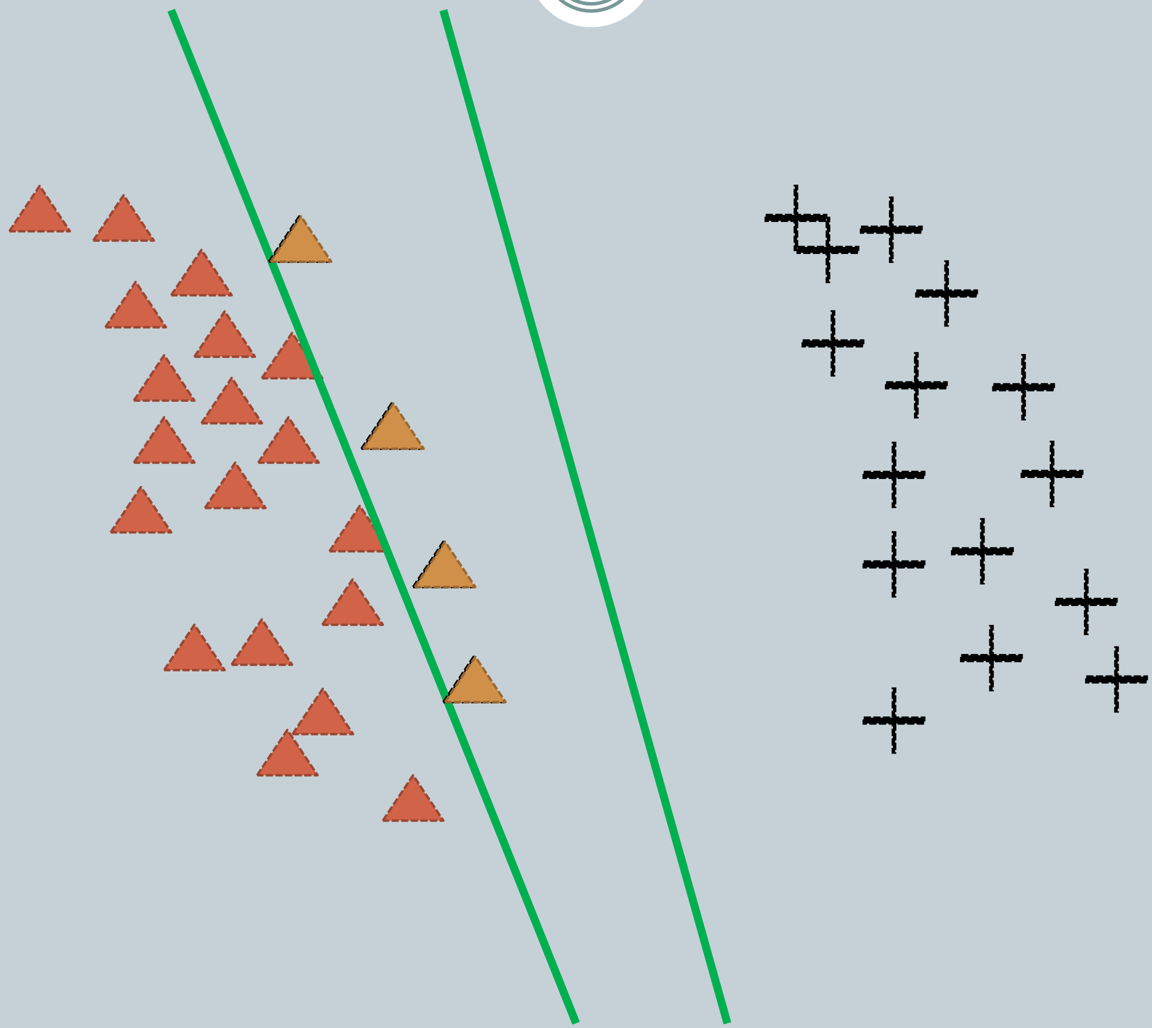
Adaline

ADAPTATIVE LINEAR NEURON
(WIDROW 1960)

Desventajas del perceptrón

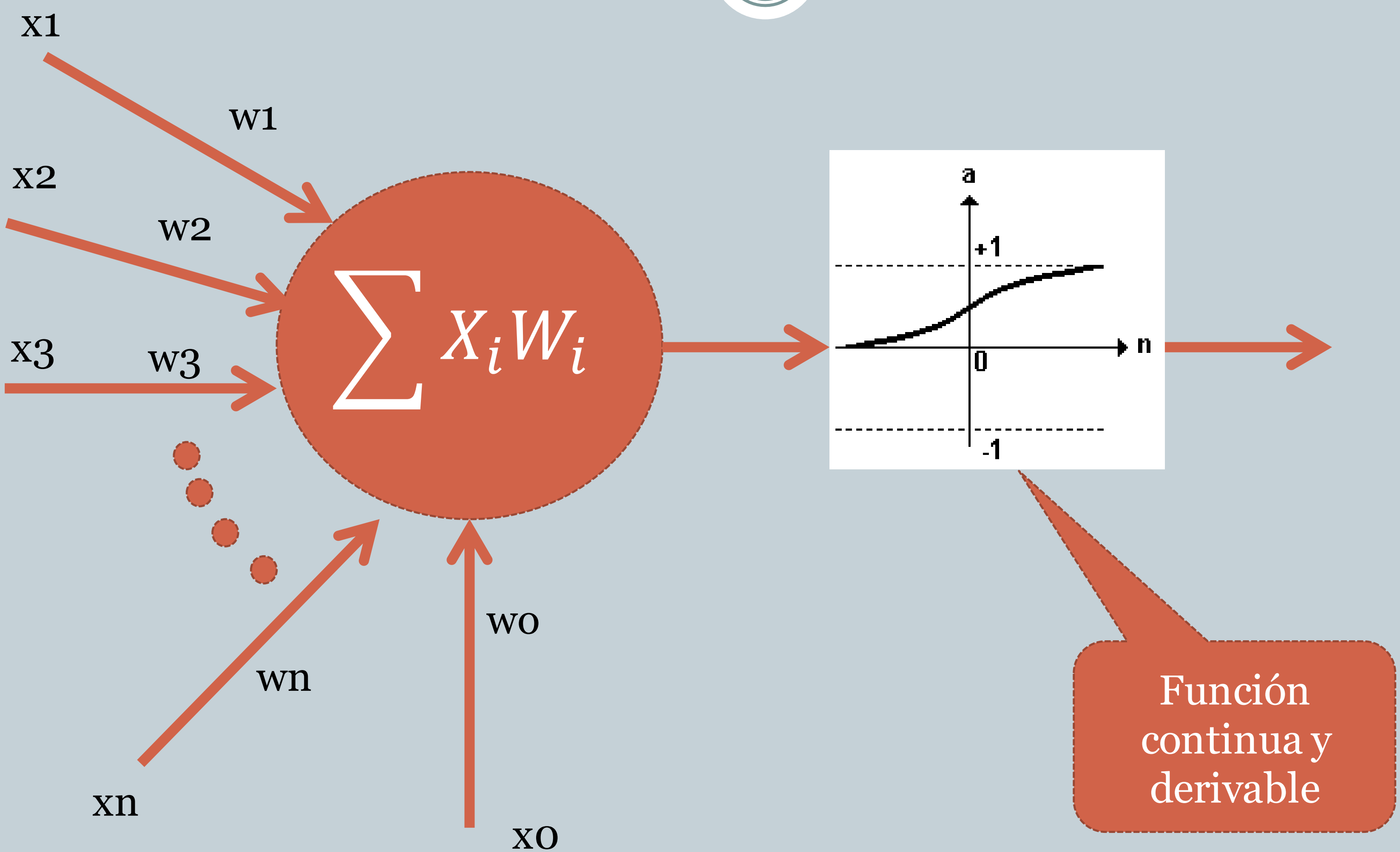


Adaline

- Al mismo tiempo que Rosenblatt trabajaba en el modelo del Perceptrón, **Widrow** introdujo el modelo de la red ADALINE y su regla de aprendizaje llamado algoritmo LMS (Least Mean Square). 1959.

