2$ y

$$\mathbf{w}'(8)\mathbf{x}(8) = -2 \quad < 0 \Rightarrow \quad \mathbf{w}(9) = \mathbf{w}(8) = \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}$$

Interación 3

Ya que dos errores ocurrieron en esta iteración, los patrones vuelven a pasar:

$t = 9 \qquad \mathbf{x}_9 = (0, 0, 1)' \in w_2$ y

$$\mathbf{w}'(9)\,\mathbf{x}(9) = 0, \qquad \Rightarrow \qquad \mathbf{w}(10) = \mathbf{w}(9) + \mathbf{x}(9) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

$t = 10 \qquad \mathbf{x}_{10} = (0, 1, 1)' \in w_1$ y

$$\mathbf{w}'(10)\,\mathbf{x}(10) = 1, \qquad > 0 \Rightarrow \qquad \mathbf{w}(11) = \mathbf{w}(10) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

$t = 11 \qquad \mathbf{x}_{11} = (1, 0, 1)' \in w_2$ y

$$\mathbf{w}'(11)\,\mathbf{x}(11) = -1, \qquad < 0 \Rightarrow \qquad \mathbf{w}(12) = \mathbf{w}(11) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

$t = 12 \qquad \mathbf{x}_{12} = (1, 1, 1)' \in w_2$ y

$$\mathbf{w}'(12)\,\mathbf{x}(12) = -1, \qquad < 0 \Rightarrow \qquad \mathbf{w}(13) = \mathbf{w}(12) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

es fácil ver que en la siguiente interacción todos los patrones quedarán clasificados correctamente

la solución es $\mathbf{w} = (-2, 0, 1)'$

$$\mathbf{x} = (x_1, x_2) \quad y \quad d(\mathbf{x}) = \mathbf{w}'\mathbf{x} = -2x_1 + 1$$

Se puede probar que el algoritmo converge.

Ejemplo 6.2[1]

Consideremos el problema de clasificar dos clases representadas por los puntos

$$w_1 = \{(0, 1), (0, 2), (1, 2)\}$$

$$w_2 = \{(1, 0), (2, 0), (2, 1)\}$$

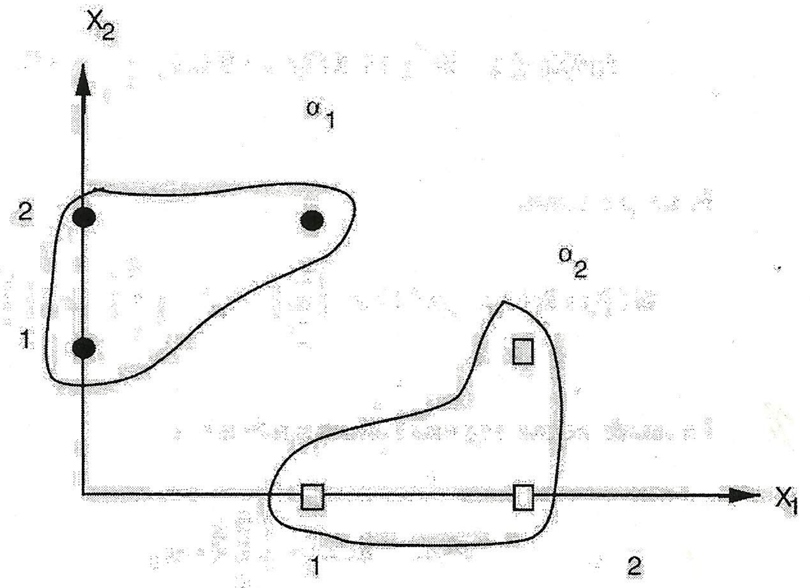

Figura 6.5 [21]

Los patrones nuestrales aumentados son:

$$\left\{ \begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}0\\2\\1\end{pmatrix}, \begin{pmatrix}1\\2\\1\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}2\\0\\1\end{pmatrix}, \begin{pmatrix}2\\1\\1\end{pmatrix} \right\}$$

[1] Dario Maravall Gómez-Allende

$$t = 1: \qquad \mathbf{x}(1) = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \in \omega_1 \text{ y}$$

$$d[\mathbf{x}(1)] = \mathbf{w}'(1)\mathbf{x}(1) = (0,0,0)\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 0 \quad \Rightarrow \quad \text{error corregir } \omega(2)$$

$$\Rightarrow \mathbf{w}(2) = \mathbf{w}(1) + c\mathbf{x}(1) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

$$t = 2: \qquad \mathbf{x}(2) = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \in \omega_1 \text{ y}$$

$$d[\mathbf{x}(2)] = \mathbf{w}'(2)\mathbf{x}(2) = (0,1,1)\begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} = 3 > 0 \Rightarrow \quad w(3) = w(2) = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

$$t = 3: \qquad \mathbf{x}(3) = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \in \omega_1 \text{ y}$$

$$d[\mathbf{x}(3)] = \mathbf{w}'(3).\mathbf{x}(3) = 3 \qquad > 0 \Rightarrow \qquad \mathbf{w}(4) = \mathbf{w}(3) = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

$$t = 4: \qquad t = 4: \mathbf{x}(4) = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \in \omega_2 \text{ y}$$

$$d[\mathbf{x}(4)] = \mathbf{w}^{1}(4)\mathbf{x}(4) = (0,1,1)\begin{pmatrix}1\\0\\1\end{pmatrix} = 1 \qquad >0 \Rightarrow$$

$$\mathbf{w}(5) = \mathbf{w}(\mathbf{x}) + c\,\mathbf{x}(4) = \begin{pmatrix}0\\1\\1\end{pmatrix} - \begin{pmatrix}1\\0\\1\end{pmatrix} = \begin{pmatrix}-1\\1\\0\end{pmatrix}$$

$t = 5 \qquad \mathbf{x}(5) = \begin{pmatrix}2\\0\\1\end{pmatrix} \in \omega_2$ y

$$d[\mathbf{x}(5)] = [-1 \;\; 1 \;\; 0]\begin{pmatrix}2\\0\\1\end{pmatrix} = -2 \quad >0 \Rightarrow \quad \mathbf{w}(6) = \mathbf{w}(5) = \begin{pmatrix}-1\\1\\0\end{pmatrix}$$

$t = 5 \qquad \mathbf{x}(5) = \begin{pmatrix}2\\0\\1\end{pmatrix} \in \omega_2$ y

$$d[\mathbf{x}(6)] = (-1,1,0)\begin{pmatrix}2\\1\\1\end{pmatrix} = -1 \quad >0 \Rightarrow \quad \mathbf{w}(7) = \mathbf{w}(6) = \begin{pmatrix}-1\\1\\0\end{pmatrix}$$

Condición de paro

Los coeficientes de w son

$\mathbf{w} = (w_1, w_2, w_0) = (-1, 1, 0)$

$$d(\mathbf{x}) = -x_1 + x_2$$