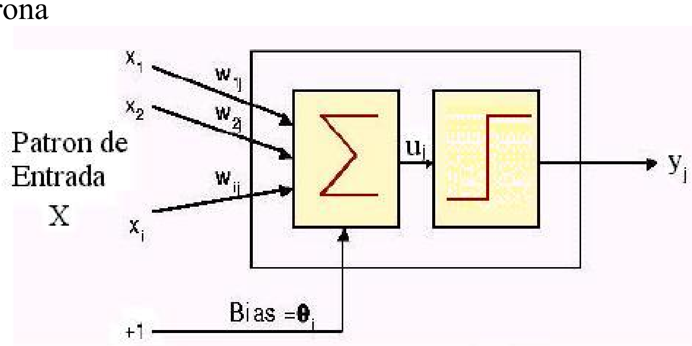
1. El Perceptrón

El perceptrón fue desarrollado por Frank Rosenblatt (Rosenblatt, 1962), es uno de los sistemas neuronales más simples, los cuales son usados como eficientes procesadores de información. Dado que algunas de sus características, tales como su comportamiento geométrico y estadístico, la filosofía de sus algoritmos de aprendizaje, el balance que se debe efectuar en su diseño entre el poder de aproximación y el de estimación, son extensibles a otros sistemas neuronales más complicados.

El perceptrón es usado para resolver problemas linealmente separables, utiliza el modelo de McCullot-Pitts como neurona.

Características Básicas

Modelo de la neurona



Regla de propagación: $u_i(t) = \sum_{j=1}^{N} x_j w_j + \theta_j$

Función de activación:

$$y_i(t+1) = f(u_i(t))$$

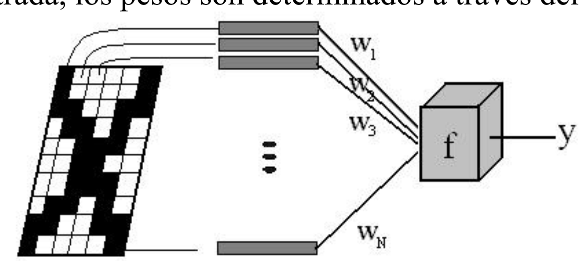
$$y_i(t+1) = \begin{cases} -1, \; si \; u_i(t) < 0 \\ +1, \; si \; u_i(t) > 0 \end{cases}$$

Estados de activación

- Bipolares (1 = activo, -1 = inactivo)
- Binarios (1 = activo, 0 = inactivo)

Arquitectura

- **Capa de pre-procesamiento** que es utilizada para la extracción de características, pudiendo implementar cualquier función pero los pesos son fijos.
- **Capa de discriminación** las neuronas de salida son usadas para discriminar los patrones de entrada, los pesos son determinados a través del aprendizaje

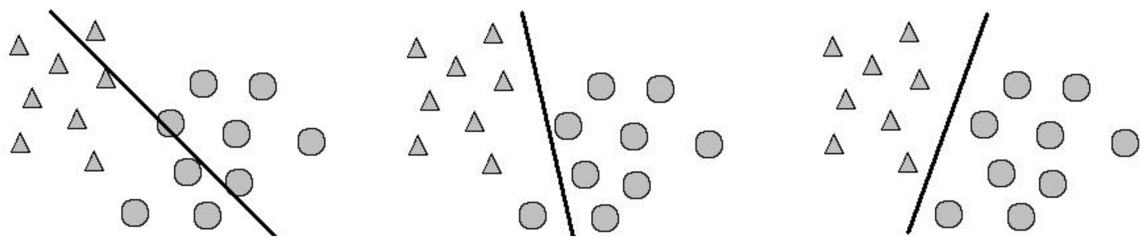


Entrenamiento

- El paradigma de entrenamiento utilizado es el supervisado utilizando el algoritmo de corrección de error el cual esta dado por:

$$\Delta w_{ij} = \begin{cases} \eta\, x_i (d_j - y_j) & si\ d \neq y \\ 0 & d = y \end{cases}$$