- Características
  - Utiliza una función de transferencia lineal.
  - Resuelve problemas linealmente separables.
  - La regla LMS es superior al aprendizaje del Perceptrón ya que minimiza el error cuadrático medio.

Adaline – Asociador lineal



Asociador lineal

- Busca minimizar

$$\xi = < \varepsilon_k^2 > = < (d_k - w^t x_k)^2 > = {}^1\!/_L [\sum_{k=1}^{L} (d_k - \sum_{i=0}^{N} x_{ik} w_i)^2]$$

donde

L : Cantidad de Patrones

N : Cantidad de neuronas de entrada

< > representa promedio

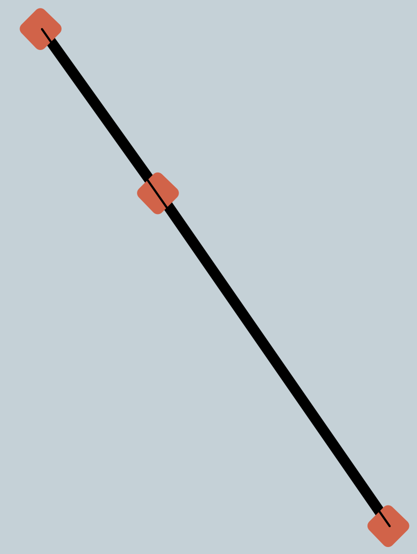
d_k : salida esperada para el patrón x_k

Asociador lineal

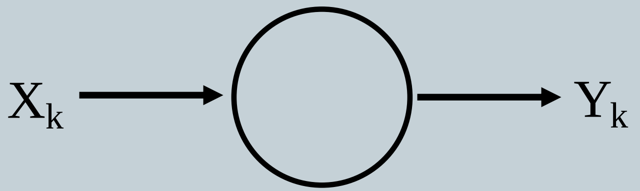
- Entrenar un asociador lineal utilizando los siguientes patrones: (-1,2), (0,0), (2,-4).

- La ecuación será de la forma

$$y = w_1 x$$



Note que **y** es la respuesta del asociador lineal.



Asociador lineal

- Se busca minimizar $\xi = <\varepsilon^2> = {}^{1}\!/\!_{L}[\sum_{k=1}^{L}(d_k - \sum_{i=0}^{N} x_{ik} w_i)^2]$

$$\xi = <\varepsilon^2> = \frac{1}{3}\left((d_1 - w_1 x_1)^2 + (d_2 - w_1 x_2)^2 + (d_3 - w_1 x_3)^2\right)$$

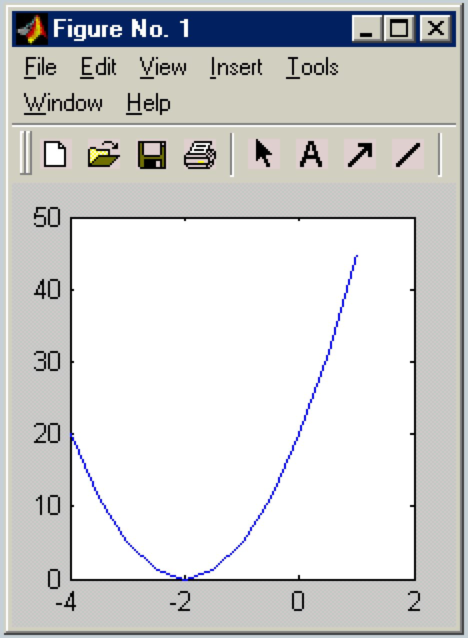
$$\xi = \frac{1}{3}\left((2 - w_1(-1))^2 + (0 - w_1 0)^2 + (-4 - w_1 2)^2\right)$$

$$\xi = \frac{1}{3}\left((2 + w_1)^2 + (-4 - 2w_1)^2\right)$$

$$\xi = \frac{1}{3}\left(4 + 4w_1 + w_1^2 + 16 + 16w_1 + 4w_1^2\right) = \frac{1}{3}(20 + 20w_1 + 5w_1^2)$$

Asociador lineal

$$\xi = \frac{1}{3}(20 + 20w_1 + 5w_1^2)$$



- Derivando

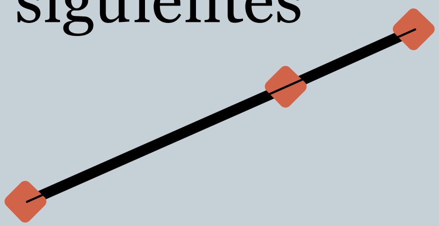
$$\frac{\partial \xi}{\partial w_1} = \frac{1}{3}(20 + 10w_1) = 0 \Rightarrow w_1 = -2$$

- Respuesta del asociador lineal

y = -2x

Asociador lineal

- Entrenar un asociador lineal utilizando los siguientes patrones: (2,3), (1,1), (-1,-3).



- La respuesta del asociador lineal será de la forma:

$$y = w_1 x + w_0$$

- Se busca minimizar:

$$\xi = \frac{1}{3}\sum_{i=1}^{3}\left(d_i - \sum_{j=0}^{1} w_j x_{ij}\right)^2 = \frac{1}{3}\sum_{i=1}^{3}\left(d_i - (w_1 x_i + w_o)\right)^2$$

Asociador lineal

- Se busca minimizar

$$\xi = \frac{1}{3}\sum_{i=1}^{3}\left(d_i - \sum_{j=0}^{1} w_j x_{ij}\right)^2 = \frac{1}{3}\sum_{i=1}^{3}\left(d_i - (w_1 x_i + w_o)\right)^2$$

$$\xi = \frac{1}{3}\left(\left(3-\left(w_1 2 + w_o\right)\right)^2 + \left(1-\left(w_1 + w_o\right)\right)^2 + \left(-3-\left(w_1(-1) + w_o\right)\right)^2\right)$$

$$\xi = \frac{1}{3}\left(\left(3-2w_1 - w_0\right)^2 + \left(1-w_1 - w_o\right)^2 + \left(-3+w_1 - w_0\right)^2\right)$$

Asociador lineal

$$\xi = \frac{1}{3}\left(\left(3-2w_1-w_0\right)^2+\left(1-w_1-w_o\right)^2+\left(-3+w_1-w_0\right)^2\right)$$

$$\frac{\partial \xi}{\partial w_1} = \frac{2}{3}\left((-2)(3-2w_1-w_0)+(-1)(1-w_1-w_o)+(-3+w_1-w_0)\right)$$

$$\frac{\partial \xi}{\partial w_1} = \frac{2}{3}\left(-6+4w_1+2w_0-1+w_1+w_0-3+w_1-w_0\right)$$

$$\frac{\partial \xi}{\partial w_1} = \frac{2}{3}\left(-10+6w_1+2w_0\right)=0 \quad \therefore \quad 3w_1+w_0-5=0$$

