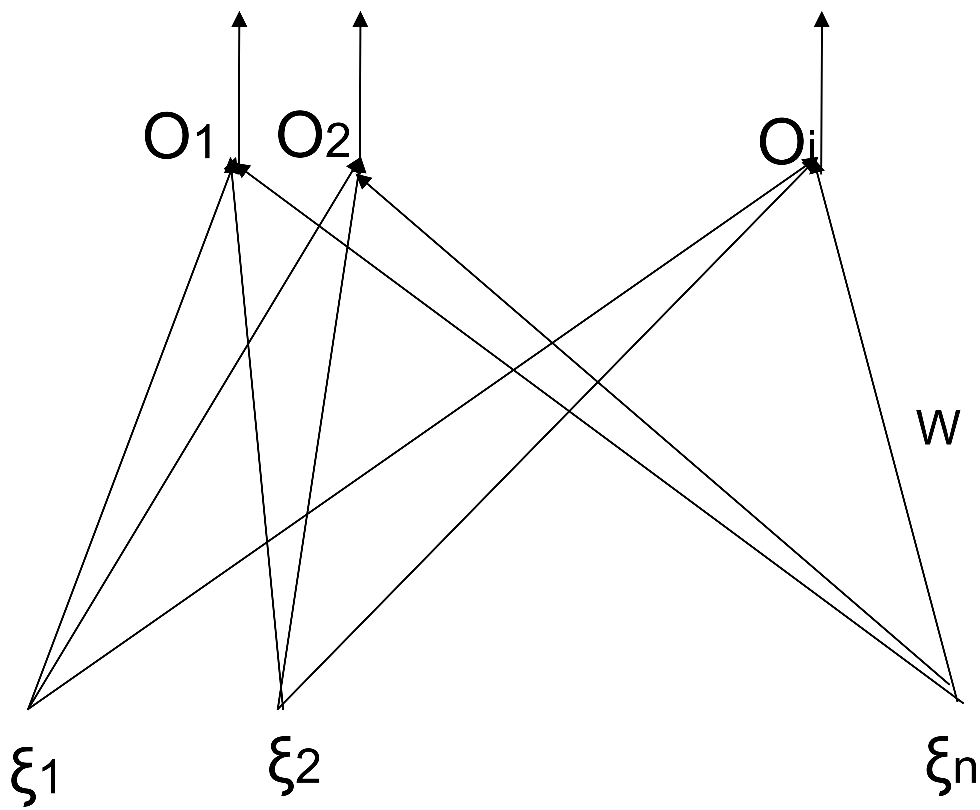
PERCEPTRON SIMPLE



$$O_i = g(h_i) = g(\Sigma\, w_{ik}\, \xi_k)$$

PERCEPTRON SIMPLE

$$O_i = g(h_i) = g\left(\sum_k w_{ik}\xi_k\right)$$

Convención: los umbrales equivalen a una neurona ξo = -1 y Wio =θi

$$O_i = g\left(\sum_{k=0}^{N} w_{ik}\xi_k\right) = g\left(\sum_{k=1}^{N} w_{ik}\xi_k - \theta_i\right)$$

$O_i^\mu = \zeta_i^\mu$ es nuestra condición de aprendizaje

$$O_i^\mu = g(h_i^\mu) = g\left(\sum_k w_{ik}\xi_k^\mu\right)$$

Ξ, O, Z booleanos o continuos

UNIDADES UMBRAL O ESCALON

g(h) = sgn (h)

Sup. ζ_i **= ±1** → buscamos $\mathrm{sgn}(\mathbf{w}\cdot\boldsymbol{\xi}_\mu) = \zeta^\mu$ para cada patrón de aprendizaje

= proyecciones de ξ sobre w coincidan en signo con ζ → la condición w ξ = 0 separa ambas clases (plano al origen, ortogonal a w) (a)

