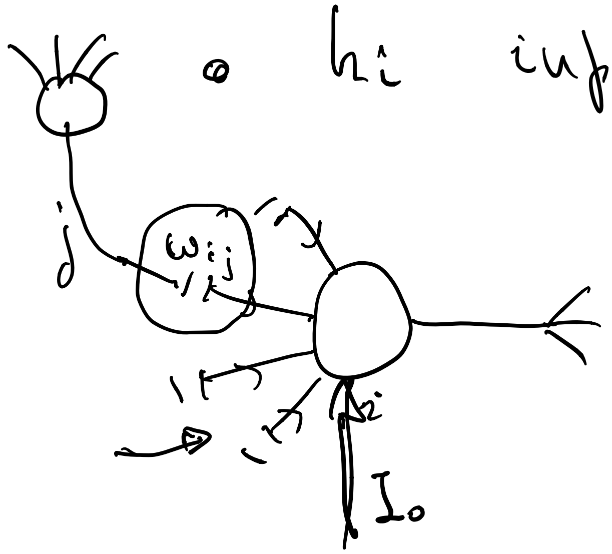
Neurona Artificial

McCulloch and Pitts 1943

- vamos a describir la neurona i
- y_i modela la señal (output)
- h_i input



$$h_i(t) = \sum_{j=1}^{N} w_{ij}\, y_j(t)$$

- puede haber un término externo

la neurona tiene un estímulo

$$h_i(t) + I_0$$

- tiene un umbral de actuación

$$\mu_i$$

- Función de activación

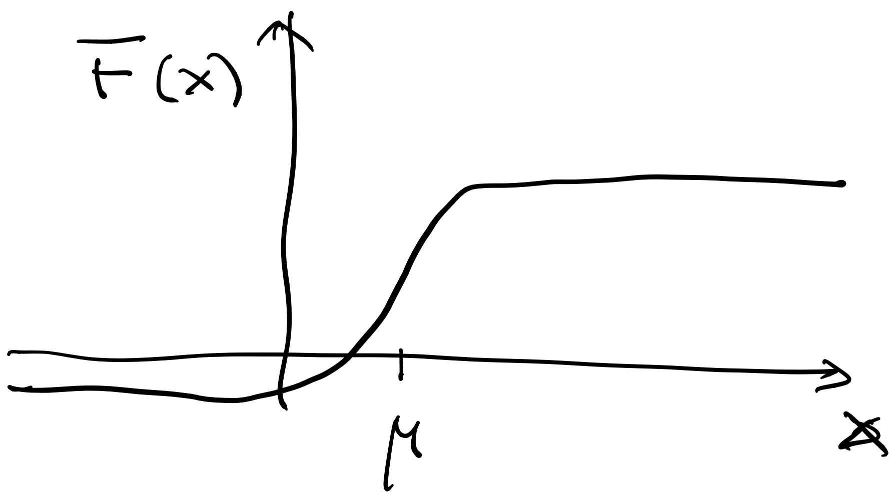
$$y_i(t+\Delta t) = F\left(h_i(t) + I_0 - \mu_i\right)$$

Asumimos que $I_0 = 0$

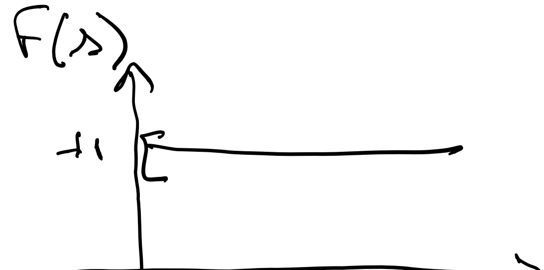
$$y_i(t+\Delta t) = F\left(h_i(t) - \mu_i\right)$$

$$F : h_i \rightarrow y_i$$

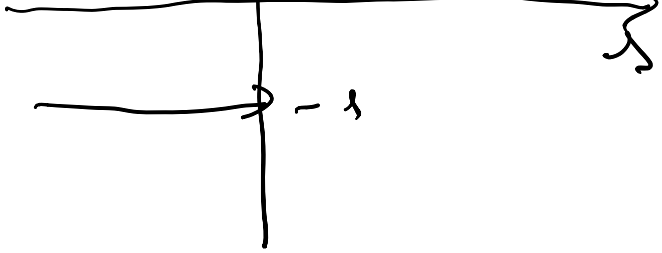
Típicamente



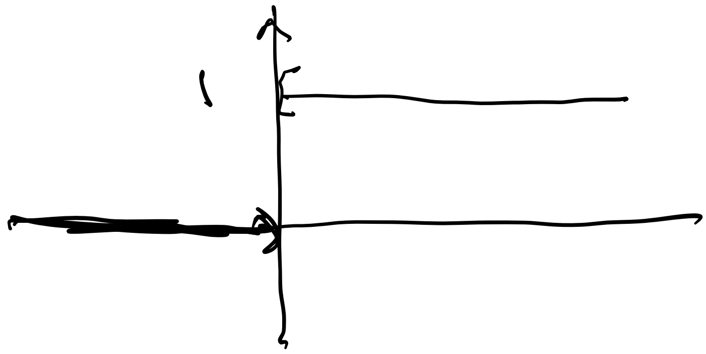
Ejemplos



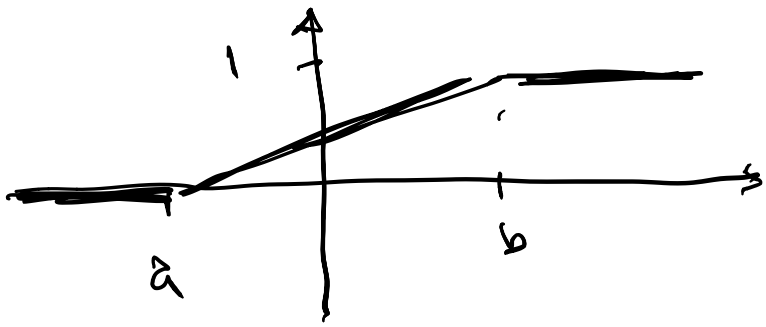
$$F(s) = \text{signo}(s) = \begin{cases} +1 & s \geq 0 \end{cases}$$