Asociador lineal

$$\xi = \frac{1}{3}\left(\left(3-2w_1-w_0\right)^2+\left(1-w_1-w_o\right)^2+\left(-3+w_1-w_0\right)^2\right)$$

$$\frac{\partial \xi}{\partial w_0} = \frac{2}{3}\left((-1)(3-2w_1-w_0)+(-1)(1-w_1-w_o)+(-1)(-3+w_1-w_0)\right)$$

$$\frac{\partial \xi}{\partial w_0} = \frac{2}{3}\left(-3+2w_1+w_0-1+w_1+w_0+3-w_1+w_0\right)$$

$$\frac{\partial \xi}{\partial w_0} = \frac{2}{3}\left(-1+2w_1+3w_0\right)=0 \quad \therefore \quad 2w_1+3w_0-1=0$$

Asociador lineal

Sólo resta resolver el sistema de ecuaciones

$$\begin{cases} 3w_1 + w_0 - 5 = 0 & \Rightarrow w_o = -3w_1 + 5 \quad (1) \\ 2w_1 + 3w_0 - 1 = 0 & (2) \end{cases}$$

Reemplazando (1) en (2)

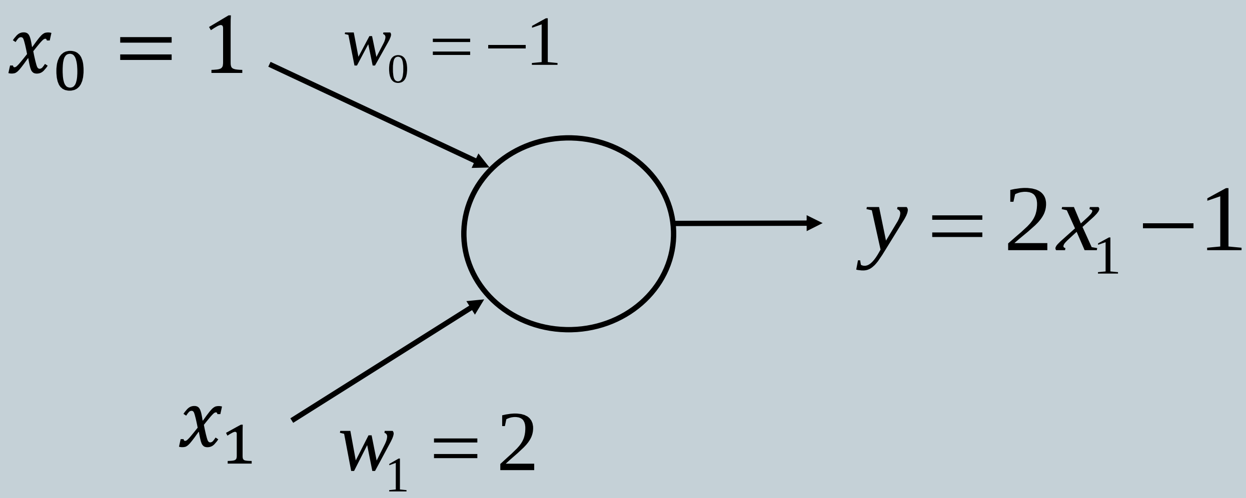
$$2w_1 - 9w_1 + 15 - 1 = 0 \quad \Rightarrow 14 - 7w_1 = 0 \Rightarrow w_1 = 2$$

De (1) $\qquad w_0 = -1$

La salida del Asociador Lineal será: $\boxed{y = 2x-1}$

Asociador lineal

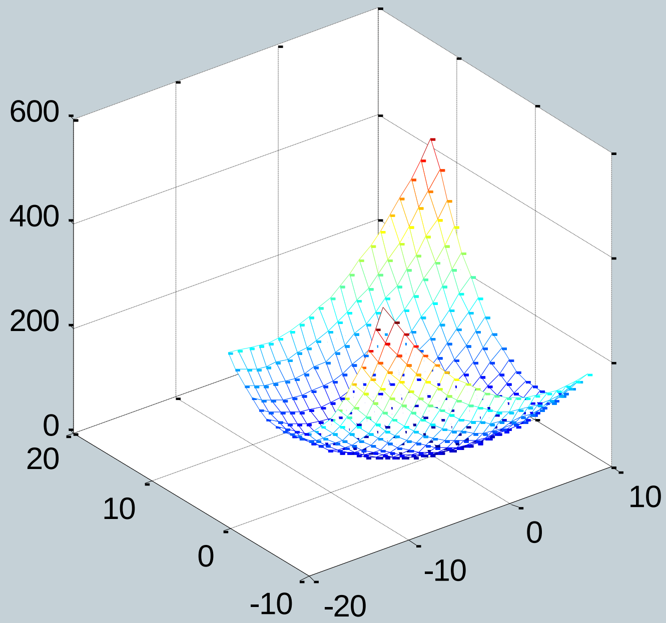
El asociador lineal a utilizar es el siguiente

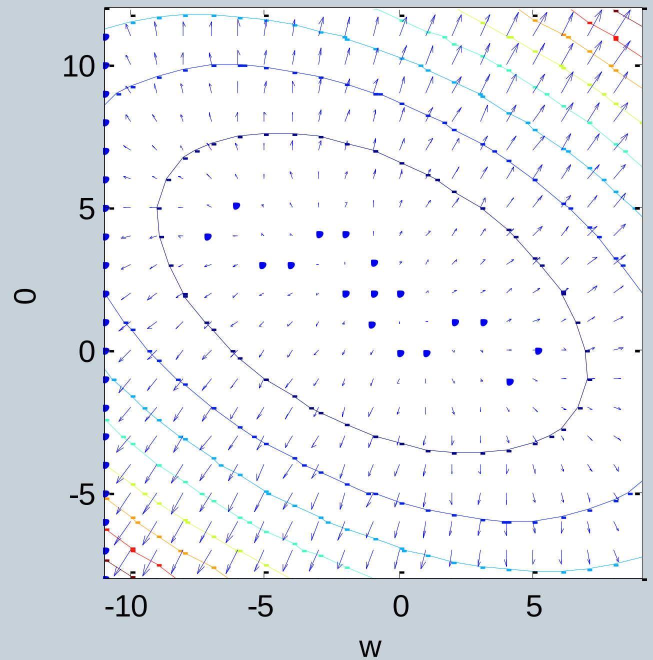


Superficie de error

La gráfica del error que minimizamos es la siguiente

$$\xi = \frac{1}{3}\left(\left(3-2w_1-w_0\right)^2+\left(1-w_1-w_o\right)^2+\left(-3+w_1-w_0\right)^2\right)$$