O, introduciendo el c.v.: $x_k^\mu \equiv \zeta^\mu \xi_k^\mu \qquad \mathbf{x}^\mu \equiv \zeta^\mu \boldsymbol{\xi}^\mu$

nuestra condición se reexpresa (b) : $\mathbf{w}\cdot\mathbf{x}^\mu > 0$

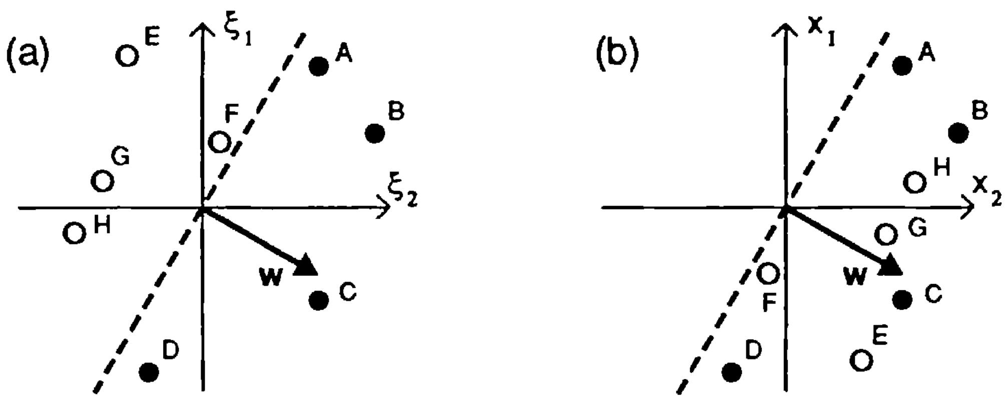


TABLE 5.1 AND function

ξ_1	ξ_2	ζ
0	0	−1
0	1	−1
1	0	−1
1	1	+1

El problema será resoluble si y sólo si es linealmente separable = existe un plano en el espacio de Ξ que separe los patrones.

Reescrito con umbral: O = sgn (w ξ – w_0) → el plano N-1 dimensional que separa es w ξ – w_0 (w_0 ordenada al origen del plano)

ALGORITMO DE APRENDIZAJE

$w_{ik}(n+1)=w_{ik}(n)+\Delta_{ik}(n)$ *con* $\Delta_{ik}(n)=\eta\,(\zeta_i - O_i)\,\xi_k$

Más general:

$$\Delta w_{ik} = \eta\,\Theta(N\kappa - \zeta_i^\mu h_i^\mu)\zeta_i^\mu \xi_k^\mu$$

