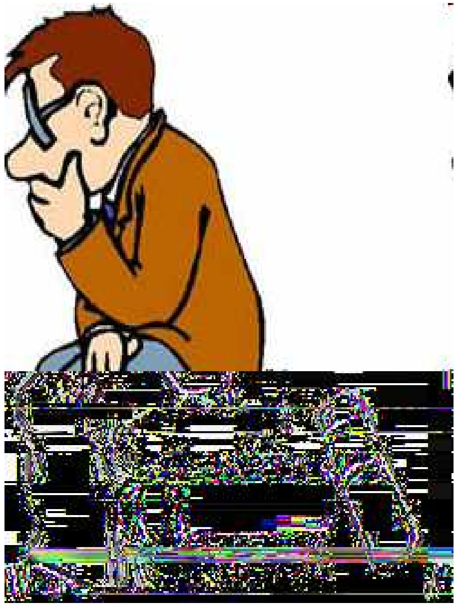
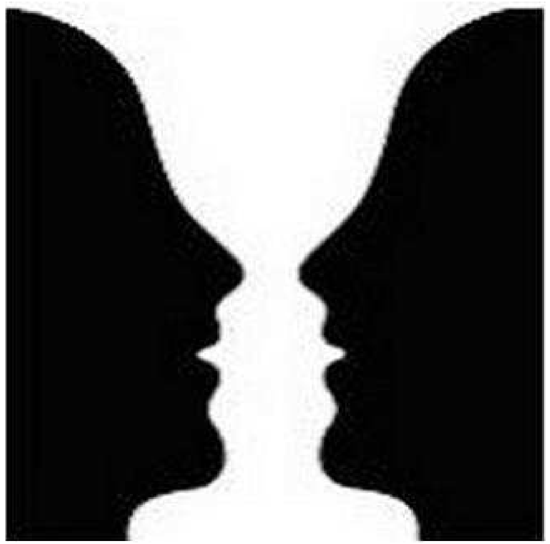
Redes Neuronales

Jorge Luis Guevara Díaz

- ¿Es posible que las máquinas puedan asociar y aprender de la experiencia?



Como Asociamos?