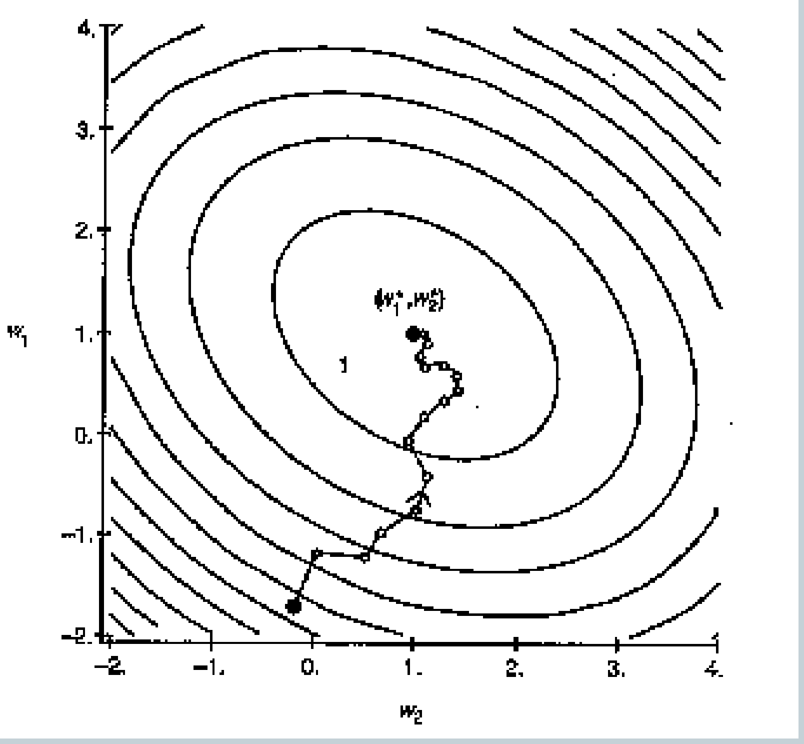




Mínimo en $\mathbf{w_o=-1}$, $\mathbf{w_1=2}$

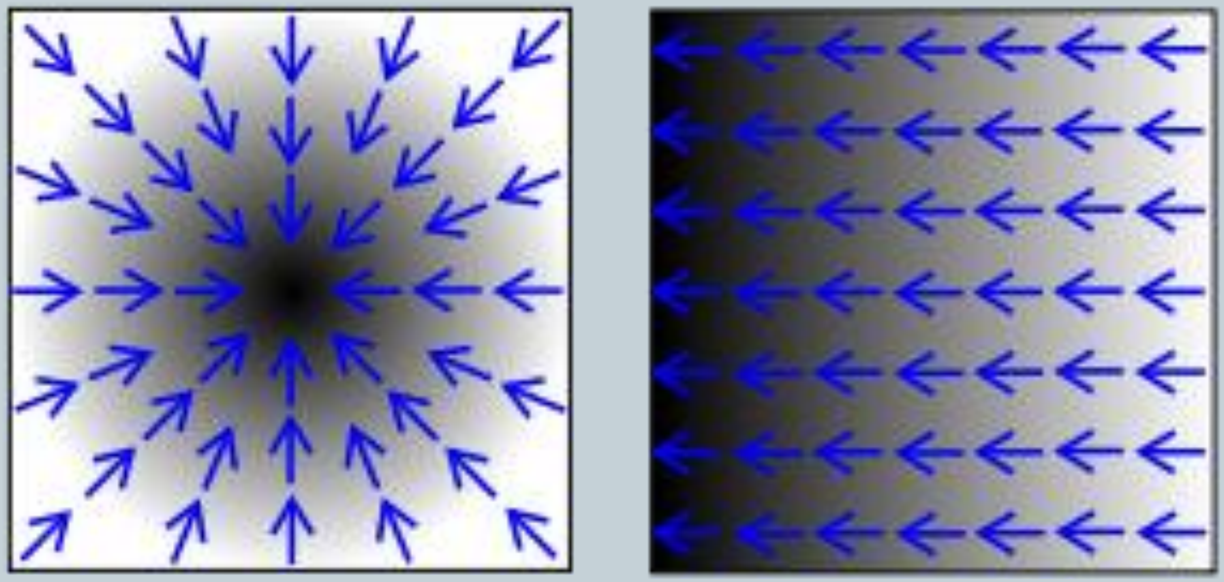
Entrenamiento del Asociador Lineal

- Se busca un método de entrenamiento que, a partir de los datos de entrada, permita calcular el vector *W*.



Concepto de Gradiente

Gradiente



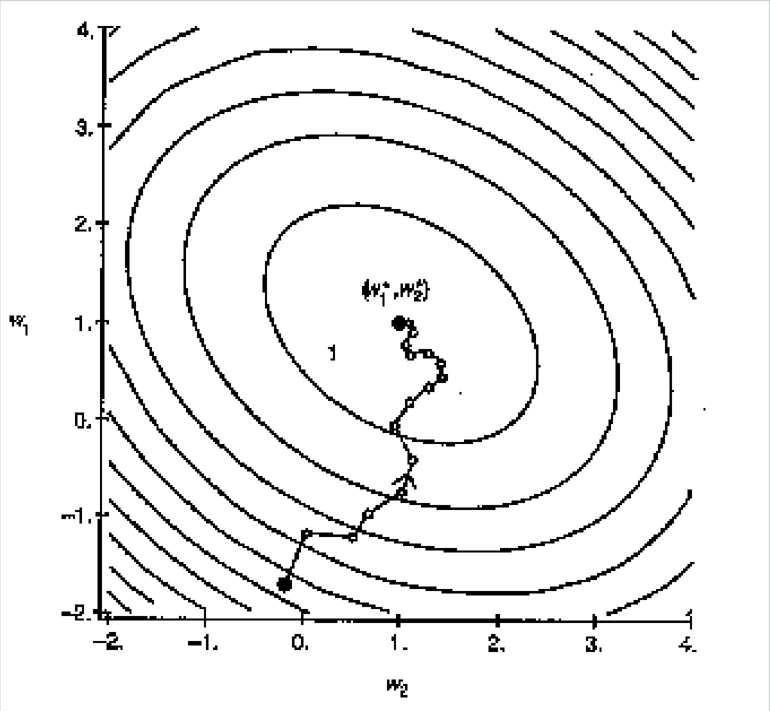
En esta imagen, el campo escalar se aprecia en blanco y negro, los cuales representan valores bajos o altos respectivamente, y el gradiente correspondiente se aprecia por flechas azules.

Minimización de funciones usando el gradiente

- Dada una función continua
  - Tomar un punto dentro del dominio de la función.
  - Calcular el vector gradiente de la función en ese punto.
  - Sumarle al punto anterior una fracción del gradiente negativo (para ir hacia el mínimo).
  - Repetir los dos pasos anteriores hasta que la diferencia entre evaluaciones consecutivas de la función sea inferior a una cierta cota.

Técnica del descenso del gradiente

$$w(t+1) = w(t) - \alpha \nabla \varepsilon_k^2(t)$$



- veamos que

$$\nabla \varepsilon_k^2(t) = -2\varepsilon_k(t) x_k$$

Técnica del gradiente

$$\nabla \varepsilon_k^2(t) = \frac{\partial \varepsilon_k^2}{\partial w} = \left[\frac{\partial (d_k - y_k)^2}{\partial w_0}; \quad \ldots; \quad \frac{\partial (d_k - y_k)^2}{\partial w_n} \right]$$

$$\nabla \varepsilon_k^2(t) = \frac{\partial \varepsilon_k^2}{\partial w} = \left[-2(d_k - y_k)\frac{\partial y_k}{\partial w_0}; \quad \ldots; \quad -2(d_k - y_k)\frac{\partial y_k}{\partial w_n} \right]$$

$$\nabla \varepsilon_k^2(t) = \frac{\partial \varepsilon_k^2}{\partial w} = -2e_k \left[\frac{\partial y_k}{\partial w_0}; \quad \ldots; \quad \frac{\partial y_k}{\partial w_n} \right]$$

$$\nabla \varepsilon_k^2(t) = \frac{\partial \varepsilon_k^2}{\partial w} = -2e_k \left[x_{0k}, \quad x_{1k}, \quad \ldots, \quad x_{nk} \right] = -2e_k x_k$$

Algoritmo iterativo de entrenamiento del Adaline

- Para cada vector de entrada
  - Aplicar el vector de entrada x_k
  - Calcular el gradiente utilizando