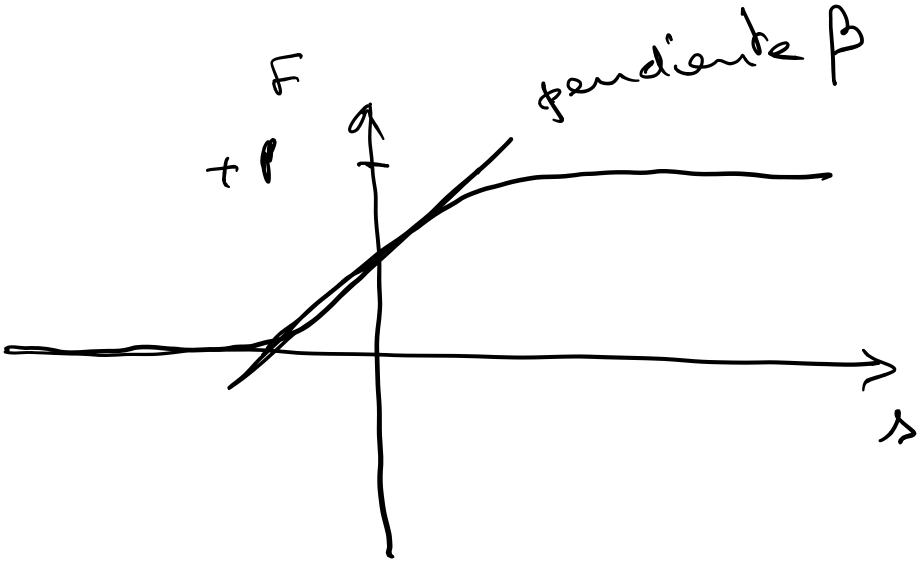
$$F(s) = \mathrm{signo}(s) = \begin{cases} \dots \\ -1 & s<0 \end{cases}$$



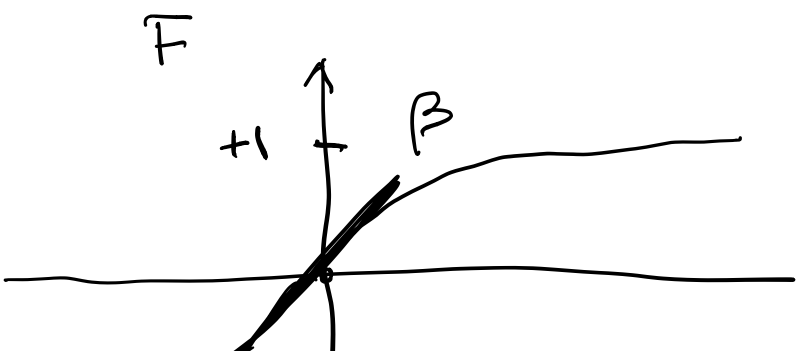
$$F(s) = H(s) = \begin{cases} +1 & s \geq 0 \\ 0 & s < 0 \end{cases}$$



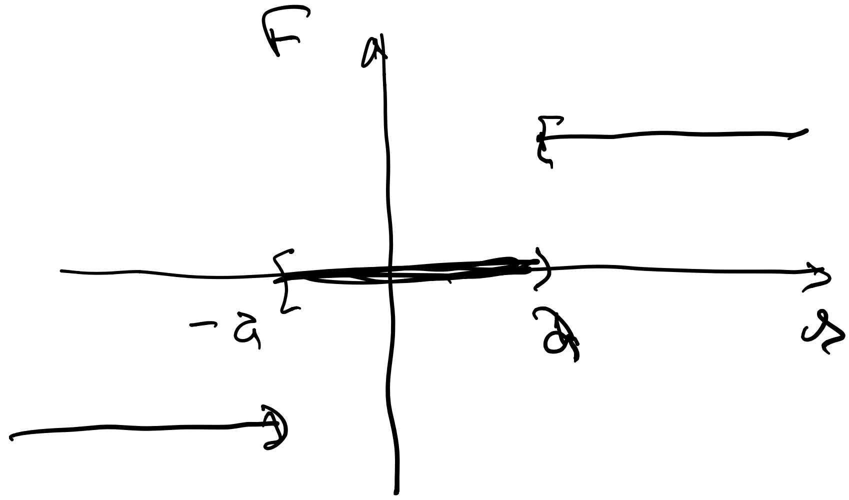
$$F(s) = \begin{cases} +1 & s > b \\ \text{recta} & a<s<b \\ 0 & s < a \end{cases}$$