Redes Neuronales

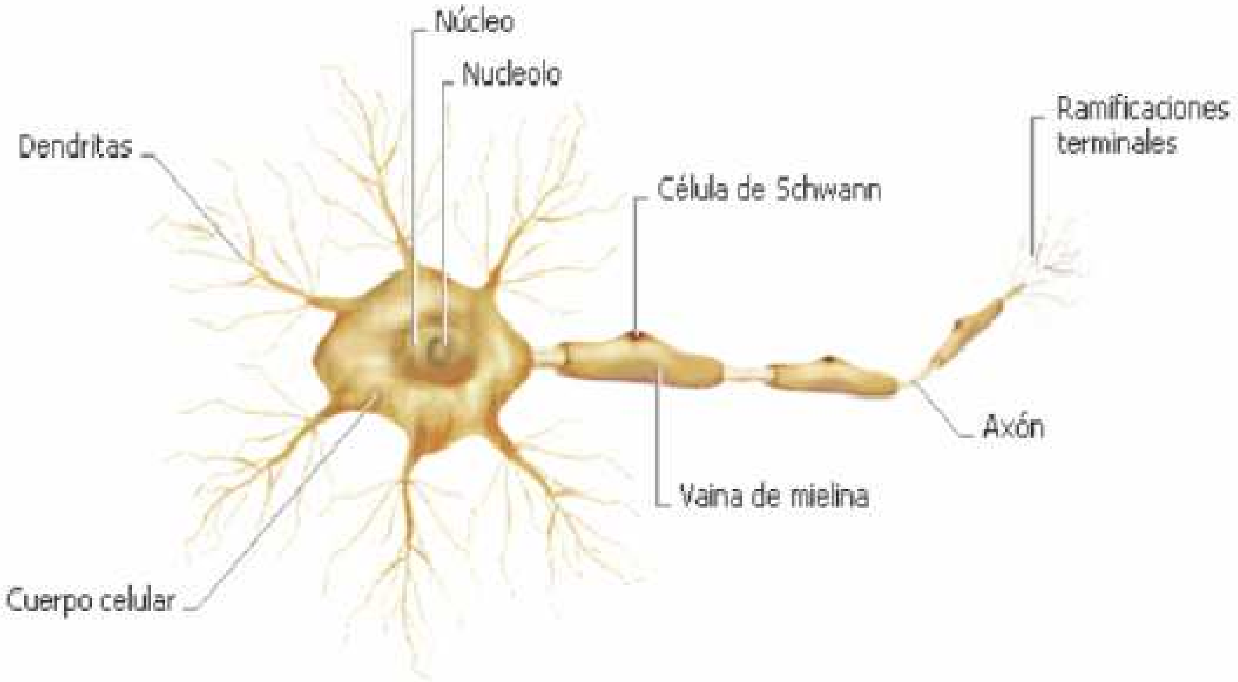


Figura 1: Una red neuronal

imagine

- Computadoras que aprendan de registros médicos y brinden tratamiento mas efectivo a nuevas enfermedades
- Casas que automáticamente ahorren energía, basado en los patrones de uso de sus habitantes

Introduccion

- Aplicaciones
  - Clasificacion de proteinas en base a su secuencia de ADN
  - Modelar reacciones químicas donde las interacciones precisas de sus reactantes es desconocida
  - Eliminacion spam
  - Biometría computacional
  - Reconocimiento de patrones
  - Minería de Datos

Algunos términos

- Función objetivo
  - Encontrar la función objetivo es la solución del problema de aprendizaje
- Hipótesis
  - Conjunto de funciones candidatas de las cuales se tratará de encontrar la función correcta
- Espacio de hipótesis
  - Conjunto de hipótesis

Algunos términos

- Algoritmo de aprendizaje
  - Es el algoritmo que toma de los datos de entrenamiento como entrada y selecciona una hipótesis del espacio de hipótesis
- Clasificación binaria
  - Problema de aprendizaje con salidas binarias, ejemplo clasificar rostros y no rostros salida si/no

Algunos términos

- Clasificación multiclase
  - Reconocer diferentes tipos de proteinas dada una secuecia de ADN
- Regresión
  - Salida dada en valores reales, por ejemplo encontrar los valores de salida que modelan una reaccion química

Tipos de aprendizaje

- Supervisado

  Se necesitan dar ejemplos previos de pertenencia a determinada clase de los valores de entrada

- No supervisado

  los datos son incluidos en una clase en particular conforme avanza el proceso de aprendizaje

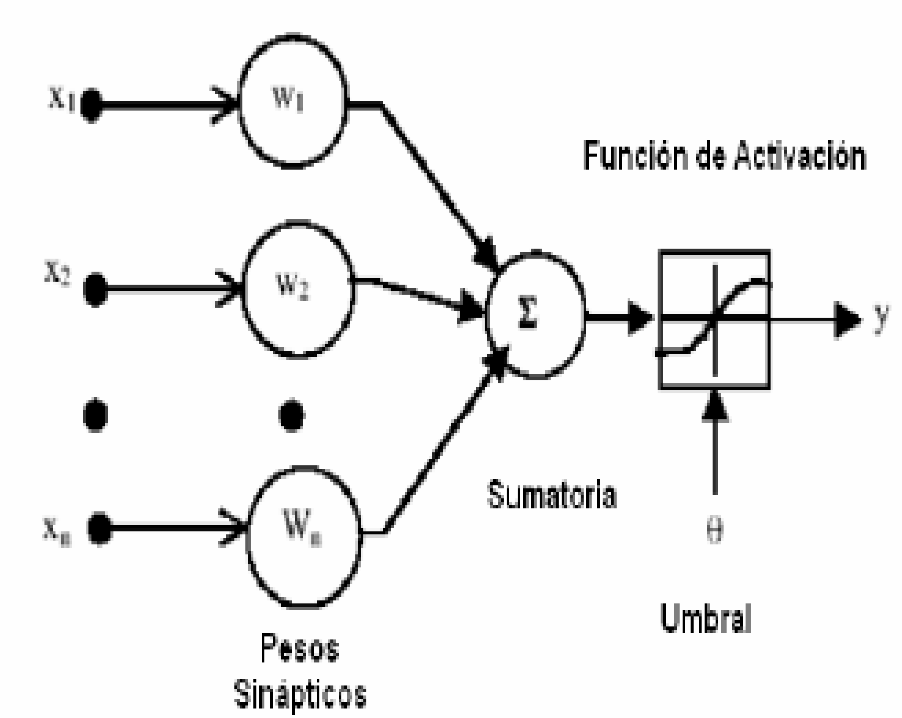
Generalizacion

- Es la habilidad de una hipótesis de clasificar correctamente datos que no pertenecen al conjunto de datos de entrenamiento

