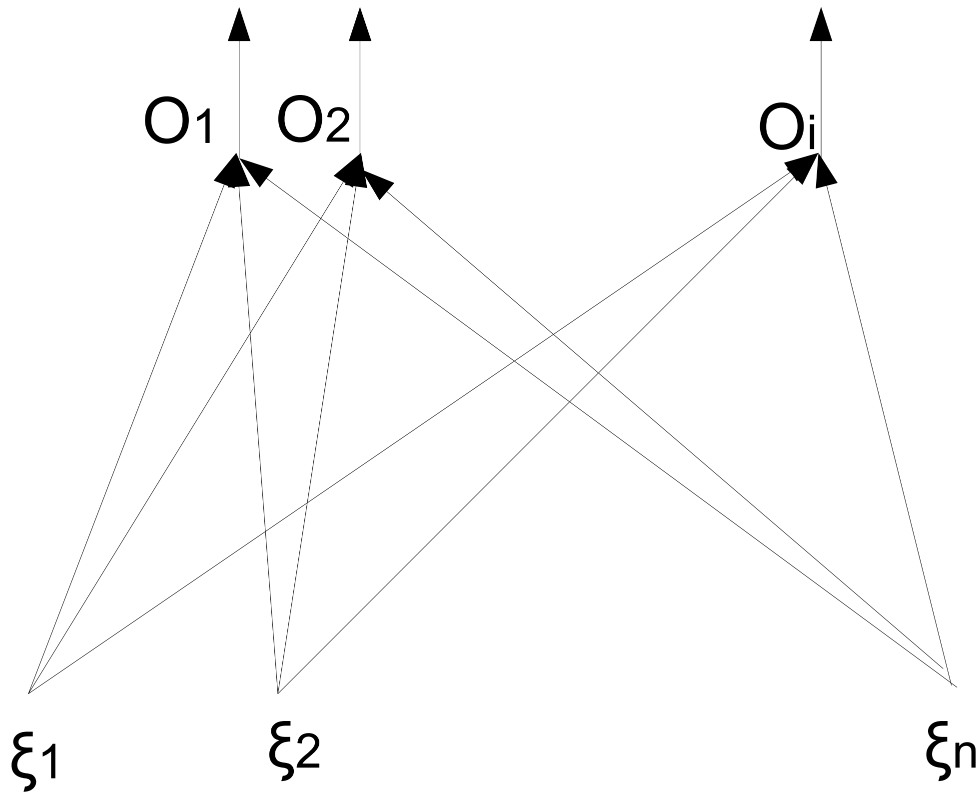
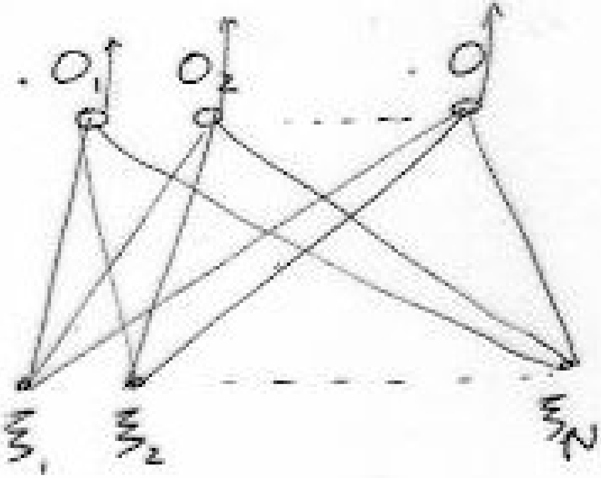
PERCEPTRON SIMPLE



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

Perceptrón Simple



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

Convención: los umbrales equivalen a una neurona $\xi_0 = -1$ y $w_{i0} = \theta_i$

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

Queremos: $\boxed{O_i^\mu = \zeta_i^\mu \qquad \mu = 1, \dots, p}$ i.e. $O_i^\mu = g(h_i^\mu) = g\left(\sum_k w_{ik}\xi_k^\mu\right)$