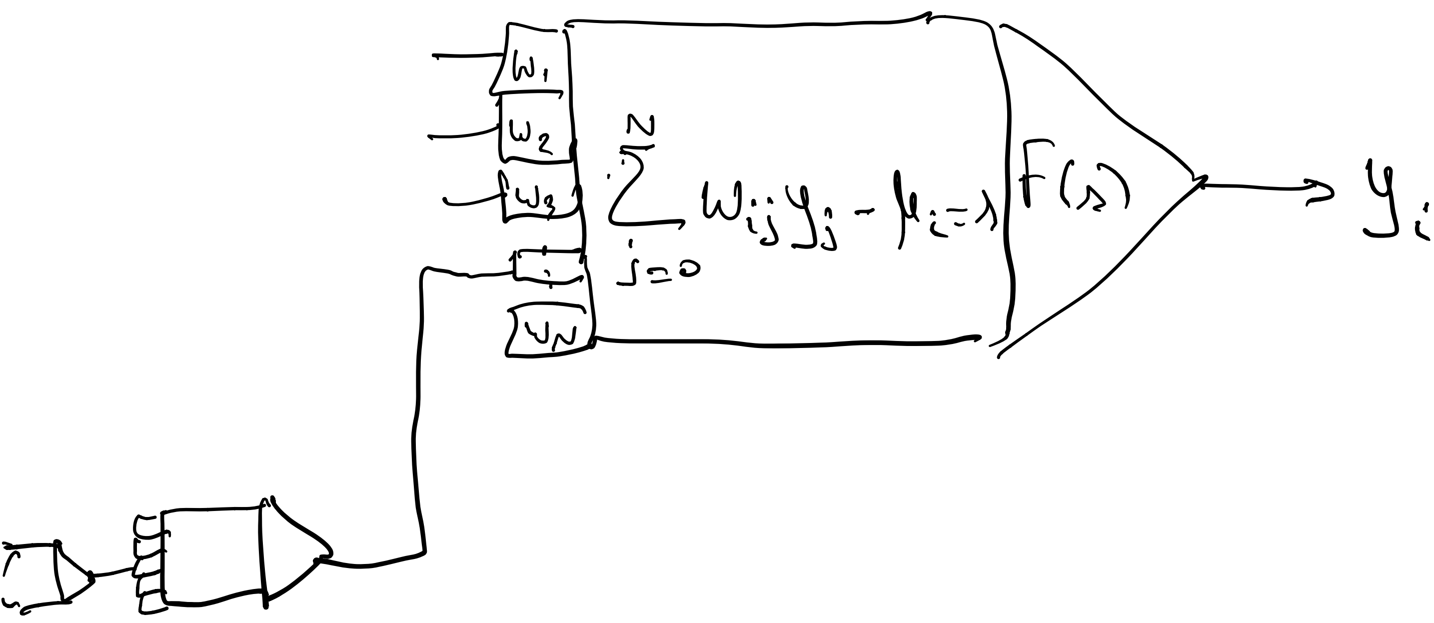
$$F(s) = \frac{1}{1+e^{-\beta s}}$$



$$F(s) = \tanh(\beta s)$$



$$F_s(s) = \begin{cases} +1 & s > a \\ 0 & -a < s < a \\ -1 & s < a \end{cases}$$



Las sinapsis son nº reales

$W_{ij} < 0$ sinapsis inhibitoria

$W_{ij} > 0$ " exhitatoria

$W_{ij} = 0$ " ausente

$|W_{ij}|$ es una medida de la eficacia

W_{ij} sinapsis entre j e i

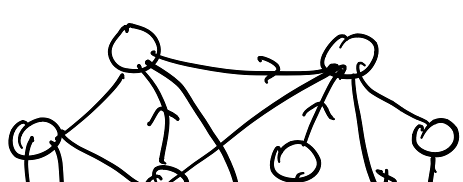
$\uparrow$

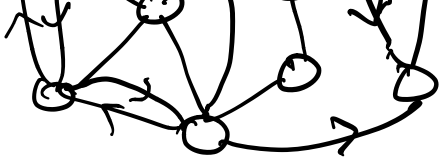
pre

El tiempo es discreto, la red se actualiza a cada Δt seg.

Arquitecturas

Redes recurrentes o atractoras

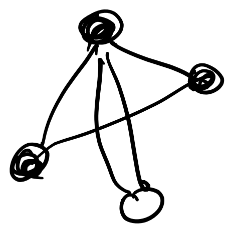




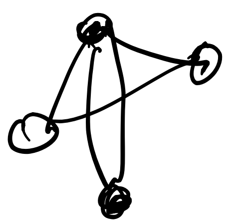
El input es el estado inicial

$$\{y_i(t=0)\}_{i=1}^{N}$$

$$y_i = \begin{cases} +1 & \bullet \\ -1 & \circ \end{cases}$$



$t=0$



$t=1$

$t = 2$

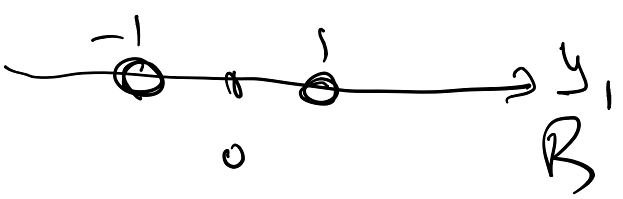
input $\{y_i(0)\}_{i=1}^{N} \to \{y_i(1)\}_{i=1}^{N} \to \{y_i(2)\}_{i=1}^{N}$

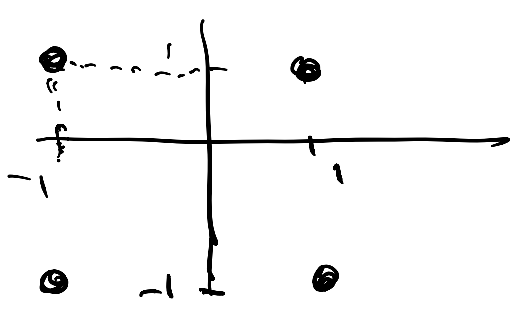
$\to \{y(t=\infty)\}_{i=1}^{N}$

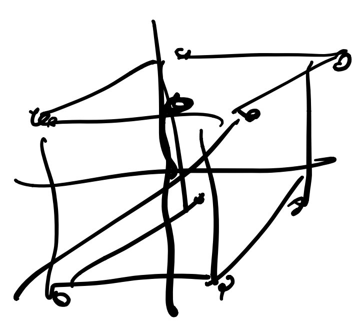
Si tengo N neuronas binarias ¿cuántos estados tengo?