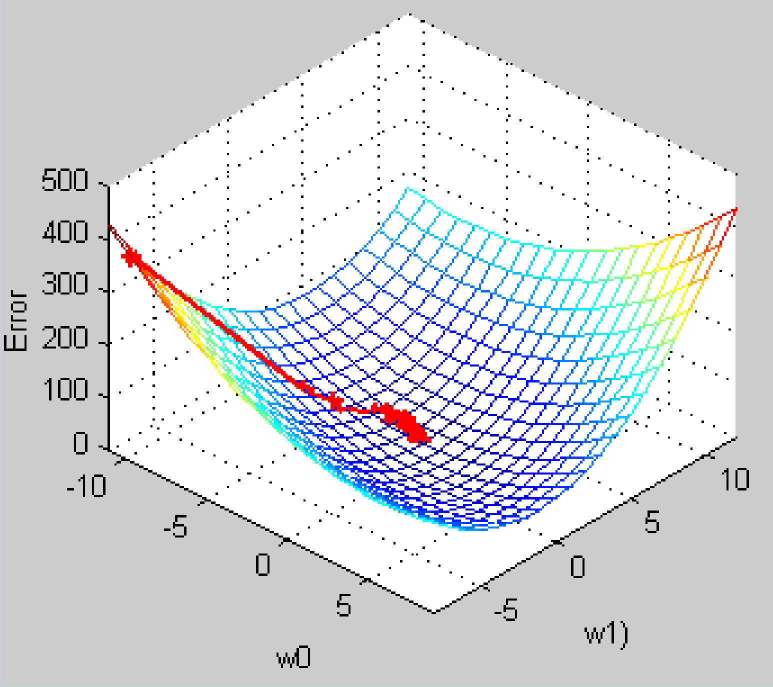
$$\nabla < \varepsilon_k^2 > \approx \nabla \varepsilon_k^2(t) = -2\varepsilon_k(t)x_k = -2(d_k - y_k)x_k$$

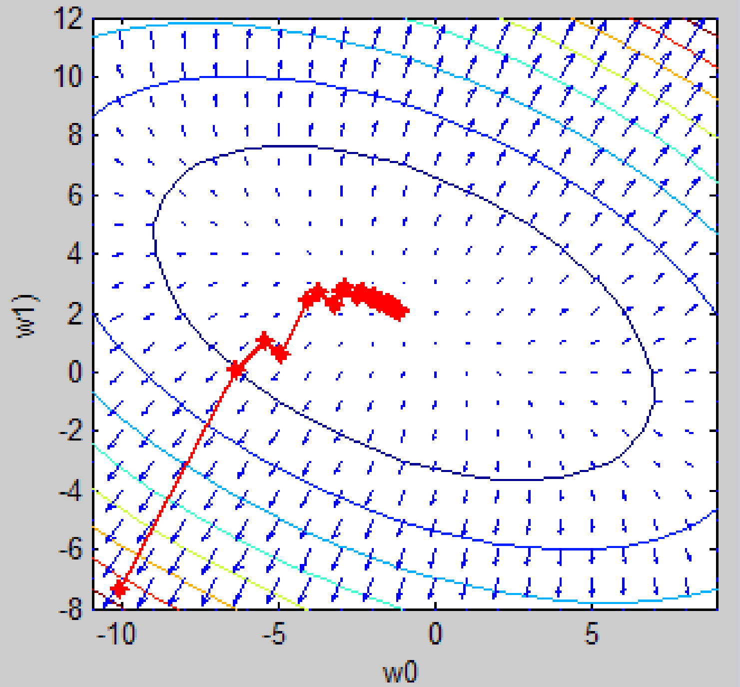
  - Actualizar el vector de pesos

$$w(t+1) = w(t) + 2\alpha(d_k - y_k)x_k$$

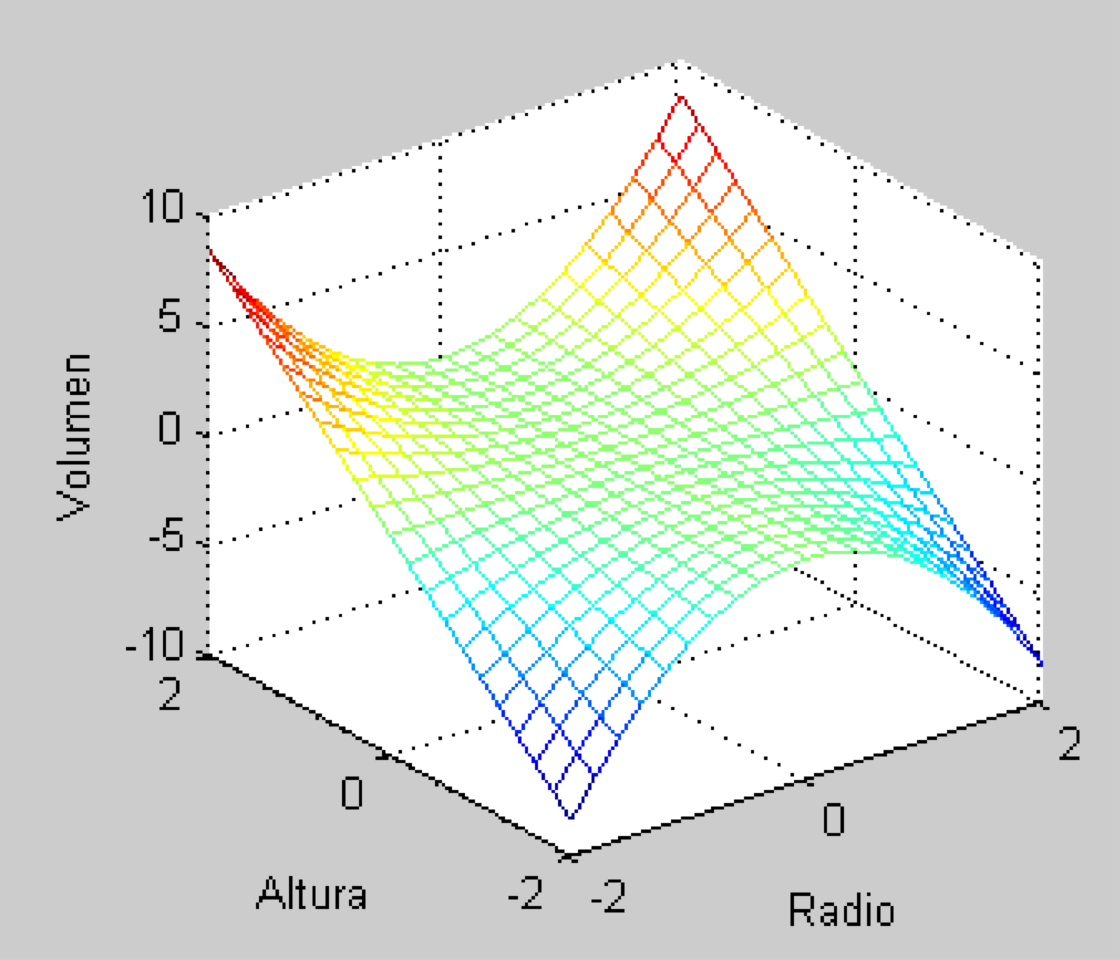
- Repetir todo hasta que el error sea aceptable