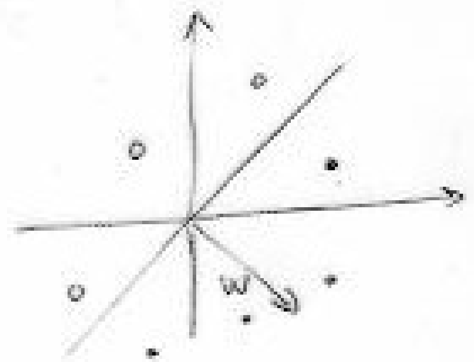
ξ, O, ζ booleanos o continuos:

- Unidades umbral o escalón: $g(h) = \mathrm{sgn}(h)$

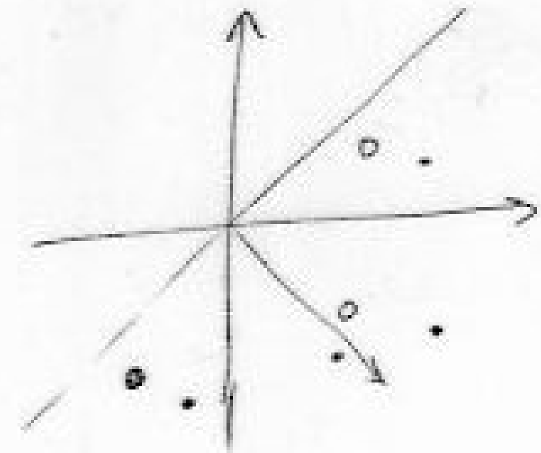
Sup. $\zeta_i^\mu = \pm 1 \longrightarrow$ buscamos: $\mathrm{sgn}(w \cdot \xi^\mu) = \zeta^\mu \ \forall \mu$

$\equiv$ proyecciones de ξ^μ sobre w coincidan en signo con ζ^μ

$\Longrightarrow$ la condición $w \cdot \xi = 0$ separa ambas clases (plano al origen, $\perp w$)



C.V.: $x^\mu \Leftarrow \xi^\mu \zeta^\mu$ $\longrightarrow$



(la clase negativa se rota sobre todos los ejes)

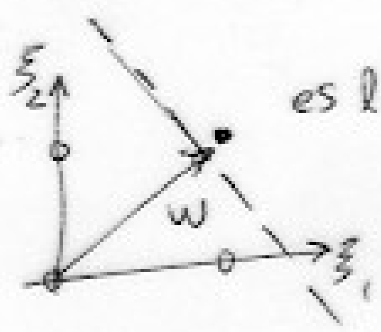
$$\boxed{\text{El problema será resoluble si y sólo si es linealmente separable: existe un plano en el espacio de } \vec{\xi} \text{ que separe los patrones}}$$

Reescrito con umbral: $O = \operatorname{sgn}(w \cdot \xi - w_0) \Longrightarrow$

el plano N-1 dim. que separa es $W \cdot \xi = w_0 \Longrightarrow w_0$ es la o.o. del plano

<u>Ejemplos:</u> AND

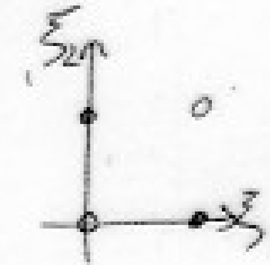
ξ_1	ξ_2	ζ
-1	-1	-1
-1	1	-1
1	-1	-1
1	1	1



es l.i.