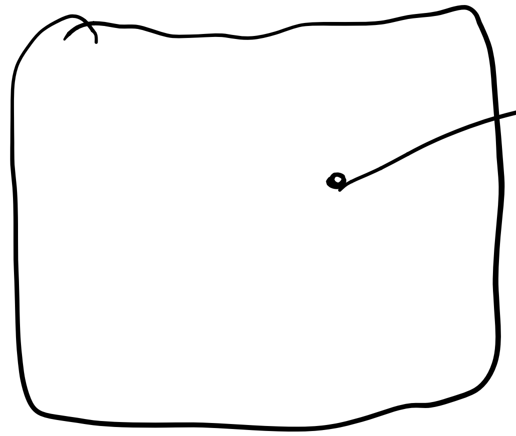
tengo $2^N = e^{\ln 2 N}$

$= e^{N(\ln(2))}$



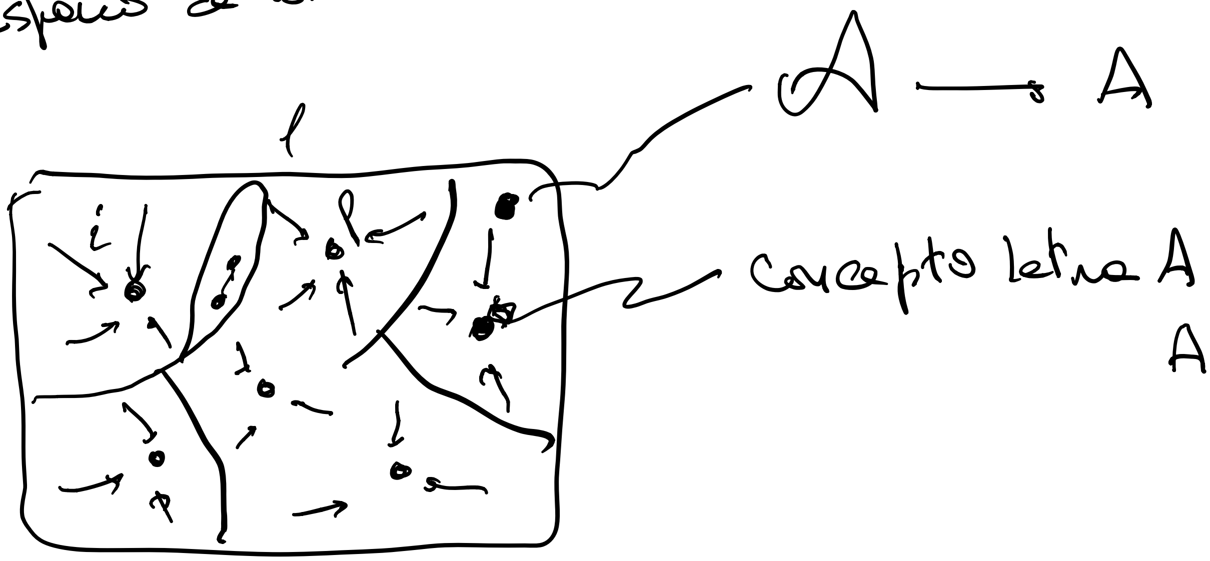






$\vec{J} = (+1, +1, -1, \ldots +1, \ldots)$

Espacio de los estados

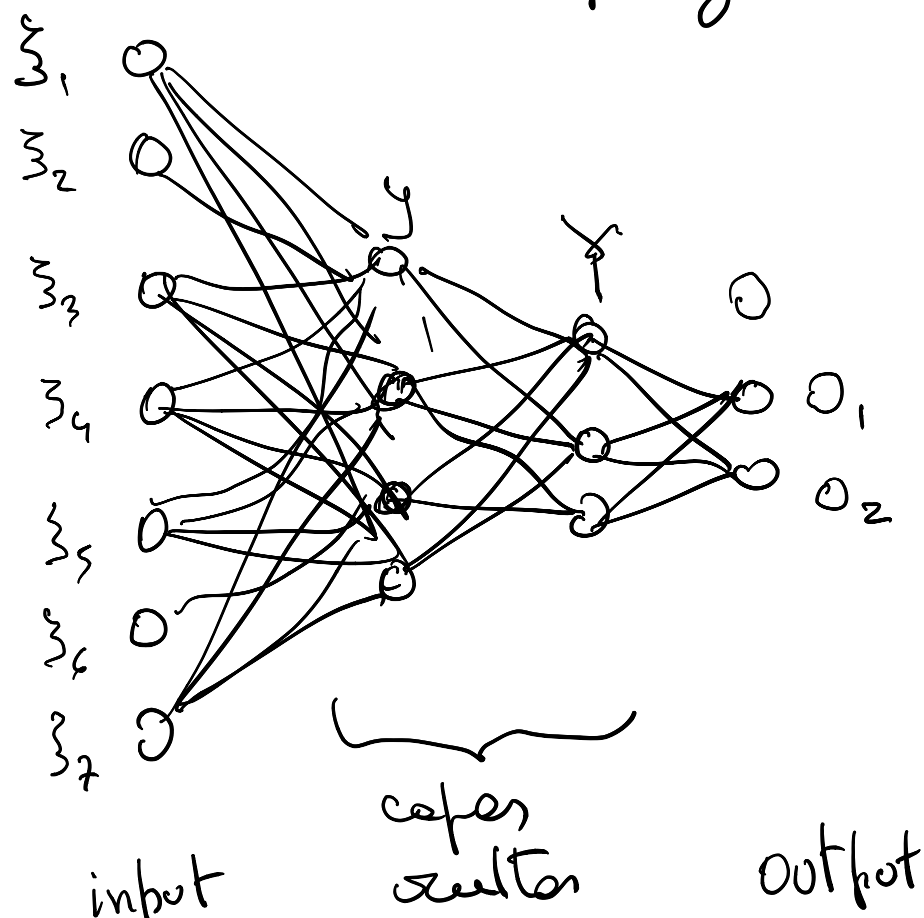


Buscamos una asociación

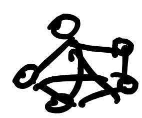
muchos a pocos
estados iniciales atractores