- Se desea optimizar esta función

Redes Neuronales Artificiales

$$EntradaNeurona_i = \sum_{j=1}^{n} x_j w_{ij}$$



$$f(u) == \begin{cases} 0 & \text{si } u < 0 \\ 1 & si \geq 0 \end{cases}$$

$$f(u) == \begin{cases} -1 & \text{si } u < -1 \\ u & -1 \leq u \leq 1 \\ 1 & si \geq 1 \end{cases}$$

$$f_u = \frac{1}{1 + e^{-au}}. \quad 0 \leq f(u) \leq 1$$

$$f_u = \frac{e^{\alpha u - e^{-\alpha u}}}{e^{\alpha u + e^{-\alpha u}}}$$

El perceptrón de Rosemblat

- El primer algorimo para aprendizaje de clasificaciones lineales fue propuesto por Rosemblat en 1956.

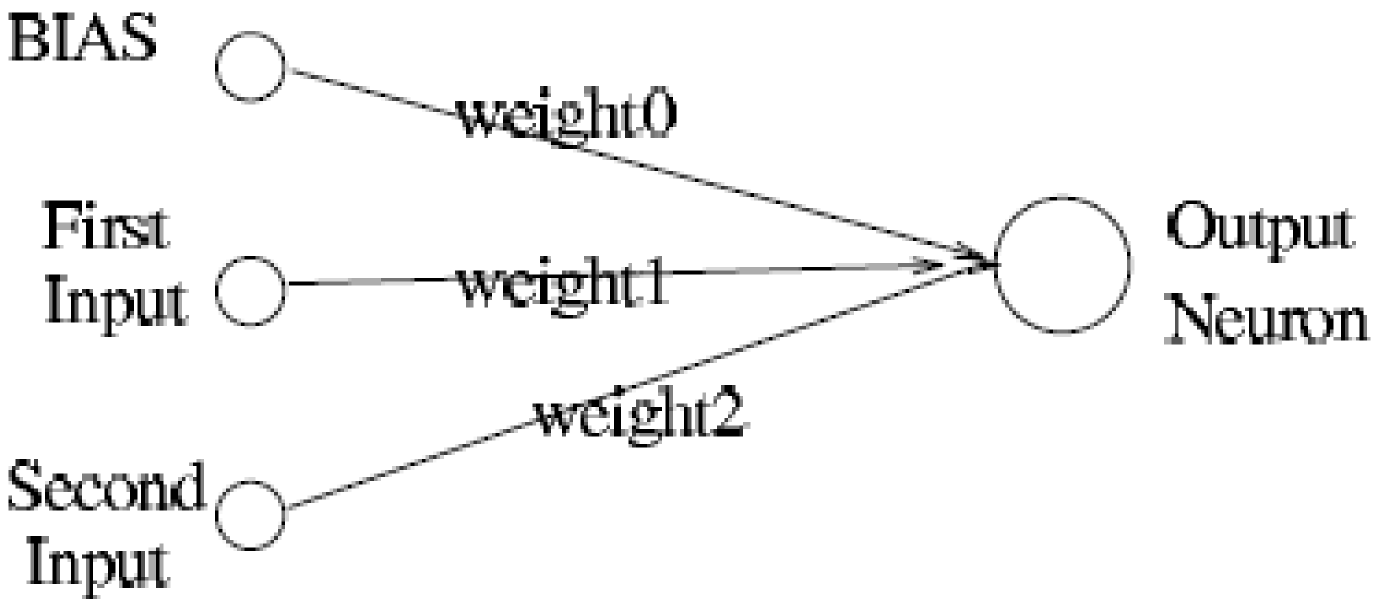
- Dado un conjunto de entrenamiento

$$S = ((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_\ell, y_\ell)) \subseteq (X \times Y)^\ell$$