Minimización de funciones usando el gradiente

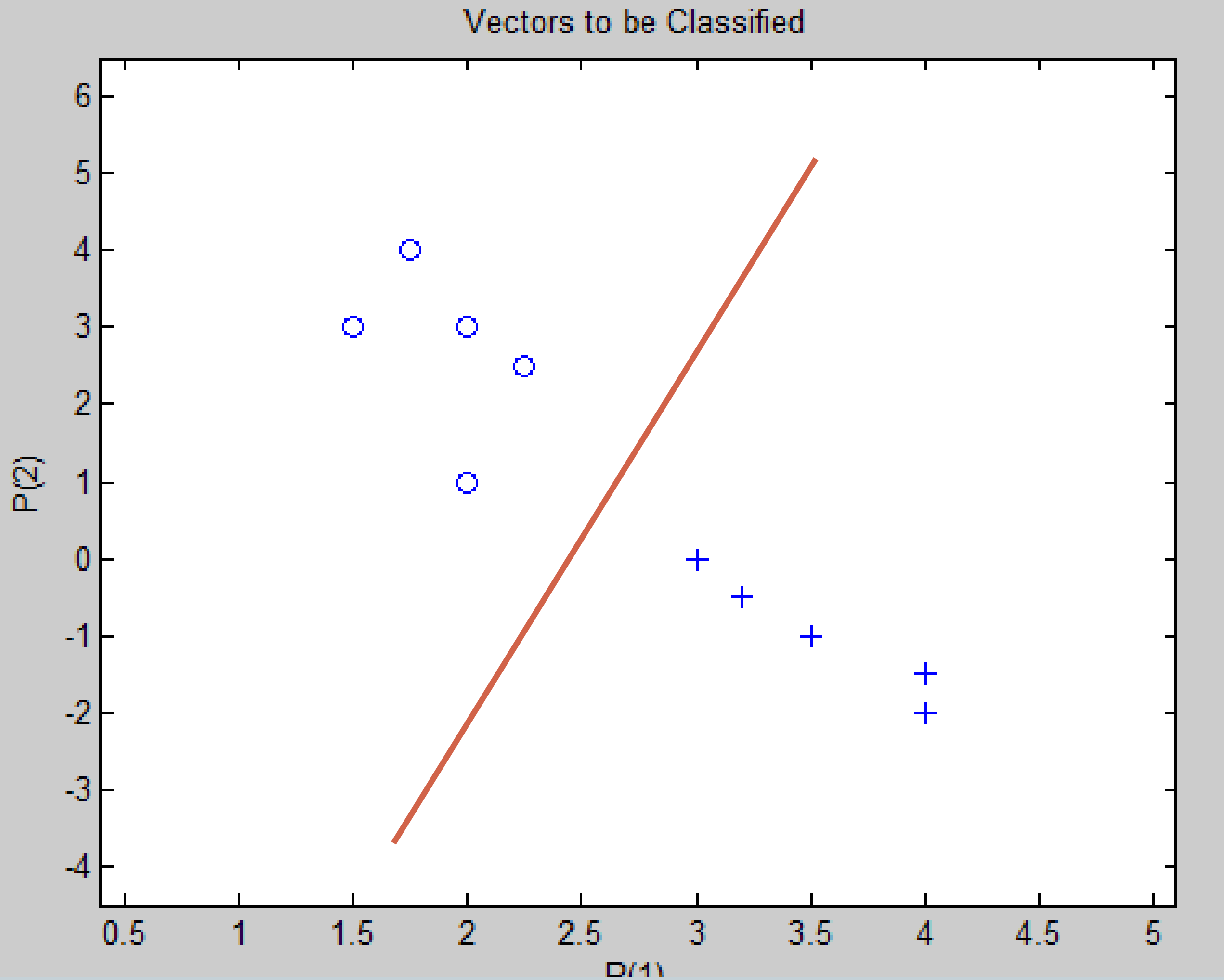




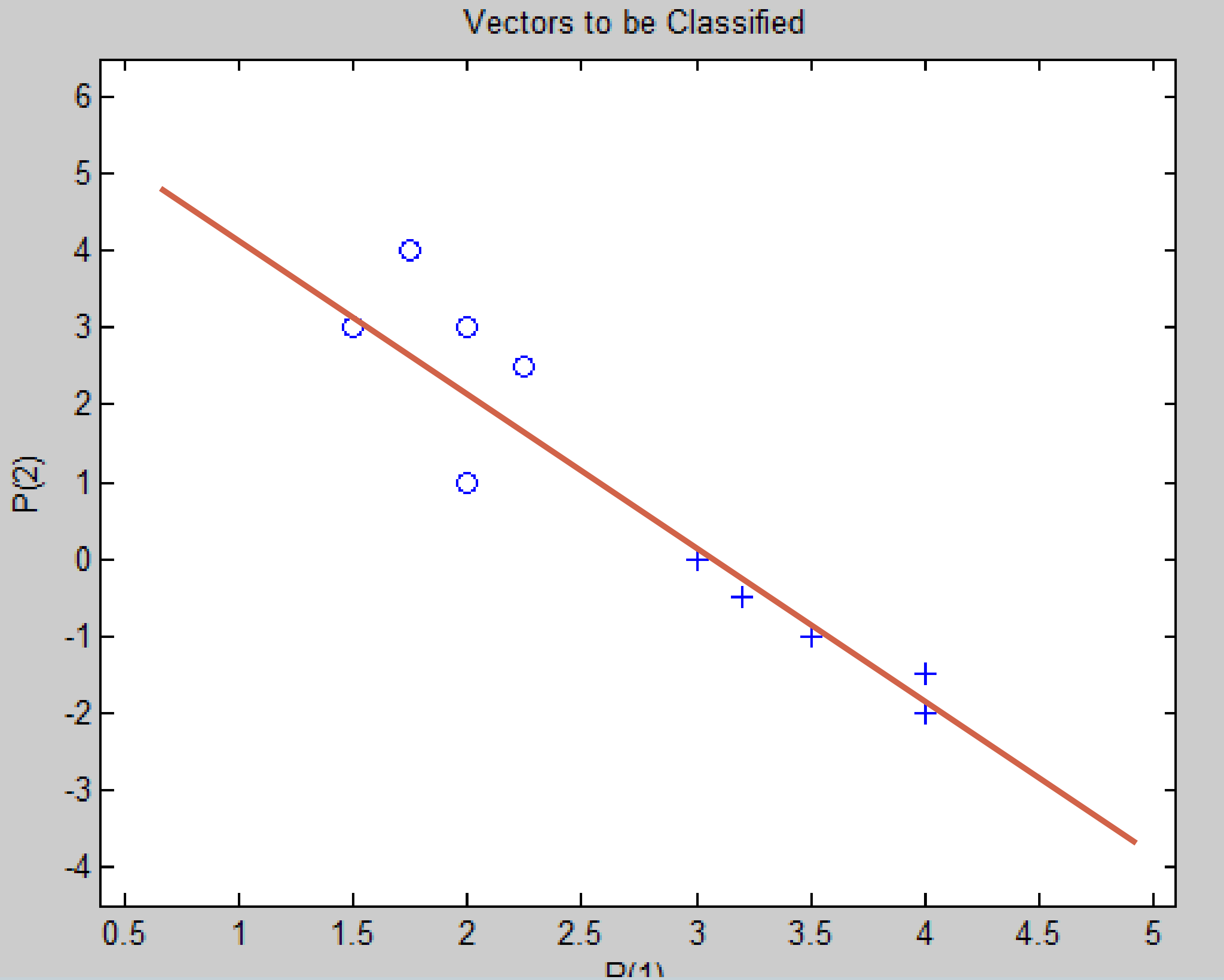
Minimización de funciones usando el gradiente