Redes feed forward

Los neuronas se organizan en capas

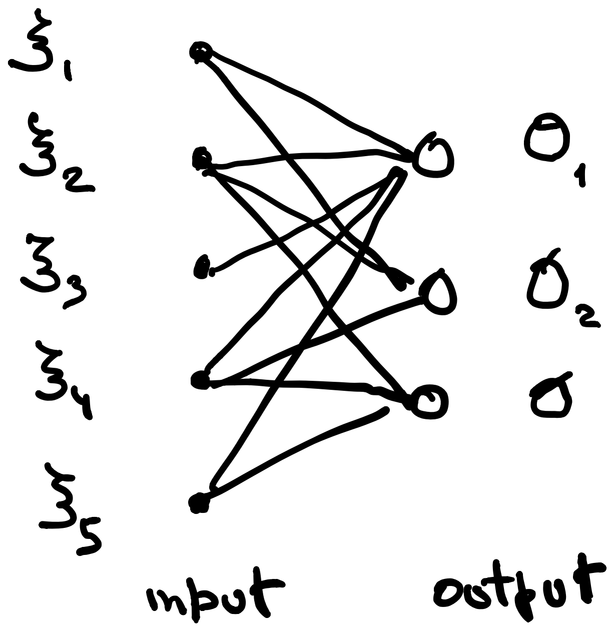



```
for t:=0 to tmax
    for i:=1 to N
        h:=0
        for k:=1 to P
            h:= h+ w(i,k)y(k)
        end
        y(i)= F(h-μ)
    end
end
```



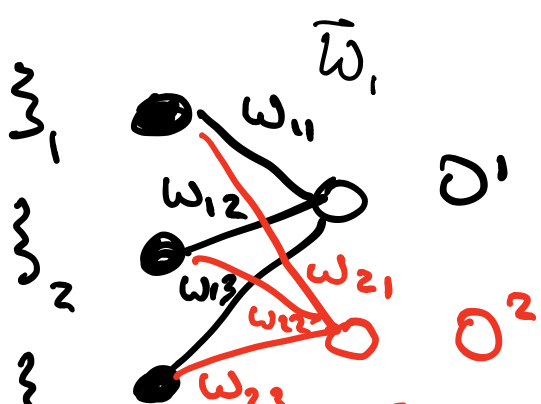
Perceptron simple

tiene una capa de entrada y otra de salida



Puedo desacoplar la red en tres redes

En general, si la capa de salida tiene M neuronas, puedo desacoplar el problema en M perceptrons simples



ξ_3 $\quad$ w_2

Entonces pensemos por ahora en una única neurona de salida por vez, y la llamamos i

Calculamos $h_i = \sum_{k=1}^{N} w_{ik} \xi_k$

Comparamos h_i con el umbral μ_i

Aplicamos la función $g(h_i - \mu_i)$.

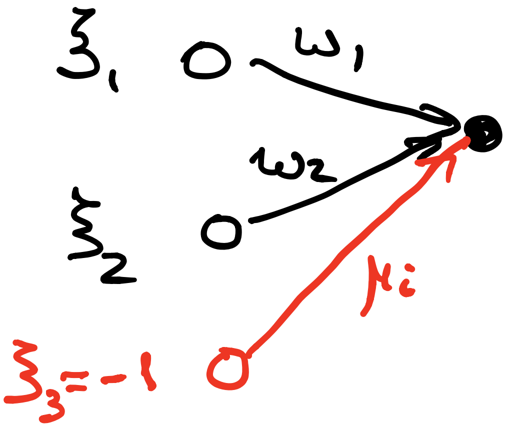
$$O_i = g(h_i - \mu_i)$$

Ahora me libero de μ_i

Agregamos una neurona $N+1$ en la entrada que siempre vale -1

$$\xi_{N+1} = -1 \quad \text{fija}$$

Defino $w_{i,N+1} = \mu_i$



$$O_i = g\left(\sum_{k=1}^{N+1} w_{ik}\,\xi_k\right)$$

Supongamos que conocemos p ejemplos correctos (relaciones input-output)