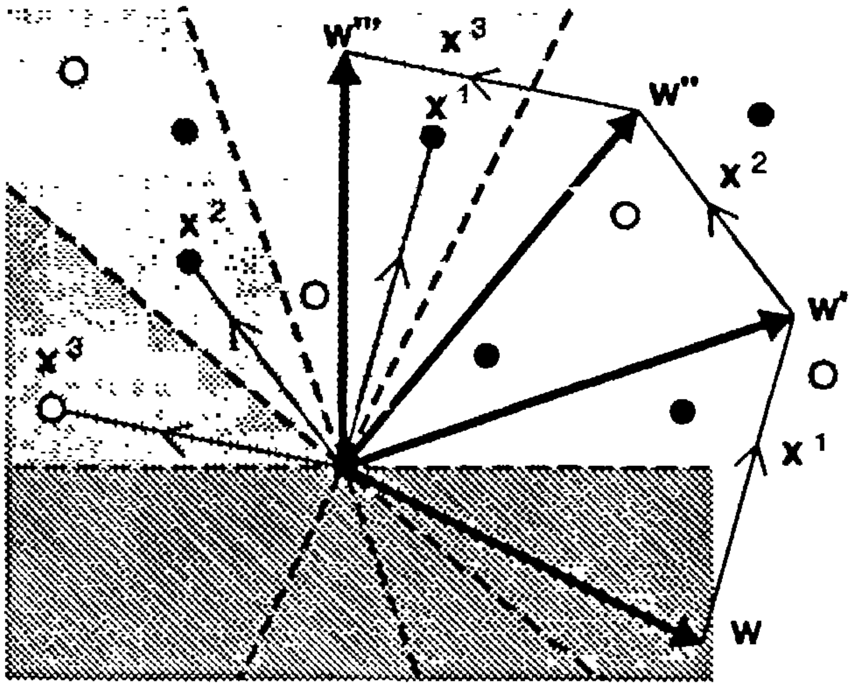
i.e la condición de aprendizaje sería

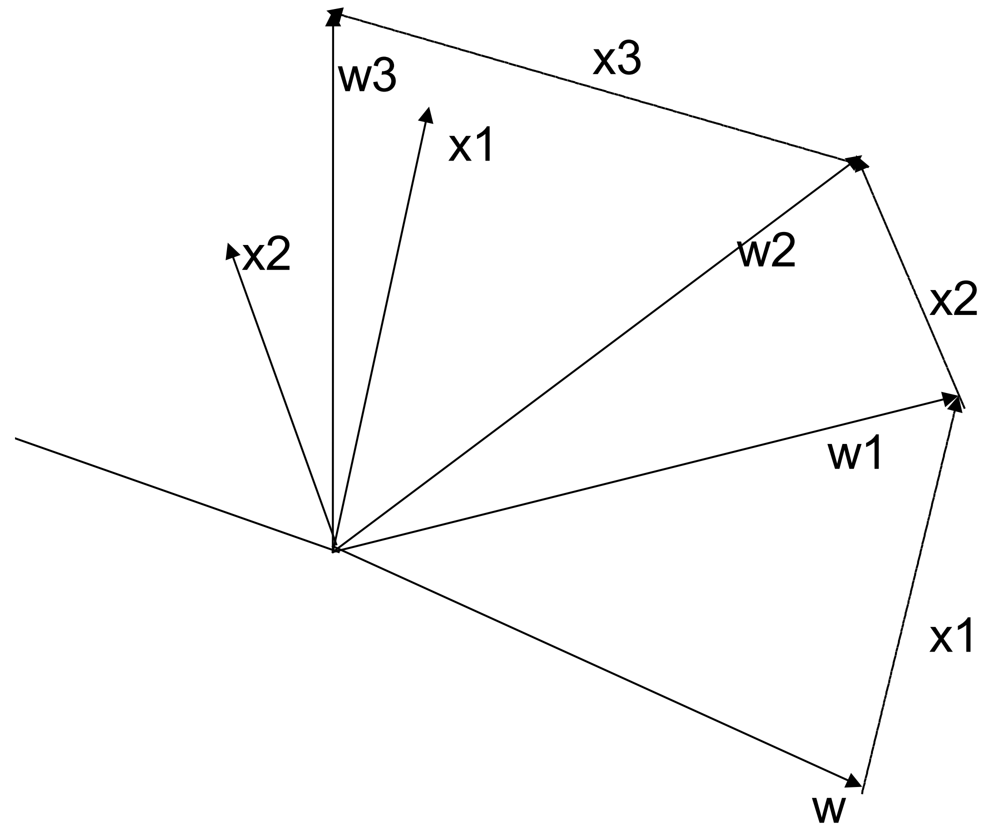
$$\zeta_i^\mu h_i^\mu \equiv \zeta_i^\mu \sum_k w_{ik}\xi_k^\mu > N\kappa$$ **en lugar de > 0**

Con una sola salida:

$$\mathbf{w}\cdot\mathbf{x}^\mu > N\kappa \qquad \Delta\mathbf{w} = \eta\,\Theta(N\kappa - \mathbf{w}\cdot\mathbf{x}^\mu)\mathbf{x}^\mu$$

→ ***todos los x a más de* Nk/lwl *del plano ortogonal a w que separa***





TEOREMA DE CONVERGENCIA (Rosenblatt)

Para un problema linealmente separable, la solución mediante la regla de aprendizaje se obtiene en un número finito de pasos.