Combinador lineal

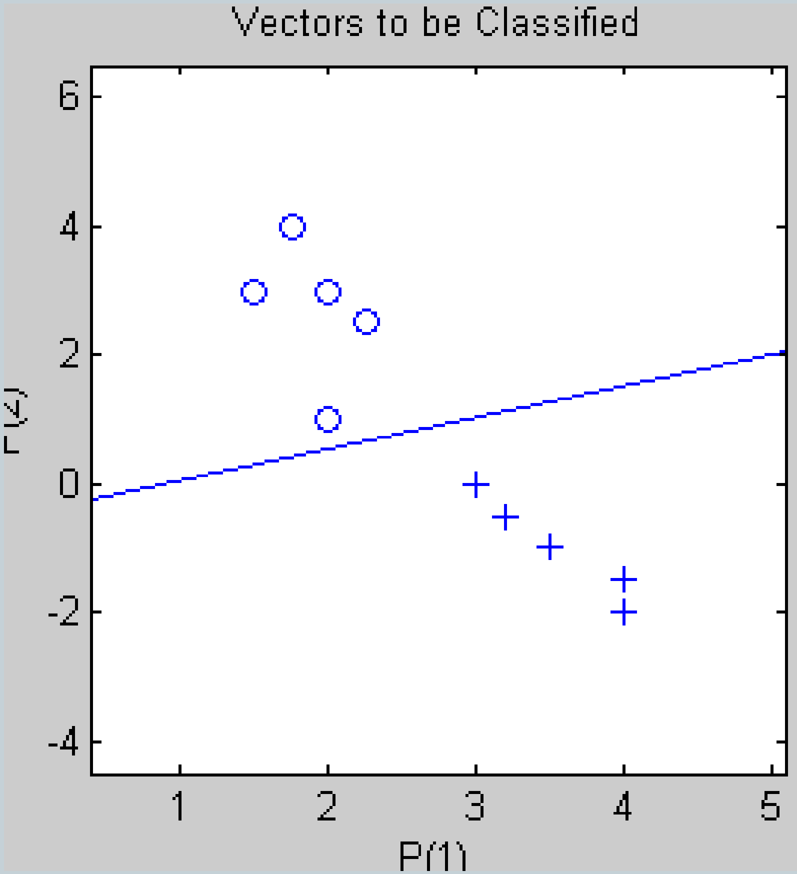


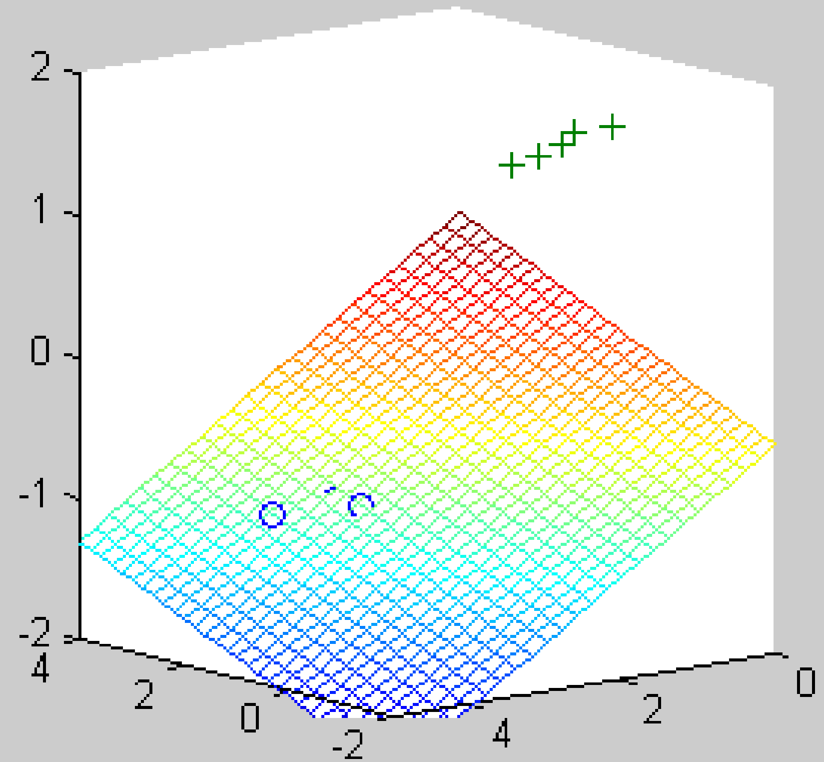
Combinador lineal



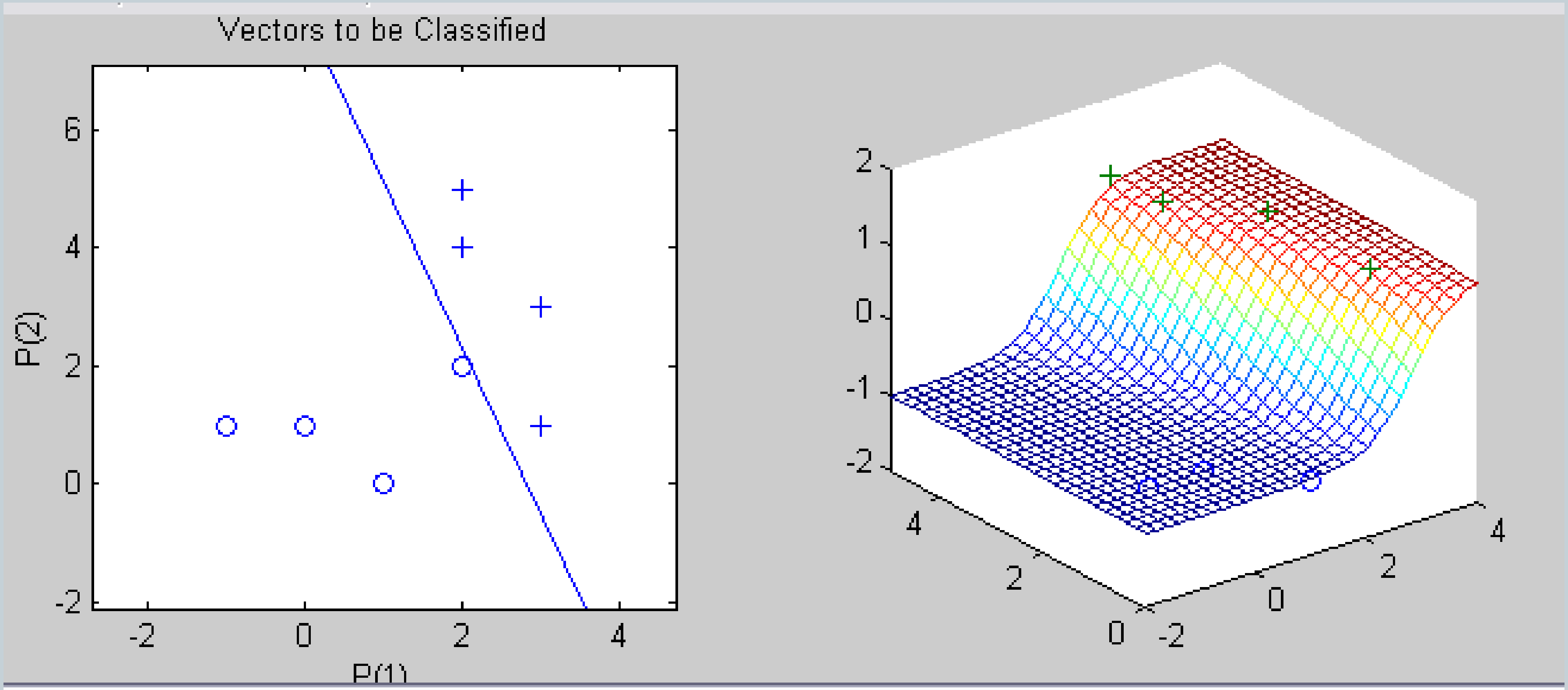
Combinador lineal

- Se agrega una nueva dimensión

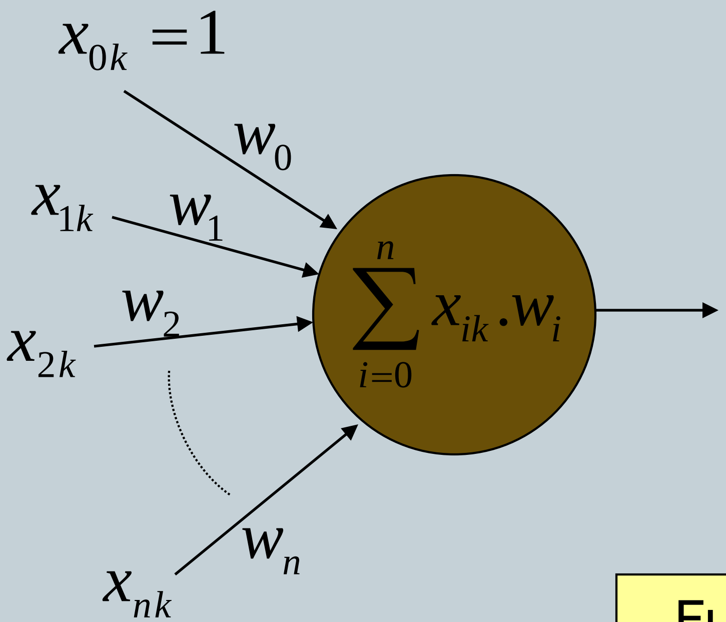




Combinador lineal



Neurona general



$$y_k = f\left(\sum_{i=0}^{n} x_{ik}.w_i\right)$$

Función derivable

¿Cómo sería la derivada del error si la neurona no es lineal?

$$\nabla \varepsilon_k^2 = \frac{\partial \varepsilon_k^2}{\partial w} = \left[\frac{\partial (d_k - y_k)^2}{\partial w_0}; \quad \ldots; \quad \frac{\partial (d_k - y_k)^2}{\partial w_n} \right]$$

$$\frac{\partial \varepsilon_k^2}{\partial w_j} = -2(d_k - y_k) \frac{\partial f_k}{\partial (neta_k)} \frac{\partial (neta_k)}{\partial w_j}$$

$$f_k^{'}$$

$$\frac{\partial (neta_k)}{\partial w_j} = \frac{\partial (\sum_{j=0}^{L} w_j x_j)}{\partial w_j} = x_j$$

$$\boxed{\frac{\partial \varepsilon_k^2}{\partial w_j} = -2(d_k - y_k) f_k'(neta_k) x_j}$$