XOR

ξ_1	ξ_2	ζ
-1	-1	-1
-1	1	1
1	-1	1
1	1	-1



no l.i. (y puede verse algebraicamente)

Algoritmo de Aprendizaje:

$$w_{ik}(n+1) = w_{ik}(n) + \Delta_{ik}(n) \quad \text{con} \quad \Delta_{ik}(n) = \eta\left(\zeta_i^\mu - O_i^\mu(n)\right)\xi_k^\mu$$

o, más general:

$$\boxed{\Delta w_{ik} = \eta\,\theta\left(N\kappa - \zeta_i^\mu h_i^\mu\right)\zeta_i^\mu\,\xi_k^\mu} \Longrightarrow$$

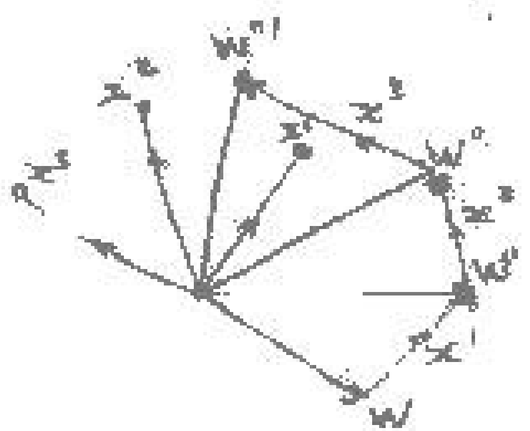
condición de cambio $\zeta_i^\mu h_i^\mu \le N\kappa$ en lugar de $\zeta_i^\mu h_i^\mu < 0$

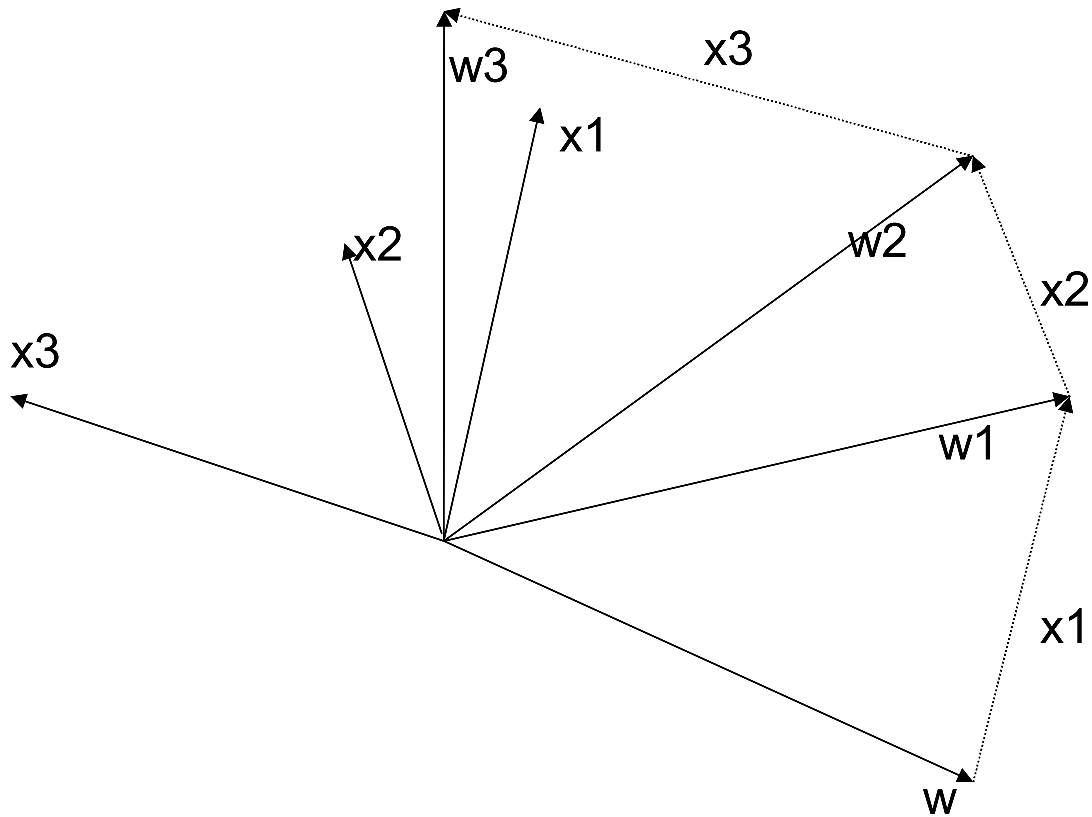
Con una sola salida:

$$w \cdot x^\mu > N\kappa \quad \text{y} \quad \boxed{\Delta w = \eta\,\theta\left(N\kappa - w\cdot x^\mu\right)x^\mu}$$