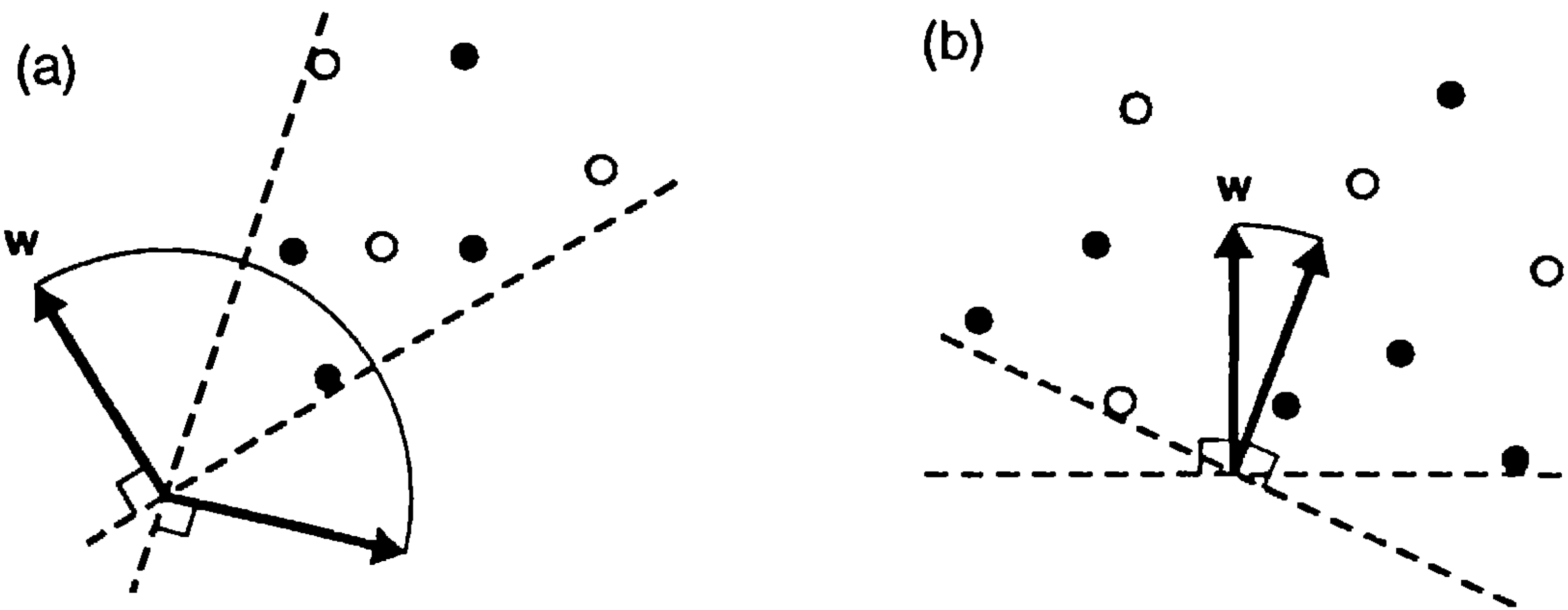
DIFICULTAD DE UN PROBLEMA (Hertz)

$$D(\mathbf{w}) = \frac{1}{|\mathbf{w}|} \min_{\mu} \mathbf{w} \cdot \mathbf{x}^{\mu} \qquad D_{\max} \equiv \max_{\mathbf{w}} D(\mathbf{w})$$

Interpretación: qué tan lejos está, para un W, el x más alejado del plano ortogonal a W, maximizado sobre todos los W.