$$\vec{\xi} = (\xi_1, \xi_2, \ \ , \xi_{N+1})$$

$$\vec{\xi}^{\mu} \longrightarrow d_i^{\mu}$$

μ etiqueta el conjunto de entrenamiento $\mu = 1, 2, \ldots, p$

Si funciona bien

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

Vamos a tratar de encontrar un conjunto $w_{i1}, w_{i2}, \ldots w_{iN}$ tal que

$$g\left(\sum_k w_{ik} \xi_k^\mu\right) = d_i^\mu \qquad \forall \mu$$

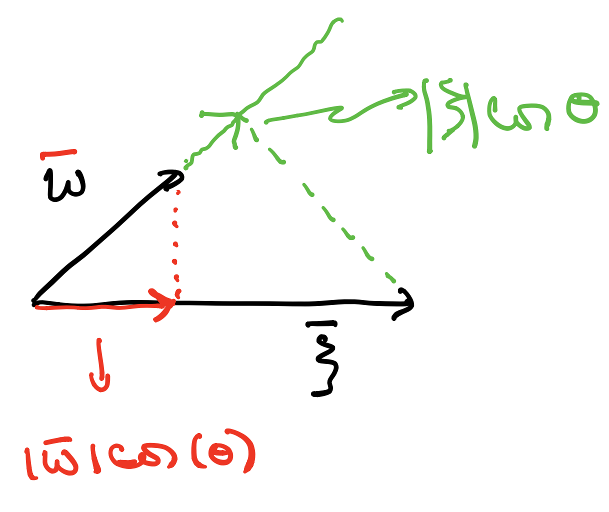
Notación $\quad \overline{w}_i = (w_{i1}, w_{i2}, \ldots, w_{i,N})$

$$\vec{\xi} = (\xi_1, \xi_2, \ldots, \xi_N)$$

$$\vec{\xi}^\mu = (\xi_1^\mu, \xi_2^\mu, \ldots, \xi_N^\mu)$$

$$\sum_K w_{ik} \xi_k^\mu = \underset{\substack{\uparrow \\ \text{producto punto}}}{\overline{w}_i \bullet \overline{\xi}^\mu} = |\overline{w}| . |\overline{\xi}^\mu| \cos(\theta)$$

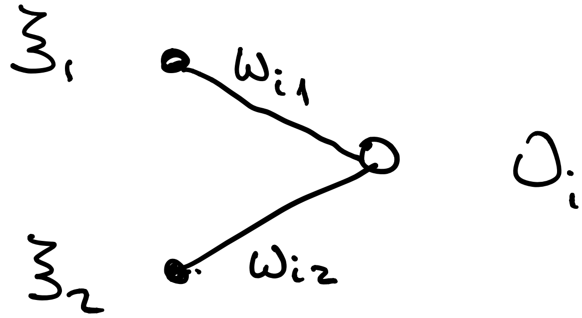
θ : ángulo que forman los vectores $\overline{w}_i$ y $\overline{\xi}^\mu$



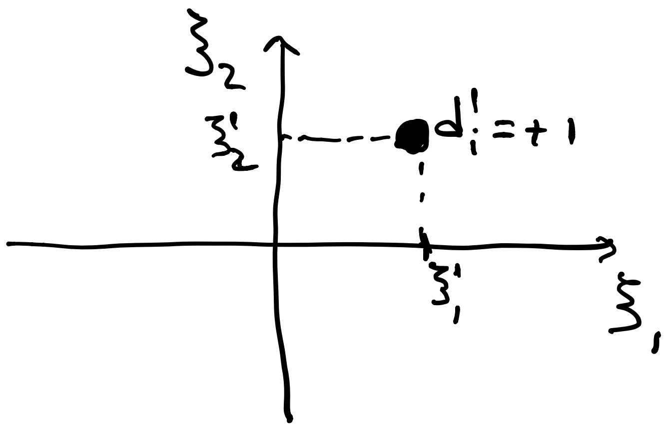
1º Caso : función signo (unidad de umbral)

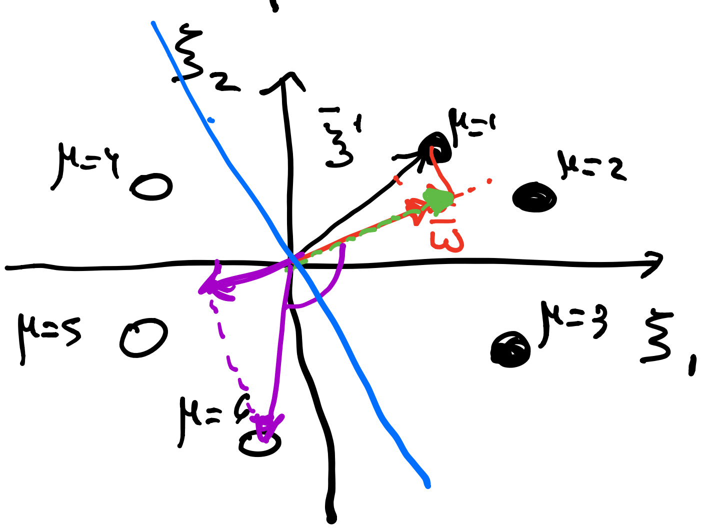
$$g(h)=\begin{cases}+1 & \text{si } h\geq 0\\ -1 & \text{si } h<0\end{cases}$$

A los efectos "gráficos" voy a suponer $N=2$.