($\rightarrow$ todos los x^μ a más de $N\kappa/|w|$ del plano $\perp w$ q' separa

$\longrightarrow$ w crece en dirección de x^μ $\Longleftrightarrow$ su proyección sobre x^μ es $< N\kappa$





x3
w3
x1
x2
w2
x2
x3
w1
x1
w

Dificultad de un problema (Hertz)

$$D_{max} \triangleq \max_{w} D(w) \quad \text{con} \quad D(w) = \frac{1}{|w|} \min_{\mu} w \cdot x^{\mu}$$

Interpretación: qué tan lejos está, para un w, el x^{μ} más alejado del plano $\perp w$, maximizado sobre todos los w.

- D_{max} grande $\longrightarrow$ problema fácil
- $D_{max} < 0$ $\longrightarrow$ problema sin solución

e.g. $D_{max}^{(AND)} = \frac{1}{\sqrt{17}}$

$D_{max}^{(XOR)} = -\frac{1}{\sqrt{3}}$

> TEOREMA DE CONVERGENCIA (Rosenblatt)
>
> Para un problema linealmente separable, la solución mediante la regla de aprendizaje se obtiene en un número finito de pasos.

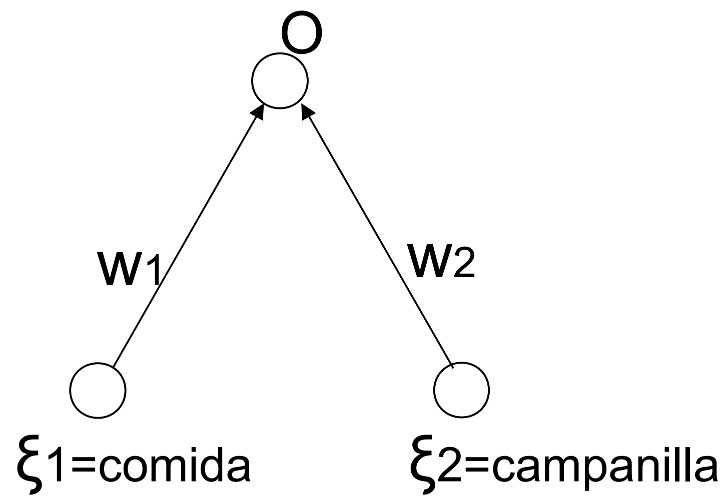
Perceptrones y regla de Hebb

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, *The organisation of behaviour*, 1949

Es el *Postulado de Hebb*

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



Si bien w2 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:

$$\boxed{E(w) = \frac{1}{2} \| \xi - O \|^2 = \frac{1}{2} \sum_{\mu} \| \xi^{\mu} - O^{\mu} \|^2}$$

Descenso por gradiente:

$$\Delta W = -\eta \nabla E \longrightarrow \Delta w_{ik} = -\eta \frac{\partial E}{\partial w_{ik}} = \eta \sum_{\mu} (\xi_i^{\mu} - O_i^{\mu}) \xi_k^{\mu} \triangleq \delta_i^{\mu} \xi_k^{\mu}$$

δ_i^{μ} error en la salida i cuando se presenta la entrada μ.