Finalidad del entrenamiento



Algoritmo de entrenamiento

1. Iniciar todas las conexiones en $w_{ij} = 0$
2. Repita

    Para par de entrenamiento (*X*, *d*)

        Calcular la salida *y*

        Si (*d* ≠ *y*) entonces

            Actualizar los pesos

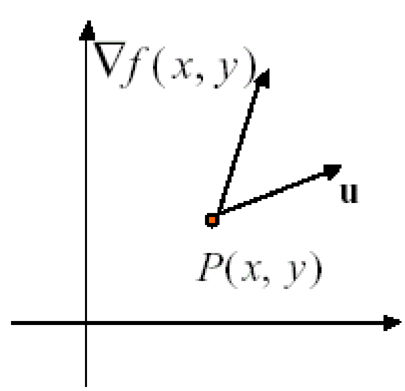
    Hasta que el error sea aceptable

NOTA: No ocurre variación en el peso de las neuronas si la salida obtenida es correcta, caso contrario cada peso es actualizado según la tasa de aprendizaje.

Gradiente de una función: $\nabla f(x,y) = \left(\frac{\partial}{\partial x} f(x,y), \frac{\partial}{\partial y} f(x,y) \right)$

Derivada direccional ($D_{\mathrm{u}} f(x,y)$) es la tasa de variación de $f(x,y)$ en la dirección definida por **u**.

$$\begin{aligned} D_{\mathrm{u}} f(x,y) &= \nabla f(x,y) \cdot \mathrm{u} \\ &= \|\nabla f(x,y)\| \|\mathrm{u}\| \cos\gamma \\ &= \|\nabla f(x,y)\| \cos\gamma \end{aligned}$$



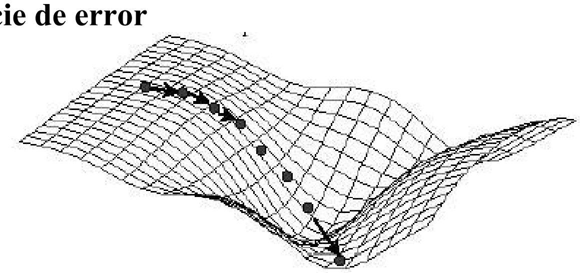
Teorema del gradiente: Sea f una función de dos variables diferenciable en el punto $P(x, y)$.

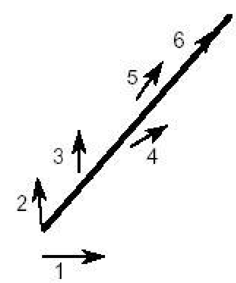
- El máximo de $D_u f(x, y)$ en $P(x, y)$ es $\|\nabla f(x, y)\|$.
- La máxima tasa de crecimiento de $f(x, y)$ en $P(x, y)$ ocurre en la dirección del gradiente $\nabla f(x, y)$

***Corolario*:** Sea f una función de dos variables diferenciable en el punto $P(x, y)$.

- El mínimo de $D_u f(x, y)$ en $P(x, y)$ es $-\|\nabla f(x, y)\|$.
- La máxima tasa de decrecimiento de $f(x, y)$ en $P(x, y)$ ocurre en la dirección del gradiente $-\nabla f(x, y)$

Superficie de error





Método del gradiente descendiente:

Cada peso sináptico i de la neurona j es actualizado al negativo de la derivada parcial del error de esta neurona con relación al peso.