Tengo p ejemplos correctos





● output $d_i^\mu=1$

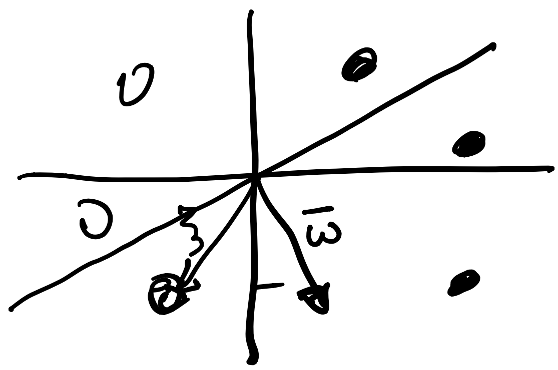
○ output $d_i^\mu=-1$

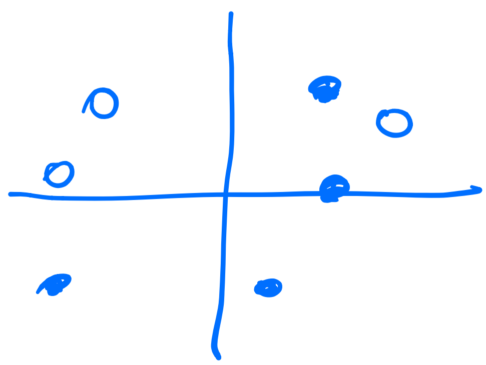
$P=6$

$$O_i = \text{signo}\left(w_{i1}\xi_1^\mu + w_{i2}\xi_2^\mu\right) = d_i^\mu$$

$$\text{signo}\left(\bar{w}\cdot\bar{\xi}\right) = d_i^\mu$$

$$\text{Si} \quad d_i^\mu = +1 \Rightarrow \bar{w}\cdot\bar{\xi} \geq 0$$





no $\exists$ $\bar{w}$

Si $\exists$ una recta (hiperplano) que separe los +1 • de los −1, entonces $\exists$ un vector $\bar{w}$ que hace funcionar bien el perceptron

Si no $\exists$ esa recta (hiperplano), no tiene solución el problema

Podemos definir $x_k^\mu = d_i^\mu \xi_k^\mu$

$$\vec{x}^\mu = d_i^\mu \bar{\xi}^\mu \qquad d_i^\mu = \pm 1$$

$\bar{x} \cdot \bar{y} \cdot a$

$\bar{x} \cdot (a\bar{y})$

$$\text{sig}\,(\bar{w} \cdot \bar{\xi}^\mu) = \pm 1 = d_i^\mu$$

$$\Downarrow$$

$$\text{signo}\,(\bar{w} \cdot \bar{\xi}^\mu)\, d_i^\mu = (d_i^\mu)^2 = 1$$

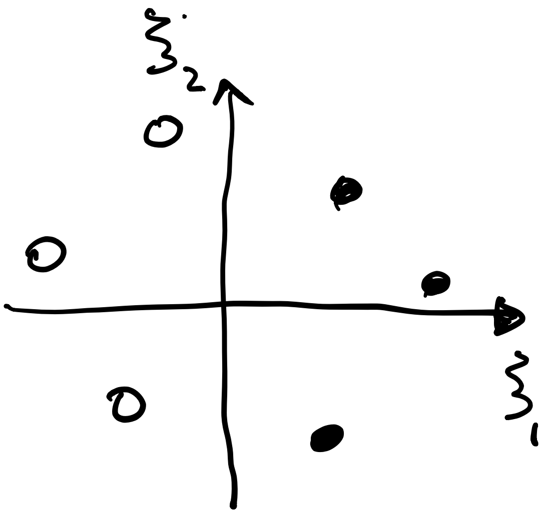
$$\text{signo}\big((\bar{w} \cdot \bar{\xi}^\mu) d_i^\mu\big) = 1$$

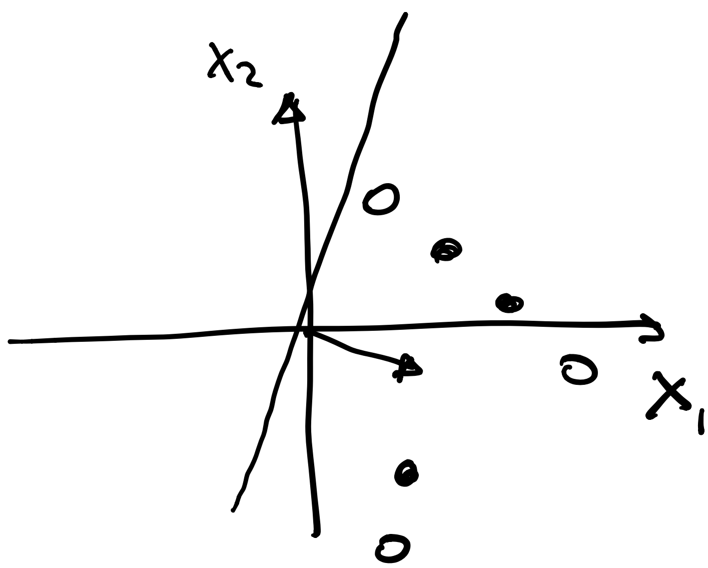
$$\text{signo}\big(\bar{w} \cdot (\bar{\xi}_i^\mu d_i^\mu)\big) = 1$$

$$\text{sigo}\,(\bar{w} \cdot \bar{x}^\mu) = 1 \quad \leftarrow$$

$$\Downarrow$$

$$\bar{w} \cdot \bar{x}^\mu > 0$$





Para que un problema de clasificación ± 1 se pueda resolver, debe ser "linealmente separable" (LS)

Si no es LS no significa que no exista una red que lo resuelva. Solo sabemos que no puede ser un perceptron simple.

"Volvemos a poner el Umbral"