- Aprendizaje

$$w_i = w_i + \eta * (y_T - y) * x_i$$

El perceptrón de Rosemblat



El perceptrón de Rosemblat

- Ejemplo
  - factor de aprendizaje =0.1

$$w_0 = 0.5, w_1 = 0.3, w_2 = 0.7$$

Bias	first input	second input	target output
1	1	1	1
1	1	-1	-1
1	-1	1	-1
1	-1	-1	-1

El perceptrón de Rosemblat

- Para patron 1
  - No cambian los pesos

- Para patron 2

$$
\begin{aligned}
w_0 &= w_0 + 0.1 * (-2) * 1 = 0.5 - 0.2 = 0.3 \\
w_1 &= w_1 + 0.1 * (-2) * 1 = 0.3 - 0.2 = 0.1 \\
w_2 &= w_2 + 0.1 * (-2) * (-1) = 0.7 + 0.2 = 0.9
\end{aligned}
$$

El perceptrón de Rosemblat

- Patron 3

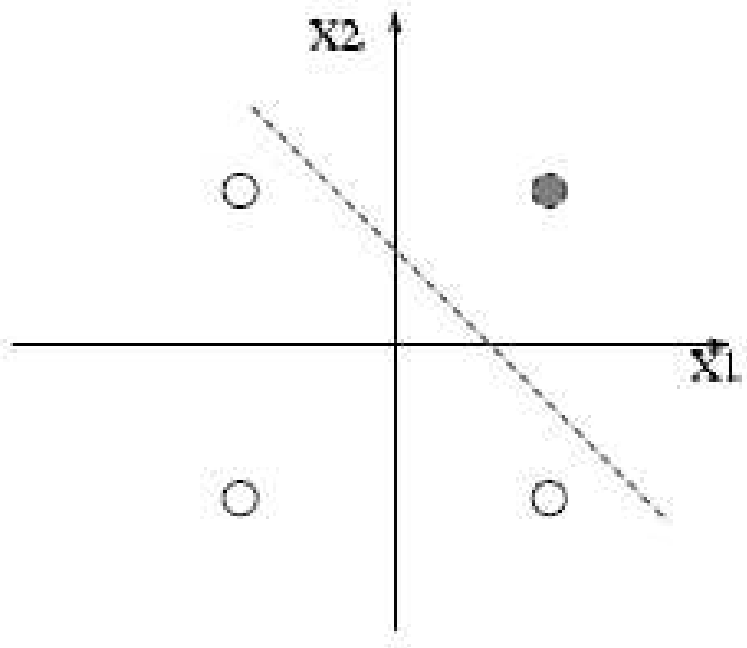
$$
\begin{aligned}
w_0 &= w_0 + 0.1 * (-2) * 1 = 0.3 - 0.2 = 0.1 \\
w_1 &= w_1 + 0.1 * (-2) * (-1) = 0.1 + 0.2 = 0.3 \\
w_2 &= w_2 + 0.1 * (-2) * 1 = 0.9 - 0.2 = 0.7
\end{aligned}
$$