Método iterativo

- Para cada vector de entrada
  - Aplicar el vector de entrada, x_k
  - Calcular el gradiente utilizando

$$\nabla < \varepsilon_k^2 > \approx \nabla \varepsilon_k^2(t) = -2\varepsilon_k(t)\, f_k^{'}(neta_k)\, x_k$$

  - Actualizar el vector de pesos

$$w(t+1) = w(t) + 2\alpha\varepsilon_k\ f_k^{'}(neta_k)\, x_k$$

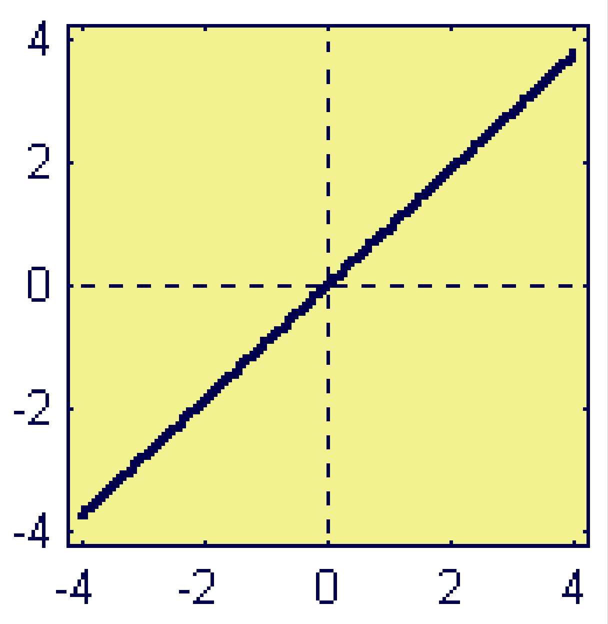
- Repetir todo hasta que el error sea aceptable

Función de Salida LINEAL

- purelin

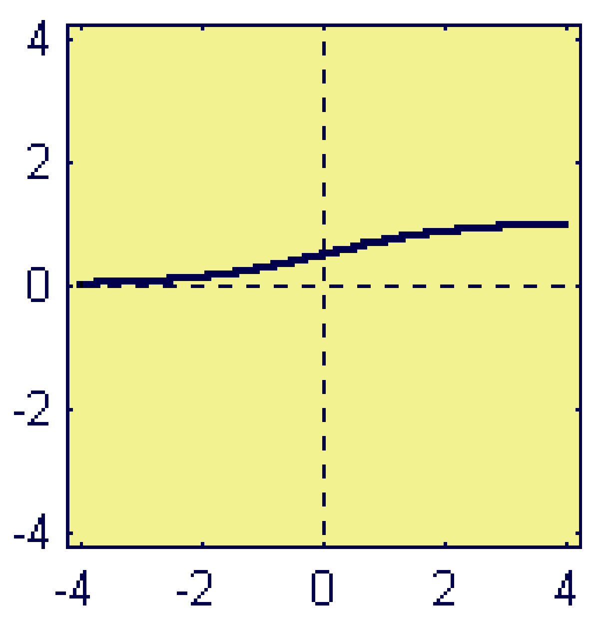
$$f(x) = x$$

$$f'(x) = 1$$



Función SIGMOIDE ∈ (0,1)

- logsig

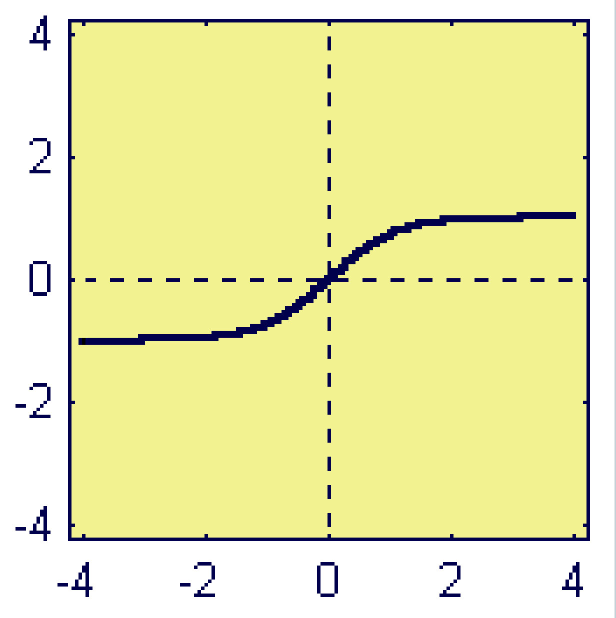


$$f(x) = \frac{1}{1+e^{-x}}$$

$$f'(x) = f(x) * (1 - f(x))$$

Función SIGMOIDE $\in$ (-1,1)

- tansig



$$f(x) = \frac{2}{1+e^{-2x}} - 1$$

$$f'(x) = 1 - f(x)^2$$

Adaline

- Bernard Widrow explicando el funcionamiento de su Adaline
- https://youtu.be/hc2Zj55j1zU
- https://youtu.be/skfNlwEbqck