$$\Delta w_{ij} = -\eta \frac{\partial E_j}{\partial w_{ij}}$$

$$\Delta w_{ij} = -\eta \frac{\partial E_j}{\partial w_{ij}} = -\eta \frac{\partial E_j}{\partial y_j} \frac{\partial y_j}{\partial w_{ij}}$$

$$E_j = \frac{1}{2}\left(d_j - y_j\right)^2 \qquad y_j = \sum x_i w_{ij} + \theta_j$$

$$\Delta w_{ij} = -\eta \left(2.\frac{1}{2}(d_j - y_j).(-1)\right).\ x_i$$

$$\Delta w_{ij} = \eta \left(d_j - y_j\right).\ x_i$$

Algoritmo de Prueba

1. Presentar el patrón X a ser reconocido
2. Calcular la salida y
3. Si (d = -1) entonces

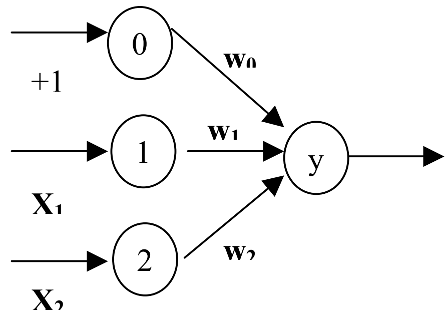
   X $\in$ clase 0

   Caso contrario

   X $\in$ clase 1

Ejemplo

AND	X_0	X_1	X_2	D
Entrada1	**1**	**0**	**0**	**0**
Entrada2	**1**	**0**	**1**	**0**
Entrada3	**1**	**1**	**0**	**0**
Entrada4	**1**	**1**	**1**	**1**



Peso inicial: $w_0 = 0$, $w_1 = 0$, $w_2 = 0$

Tasa de aprendizaje: $\eta = 0.5$

1)

Entrada 1: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(0\times1+0\times0+0\times0) = f(0) = 0$ $y_{out} = d$

Entrada 2: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(0\times1+0\times1+0\times0) = f(0) = 0$ $y_{out} = d$

Entrada 3: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(0\times1+0\times0+0\times1) = f(0) = 0$ $y_{out} = d$

Entrada 4: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(0\times1+0\times1+0\times1) = f(0) = 0$ $y_{out} \neq d$

$w_0 = w_0+(d-y_{out})x_0 = 0 + 0.5 \times(1-0) \times1 = 0.5$

$w_1 = w_1+(d-y_{out})x_1 = 0 + 0.5 \times(1-0) \times1 = 0.5$

$w_2 = w_2+(d-y_{out})x_2 = 0 + 0.5 \times(1-0) \times1 = 0.5$

2)

Entrada 1: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(0.5\times1+0.5\times0+0.5\times0) = f(0.5) = 1$ $y_{out} \neq d$

$w_0 = w_0+(d-y_{out})x_0 = 0.5 + 0.5 \times(0-1) \times1 = 0$

$w_1 = w_1+(d-y_{out})x_1 = 0.5 + 0.5 \times(0-1) \times0 = 0.5$

$w_2 = w_2+(d-y_{out})x_2 = 0.5 + 0.5 \times(0-1) \times0 = 0.5$

Entrada 2: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(0\times1+0.5\times0+0.5\times1) = f(0.5) = 1$ $y_{out} \neq d$

$w_0 = w_0+(d-y_{out})x_0 = 0 + 0.5 \times(0-1) \times1 = -0.5$

$w_1 = w_1+(d-y_{out})x_1 = 0.5 + 0.5 \times(0-1) \times0 = 0.5$

$w_2 = w_2+(d-y_{out})x_2 = 0.5 + 0.5 \times(0-1) \times1 = 0$

Entrada 3: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(-0.5\times1+0.5\times1+0\times0) = f(0) = 0$ $y_{out} = d$

Entrada 4: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(-0.5\times1+0.5\times1+0\times1) = f(0) = 0$ $y_{out} \neq d$

$w_0 = w_0+(d-y_{out})x_0 = -0.5 + 0.5 \times(1-0) \times1 = 0$

$w_1 = w_1+(d-y_{out})x_1 = 0.5 + 0.5 \times(1-0) \times1 = 1$

$w_2 = w_2+(d-y_{out})x_2 = 0 + 0.5 \times(1-0) \times1 = 0.5$

3)