→ regla delta (nuevamente, idem unidades escalón)

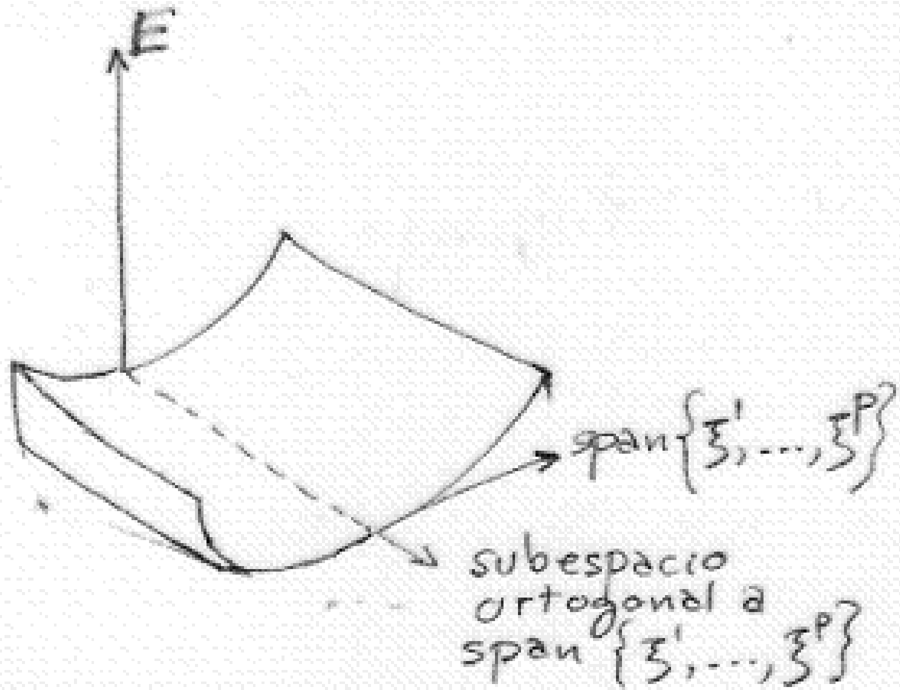
→ coinciden el enfoque <u>hebbiano</u> y el del <u>gradiente</u>

$E(w)$ es forma cuadrática $\longrightarrow$ único mínimo

Si $\{\xi^\mu\}$ l.i. (= hay solución) $\longrightarrow$ mínimo en $E=0$

$$\boxed{\text{Toda dirección } \vec{\xi}^* \perp \vec{\xi}^\mu \ \forall \mu \implies E(w) \text{ constante}}$$

$$E(w + \underbrace{\vec{a}^T}_{\text{vector columna}} \vec{\xi}^*) = \frac{1}{2}\sum \|\xi^\mu - O^\mu\|^2 = \frac{1}{2}\sum_{\mu i}\Big(\xi_i^\mu - \sum_k (w_{ik} + \cancel{a_i \xi_k^*})\xi_k^\mu\Big)^2 = E(w)$$



- Cualquier componente de los pesos ortogonal a los ξ^μ no es cambiado por el aprendizaje

- Siempre se llega a una solución con η suficientemente pequeño

UNIDADES NO LINEALES

y continuas (diferenciables) en general

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

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

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

Para sigmoideas típicas $\left[\begin{array}{l} g'(h) \text{ grande cuando } \lvert h \rvert \text{ pequeño} \\ g'(h) \rightarrow 0 \text{ cuando } h \rightarrow \pm\infty \end{array} \right.$

$\rightarrow$ se cambian las conexiones que alimentan entradas con $\lvert h_i^\mu \rvert$ pequeño (las "indecisas")

Típicamente:

- $g(h) = \tanh(\beta h) = \dfrac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} \in (-1, 1) \rightarrow g'(h) = \beta(1 - g^2)$
- $g(h) = f_\beta(h) = \dfrac{1}{1 + e^{-2\beta h}} \in (0, 1)$ [logística] $\rightarrow g'(h) = 2\beta g(1 - g)$

$\rightarrow$ no hace falta recalcular $g'(h_i^\mu)$ cuando ya se ha calculado $O_i^\mu = g(h_i^\mu)$

Condición de existencia de solución: independencia lineal
(g inversible $\Longrightarrow$ $g(h) = 5 \Longleftrightarrow h = g^{-1}(5)$)

Convergencia del descenso por gradiente: ahora puede haber mínimos locales (además del global en $E=0$)
si los valores objetivo caen fuera del rango de g