*D*max grande → problema fácil

*D*max < 0 → problema sin solución E.g.: *D*max(AND) = 1/√17 *D*max(XOR) =-1/√3



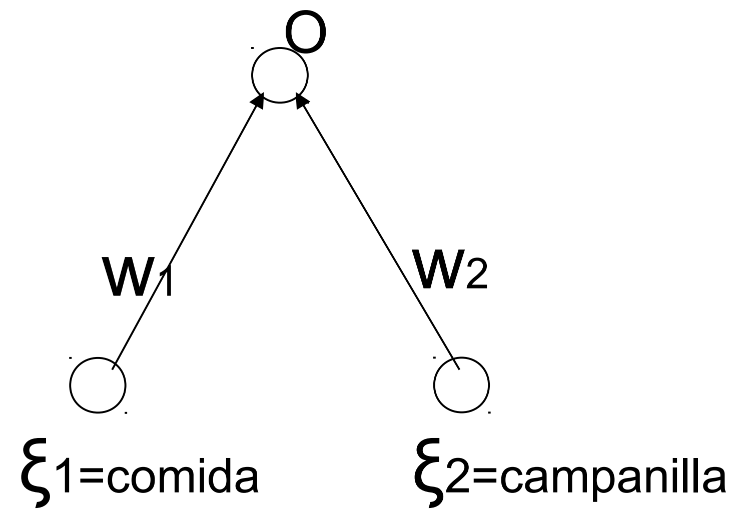
PERCEPTRONES Y REGLA DE HEBB

Cuando un axón de una célula A está suficientemente próximo para excitar a una célula B o toma parte en su disparo de forma persistente, tiene lugar algún proceso de crecimiento o algún cambio metabólico en una de las células, o en las dos, de tal modo que la eficiencia de A, como una de las células que desencadenan el disparo de B, se ve incrementada.

D. Hebb, *Organization of behaviour*, 1949

Es el *Postulado de Hebb.*

Apliquémoslo al ***experimento de Pavlov*****:** condicionamiento estímulo-respuesta, vía un perceptrón simple



Si bien w_2 no es al principio lo suficientemente excitatorio por sí solo, como

$$\Delta w_2 \approx \zeta \; \xi_2$$

si presentamos persistentemente la asociación (ξ,ζ) al sistema, el peso w_2 va a llegar a reforzarse hasta el punto de ser suficiente sin ξ_1

Aprendizaje = modificación en las sinapsis

Hebb: ¿aumento del área de la unión sináptica? o (más reciente) ¿ incremento de la velocidad con que se libera el neurotransmisor en la célula presináptica?

UNIDADES LINEALES

g*(*h*) = *h *y propongamos:*

$$E[\mathbf{w}] = \frac{1}{2}\sum_{i\mu}(\zeta_i^\mu - O_i^\mu)^2 = \frac{1}{2}\sum_{i\mu}\Big(\zeta_i^\mu - \sum_k w_{ik}\xi_k^\mu\Big)^2 .$$

Descenso por gradiente:

$$\begin{aligned}\Delta w_{ik} &= -\eta\frac{\partial E}{\partial w_{ik}} \\ &= \eta\sum_\mu(\zeta_i^\mu - O_i^\mu)\xi_k^\mu .\end{aligned}$$

$$\Delta w_{ik} = \eta(\zeta_i^\mu - O_i^\mu)\xi_k^\mu \qquad\qquad \Delta w_{ik} = \eta\delta_i^\mu\xi_k^\mu$$

$\delta_i^\mu = \zeta_i^\mu - O_i^\mu$ *.es el error en la salida i cuando se presenta la entrada μ → regla delta (nuevamente, idem unidades escalón) → coinciden el enfoque hebbiano y el del gradiente*