Entrada 1: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(0\times1+1\times0+0.5\times0) = f(0) = 0$ $y_{out} = d$

Entrada 2: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(0\times1+1\times0+0.5\times1) = f(0.5) = 1$ $y_{out} \neq d$

$w_0 = w_0+(d-y_{out})x_0 = -0.5 + 0.5 \times(0-1) \times1 = -1$

$w_1 = w_1+(d-y_{out})x_1 = 1 + 0.5 \times(0-1) \times0 = 1$

$w_2 = w_2+(d-y_{out})x_2 = 0.5 + 0.5 \times(0-1) \times1 = 0$

Entrada 3: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(-1\times1+1\times1+0\times0) = f(0) = 0$ $y_{out} = d$

Entrada 4: $y_{out} = f(w_0x_0 + w_1x_1 + w_2x_2) = f(-1\times1+1\times1+0\times1) = f(0) = 0$ $y_{out} \neq d$

$w_0 = w_0+(d-y_{out})x_0 = -1 + 0.5 \times(1-0) \times1 = -0.5$

$w_1 = w_1 + (d - y_{out}) x_1 = 1 + 0.5 \times (1-0) \times 1 = 1.5$
$w_2 = w_2 + (d - y_{out}) x_2 = 0 + 0.5 \times (1-0) \times 1 = 0.5$

4)
Entrada 1: $y_{out} = f(w_0 x_0 + w_1 x_1 + w_2 x_2) = f(-0.5 \times 1 + 1.5 \times 0 + 0.5 \times 0) = f(-0.5) = 0$ $y_{out} = d$
Entrada 2: $y_{out} = f(w_0 x_0 + w_1 x_1 + w_2 x_2) = f(-0.5 \times 1 + 1.5 \times 0 + 0.5 \times 1) = f(0) = 0$ $y_{out} = d$
Entrada 3: $y_{out} = f(w_0 x_0 + w_1 x_1 + w_2 x_2) = f(-0.5 \times 1 + 1.5 \times 1 + 0.5 \times 0) = f(1) = 1$ $y_{out} \neq d$
$w_0 = w_0 + (d - y_{out}) x_0 = -0.5 + 0.5 \times (0-1) \times 1 = -1$
$w_1 = w_1 + (d - y_{out}) x_1 = 1.5 + 0.5 \times (0-1) \times 1 = 1$
$w_2 = w_2 + (d - y_{out}) x_2 = 0.5 + 0.5 \times (0-1) \times 0 = 0.5$
Entrada 4: $y_{out} = f(w_0 x_0 + w_1 x_1 + w_2 x_2) = f(-1 \times 1 + 1 \times 1 + 0.5 \times 1) = f(0.5) = 1$ $y_{out} = d$

5)
Entrada 1: $y_{out} = f(w_0 x_0 + w_1 x_1 + w_2 x_2) = f(-1 \times 1 + 1 \times 0 + 0.5 \times 0) = f(-1) = 0$ $y_{out} = d$
Entrada 2: $y_{out} = f(w_0 x_0 + w_1 x_1 + w_2 x_2) = f(-1 \times 1 + 1 \times 0 + 0.5 \times 1) = f(-0.5) = 0$ $y_{out} = d$
Entrada 3: $y_{out} = f(w_0 x_0 + w_1 x_1 + w_2 x_2) = f(-1 \times 1 + 1 \times 1 + 0.5 \times 0) = f(0) = 0$ $y_{out} = d$
Entrada 4: $y_{out} = f(w_0 x_0 + w_1 x_1 + w_2 x_2) = f(-1 \times 1 + 1 \times 1 + 0.5 \times 1) = f(0.5) = 1$ $y_{out} = d$

$$w_0 = -1,\ w_1 = 1,\ w_2 = 0.5$$

Interpretación geométrica