- Patron 4
  - No cambia
- Patron 2

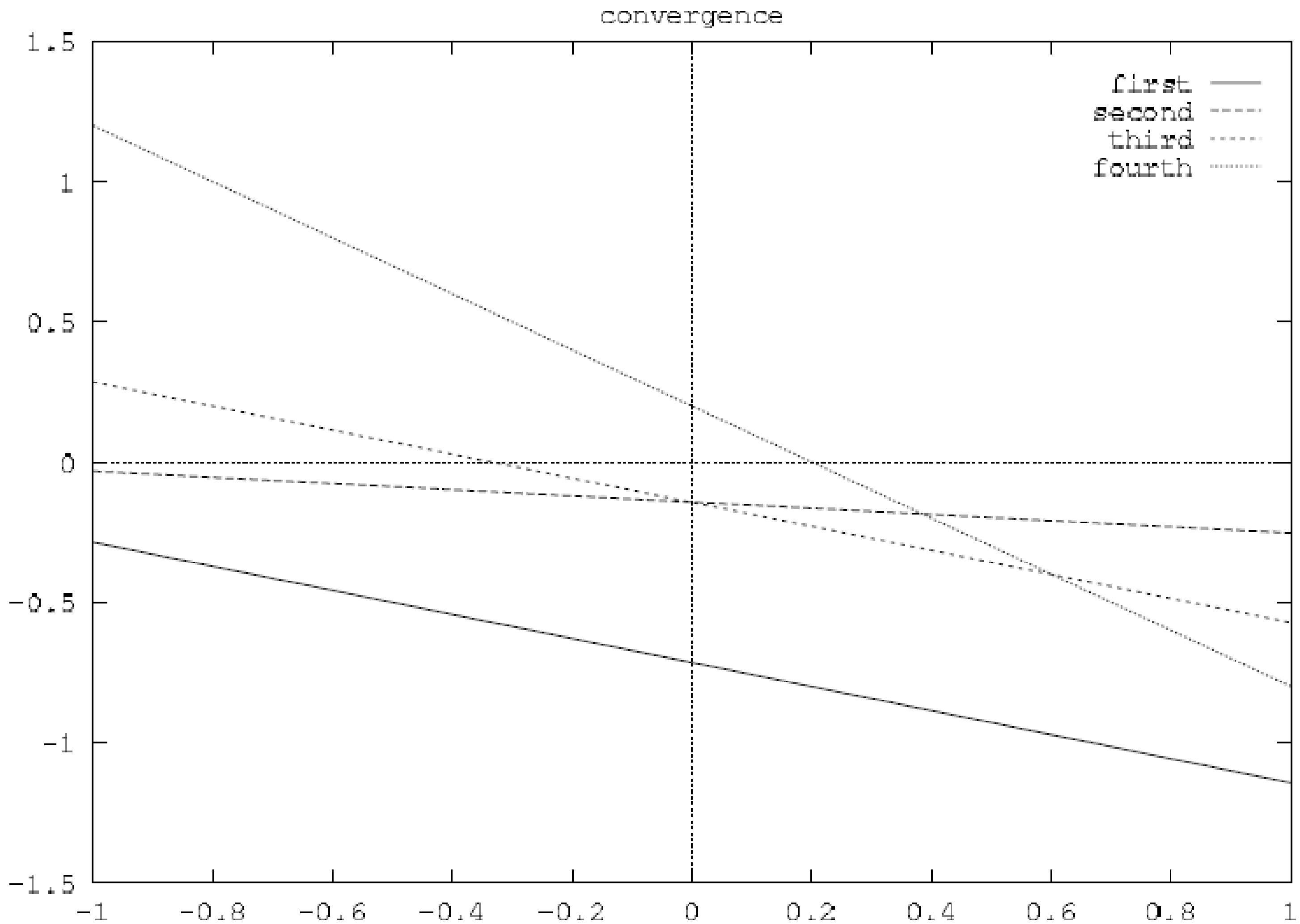
$$
\begin{aligned}
w_0 &= w_0 + 0.1 * (-2) * 1 = 0.1 - 0.2 = -0.1 \\
w_1 &= w_1 + 0.1 * (-2) * (-1) = 0.3 + 0.2 = 0.5 \\
w_2 &= w_2 + 0.1 * (-2) * 1 = 0.7 - 0.2 = 0.5
\end{aligned}
$$

El perceptrón de Rosemblat

$$0.5x_1 + 0.5x_2 - 0.1 = 0$$



El perceptrón de Rosemblat



Ejercicio

- Clasificar los siguientes nueces :

Nuez	Tipo A-1	Tipo A-2	Tipo A-3	Tipo A-4	Tipo A-5	Tipo A-6
Largo	2.2	1.5	0.6	2.3	1.3	0.3
Peso	1.4	1.0	0.5	2.0	1.5	1.0

- Entrenar la red con los siguientes vectores de entrada:

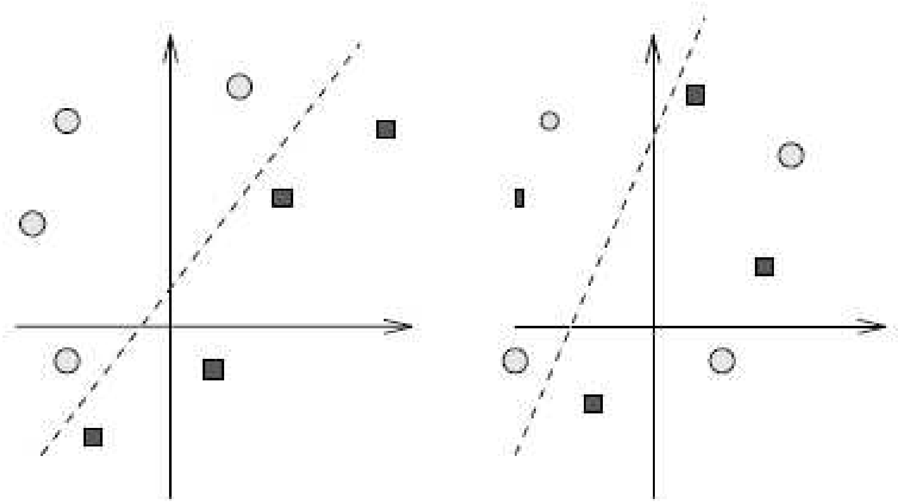
$(1, 2.2, 1.4)$
$(1, 1.5, 1.0)$
$(1, 0.6, 0.5)$