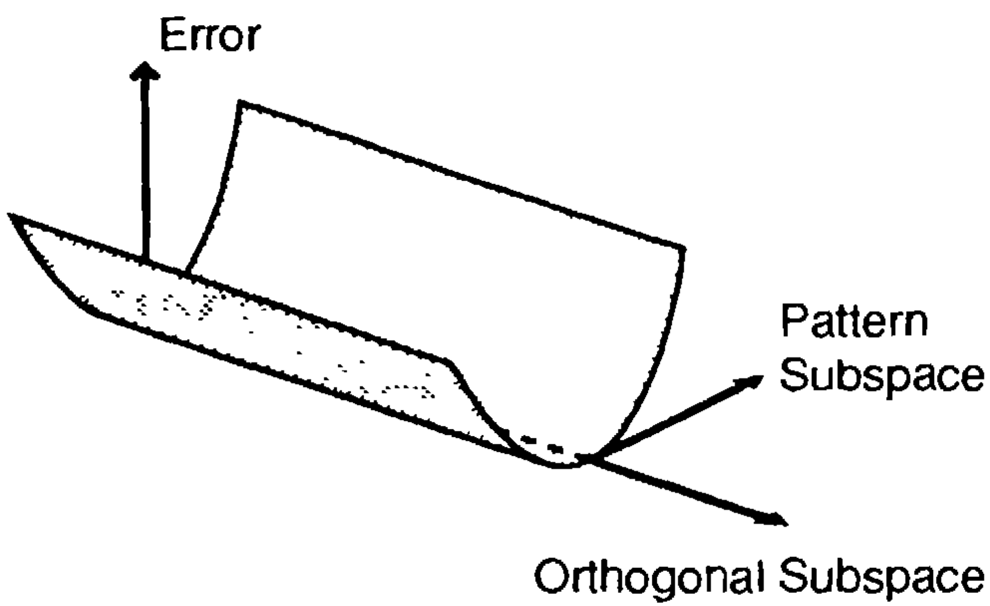
***E*(w)** *es forma cuadrática → único mínimo*

Si los **{ ξ }** del conjunto de entrenamiento son l.i. (=hay solución)

→ mínimo en $E = 0$

Toda dirección ξ ortogonal a { ξ } (patrones del conjunto de aprendizaje)

→ E(w) constante



- Cualquier componente de los pesos ortogonal a los { ξ } no es cambiado por el aprendizaje

- Siempre se llega a una solución con η suficientemente pequeño

UNIDADES NO LINEALES

y continuas (diferenciables) en general

$$E[\mathbf{w}] = \frac{1}{2}\sum_{i\mu}(\zeta_i^\mu - O_i^\mu)^2 = \frac{1}{2}\sum_{i\mu}\left[\zeta_i^\mu - g\left(\sum_k w_{ik}\xi_k^\mu\right)\right]^2 .$$

$$\frac{\partial E}{\partial w_{ik}} = -\sum_\mu [\zeta_i^\mu - g(h_i^\mu)]g'(h_i^\mu)\xi_k^\mu$$

→ $\Delta w_{ik} = \eta\delta_i^\mu\xi_k^\mu$. con $\delta_i^\mu = [\zeta_i^\mu - O_i^\mu]g'(h_i^\mu)$ cambio debido al patrón μ

Para sigmoideas típicas : $g'(h)$ grande cuando $|h|$ pequeño
$g'(h) \to 0$ cuando $h \to +\infty$ o $-\infty$

→ *se cambian las conexiones que alimentan entradas con h_i pequeño (las "indecisas")*

Típicamente:

$$g(h) = f_\beta(h) = \frac{1}{1+\exp(-2\beta h)} \quad g'(h) = 2\beta g(1-g)$$

$$g(h) = \tanh \beta h \quad g'(h) = \beta(1-g^2)$$

→ *no hace falta recalcular $g'(h_i)$ cuando ya se ha calculado $O_i = g(h_i)$*

<u>*Condición de existencia de solución*</u>*:* **independencia lineal**

ya que g inversible → $g(h) = \zeta \quad \leftrightarrow \quad h = g\text{-}1(\zeta)$

<u>*Convergencia del descenso por gradiente*</u>: ahora puede haber **mínimos locales** (además del global en E = 0, si lo hay), si los valores objetivo caen fuera del rango de g.

UNA APLICACION: RECONOCIMIENTO DE PATRONES

```
•#•
#•#
###
#•#
#•#
```
(-1,-1,-1)

```
##•
#•#
##•
#•#
##•
```
(-1,-1, 1)

```
•##
#••
#••
#••
•##
```
(-1, 1,-1)

```
##•
#•#
#•#
#•#
##•
```
(-1, 1, 1)

```
###
#••
##•
#••
###
```
(1,-1,-1)

```
###
#••
##•
#••
#••
```
(1,-1, 1)

```
•##
#••
#•#
#•#
•##
```
(1, 1,-1)

```
#•#
#•#
###
#•#
#•#
```
(1, 1, 1)