Asociar 1 clase A

$(1, 2.3, 2.0)$
$(1, 1.3, 1.5)$
$(1, 0.3, 1.0)$

Asociar -1 clase B

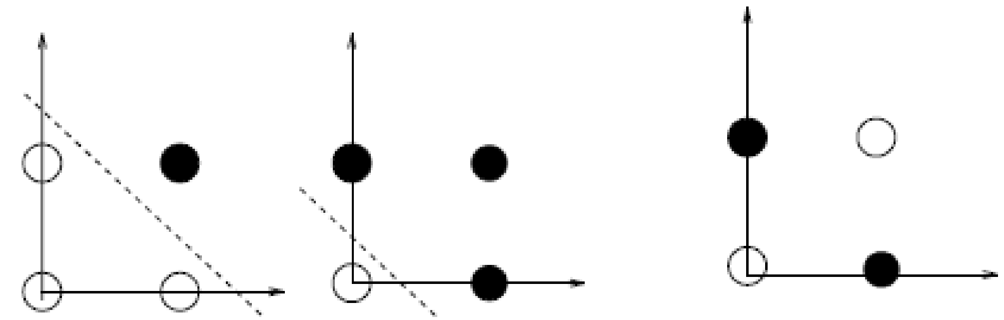
El perceptrón de Rosemblat



X	1	1	0	0
Y	1	0	1	0
X and Y	1	0	0	0

X	1	1	0	0
Y	1	0	1	0
X OR Y	1	1	1	0

X	1	1	0	0
Y	1	0	1	0
X XOR Y	0	1	1	0



El perceptrón de Rosemblat forma primal

- Dado un conjunto linealmente separable S y una tasa de aprendizaje $\eta \in \mathbb{R}^+$

$$\mathbf{w}_0 \leftarrow \mathbf{0};\ b_0 \leftarrow 0;\ k \leftarrow 0$$

$$R \leftarrow \max_{1 \le i \le \ell} \|\mathbf{x}_i\|$$

repeat

  for $i = 1$ **to** ℓ

    if $y_i(\langle \mathbf{w}_k \cdot \mathbf{x}_i \rangle + b_k) \le 0$ **then**

      $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k + \eta y_i \mathbf{x}_i$

      $b_{k+1} \leftarrow b_k + \eta y_i R^2$

      $k \leftarrow k + 1$

    end if

  end for

Hasta que no haya ningun error dentro del for

return $(\mathbf{w}_k, b_k)$ donde ***k*** es el número de errores

El perceptrón de Rosemblat forma Dual

- Como se podrá observar el vetcor de pesos ***w*** está expresado como combinación lineal de ***x*** y de ***y***

- Una representación alternativa del perceptrón se tiene teniendo en cuenta lo siguiente:

$$\mathbf{w} = \sum_{i=1}^{\ell} \alpha_i y_i \mathbf{x}_i,$$

El perceptrón de Rosemblat forma Dual

- La función de decisión puede ser escrita rn coordenadas duales

$$
\begin{aligned}
h(\mathbf{x}) &= \operatorname{sgn}\left(\langle \mathbf{w} \cdot \mathbf{x} \rangle + b\right) \\
&= \operatorname{sgn}\left(\left\langle \sum_{j=1}^{\ell} \alpha_j y_j \mathbf{x}_j \cdot \mathbf{x} \right\rangle + b\right) \\
&= \operatorname{sgn}\left(\sum_{j=1}^{\ell} \alpha_j y_j \langle \mathbf{x}_j \cdot \mathbf{x} \rangle + b\right),
\end{aligned}
$$

El perceptrón de Rosemblat forma Dual

- Dado un conjunto de entrenamiento ***S***

$\alpha \leftarrow \mathbf{0}; b \leftarrow 0$

$R \leftarrow \max_{1 \leq i \leq \ell} ||\mathbf{x}_i||$

repeat

$\quad$ **for** $i = 1$ **to** ℓ

$\quad\quad$ **if** $y_i \left(\sum_{j=1}^{\ell} \alpha_j y_j \langle \mathbf{x}_j \cdot \mathbf{x}_i \rangle + b \right) \leq 0$ **then**

$\quad\quad\quad \alpha_i \leftarrow \alpha_i + 1$

$\quad\quad\quad b \leftarrow b + y_i R^2$

$\quad\quad$ **end if**

$\quad$ **end for**

Hasta que no haya ningun error dentro del for

return $(\boldsymbol{\alpha}, b)$ que define la función ***h(x)***

El perceptrón de Rosemblat forma Dual

- Esta representación alternativa del perceptron con su función de decisión tiene propiedades interesantes, por ejemplo los puntos que son más dificiles de aprender tienen un valor alto de alfa

Convergencia del Perceptrón

- Prueba de Novikoff
  - Sea ***S*** un conjunto de entrenamiento no trivial

$$R = \max_{1 \le i \le \ell} \|\mathbf{x}_i\| .$$

  - Suponer que exite un vector $\boldsymbol{w}_{opt}$ $: \|\mathbf{w}_{opt}\| = 1$
  - Además $y_i(\langle \mathbf{w}_{opt} \cdot \mathbf{x}_i \rangle + b_{opt}) \ge \gamma$
  - El número de mistakes en el algoritmo del preceptrón es a lo mucho $\left(\frac{2R}{\gamma}\right)^2$

Ejercicio

- Cuales son los valores de w, b y k para la función AND? En la forma primal

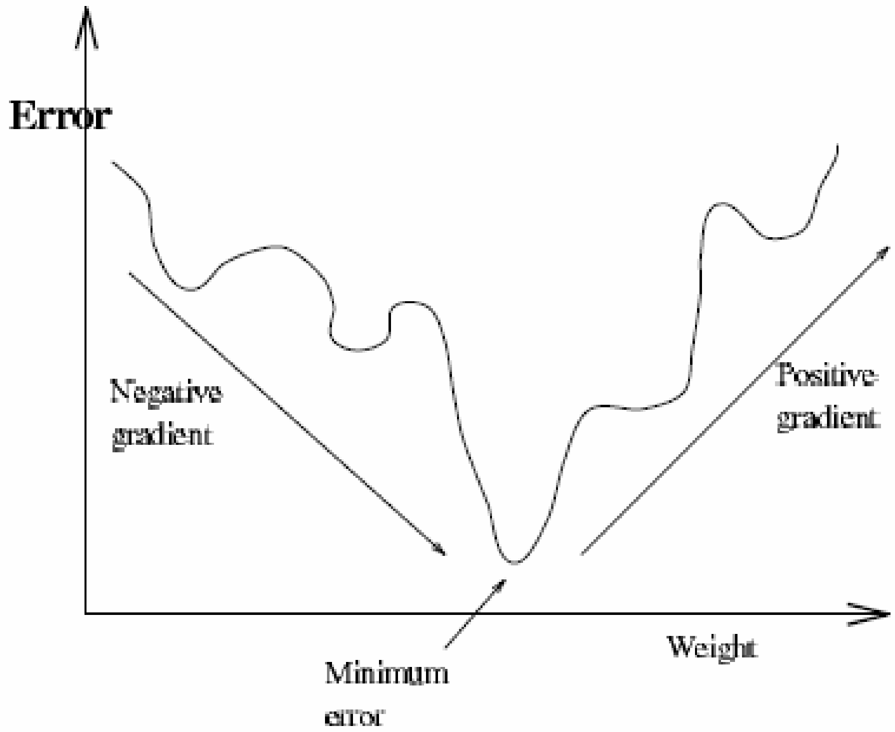
- Cuales son los valores de alfa, b y k para la función AND en forma dual

Red Adaline

- Es una generalización del perceptron,
- Adaline, significa (ADAptative LINear Element)
- Aplicaciones:
  - Filtros de ruido
  - Filtros adaptativos

Red Adaline

- Regla de aprendizaje
  - Aprendizaje LMS o regla Delta o regla Widrow hoff



Red Adaline

- Regla de aprendizaje
  - Aprendizaje LMS o regla Delta o regla Widrow hoff

$$w_{t+1} = w_t + \Delta w$$

$$donde$$

$$\Delta w = \eta\left(-\frac{\partial E}{\partial w_j}\right)$$

Red Adaline

- Regla de aprendizaje
  - Aprendizaje LMS o regla Delta o regla Widrow hoff

Aprendizaje batch: se presentan todos los patrones a la vez y luego se actualizan los pesos

$$\Delta w = \eta \delta x_j$$

$$\Delta w = \eta \sum_p (d - y) f' x_j$$

Aprendizaje online: Se presenta solo un patron a la vez y luego se realiza la actualización de pesos

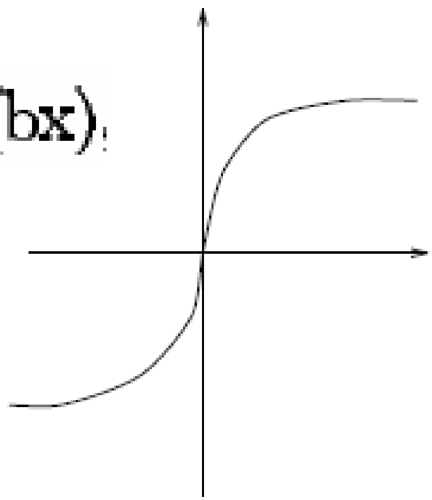
$$\Delta w = \eta \delta x_j$$

$$\Delta w = \eta (d - y) f' x_j$$

Red Adaline

- SI f(x) es tanh su derivada es

$$f(x) = \tanh(bx)$$

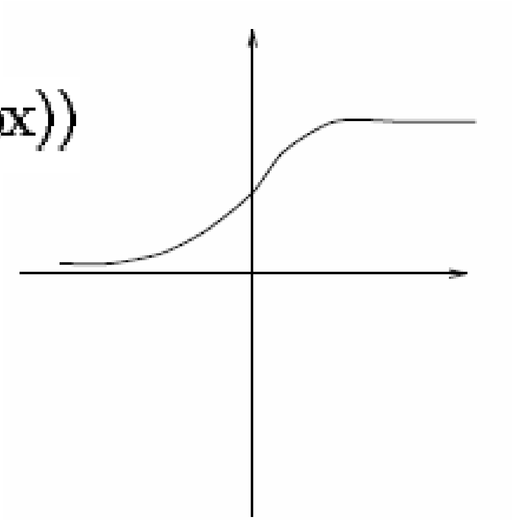


$$f'(a) = b(1 - f(a)*f(a))$$

Red Adaline

- SI f(x) es sigmoidal su derivada es

$$f(x) = 1/(1+\exp(-bx))$$



$$f'(a)=bf(a)(1 - f(a))$$

Red Adaline

- Ejemplo para n=1

$$w_0 = 0.7,\ w_1 = 0.5\ ,\ w_2 = 0.5.$$

$$\begin{aligned} Act &= w_0 * 1 + w_1 * 0.6 + w_2 * 0.5 \\ &= 0.7 + 0.3 + 0.25 = 1.25 \end{aligned}$$

$$\underline{y} = \tanh(\text{act}) = \tanh(1.25) \approx 0.8$$

Red Adaline

- Actualizando pesos

$$
\begin{aligned}
\Delta w_0 &= 1 * (1 - 0.8) * (1 - 0.8 * 0.8) * 1 \approx 0.06 \\
w_0 &\rightarrow 0.76 \\
\Delta w_1 &= 1 * (1 - 0.8) * (1 - 0.8 * 0.8) * 0.6 \approx 0.04 \\
w_1 &\rightarrow 0.54 \\
\Delta w_2 &= 1 * (1 - 0.8) * (1 - 0.8 * 0.8) * 0.5 \approx 0.03 \\
w_2 &\rightarrow 0.53
\end{aligned}
$$

Red Adaline

- Criterios de parada
  - Cuando para todos los patrones el error es menor que 0.1
  - Cuando el promedio de los ultimas epocas fue menor que un valor
  - Cuando el error por epoca es menor que un determinado valor