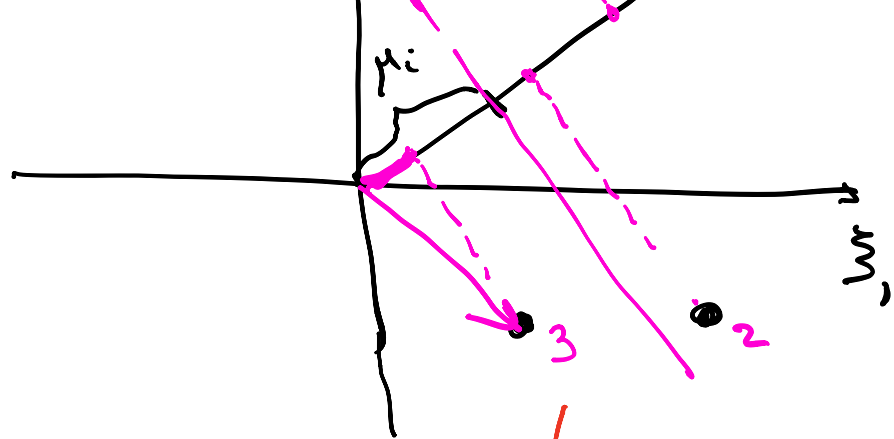
Si hay umbral

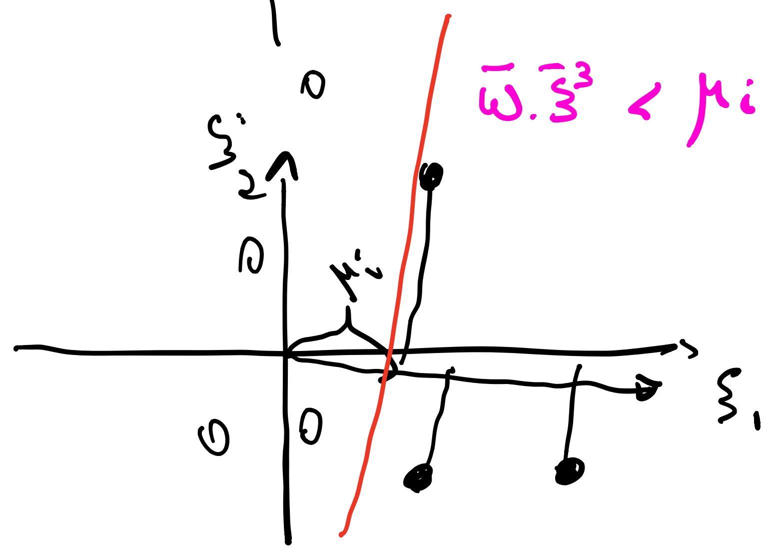
$$O_i = \text{signo}\left(\bar{w} \cdot \bar{\xi} - \mu_i\right)$$

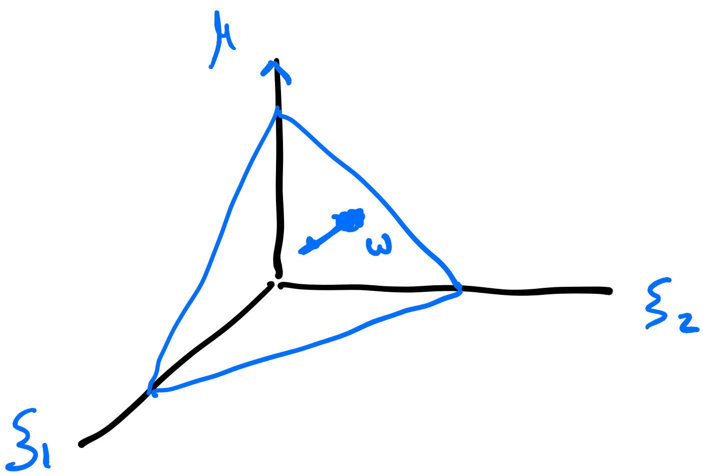
$\bar{w} \cdot \bar{\xi} = \mu_i$ define la separatriz

si $d_i^\mu = 1$ $\bullet$ $\quad \bar{w} \cdot \bar{\xi} > \mu_i$

$d_i^\mu = -1$ $\circ$ $\quad \bar{w} \cdot \bar{\xi} < \mu_i$

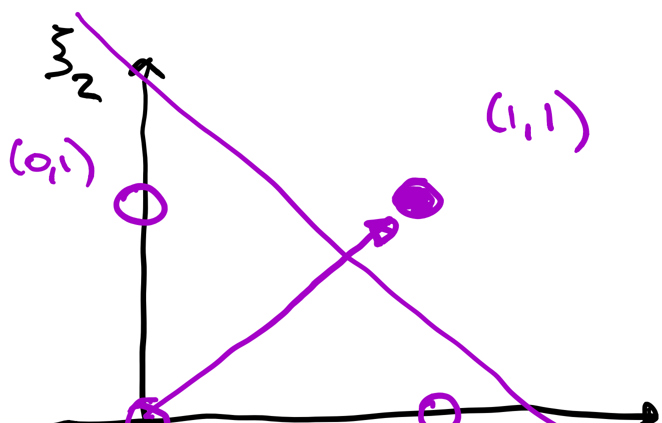


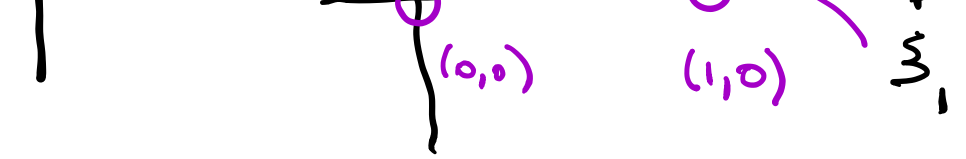




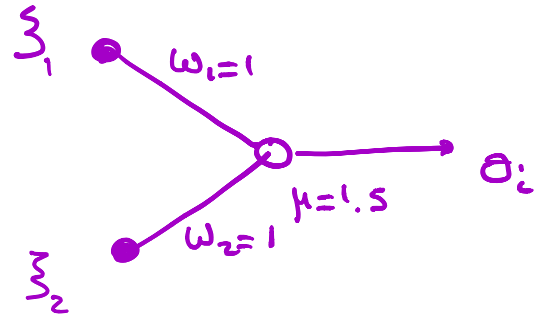
La función AND

	ξ_1	ξ_2	d
$\mu=1$	0	0	-1
$\mu=2$	0	1	-1
$\mu=3$	1	0	-1
$\mu=4$	1	1	1





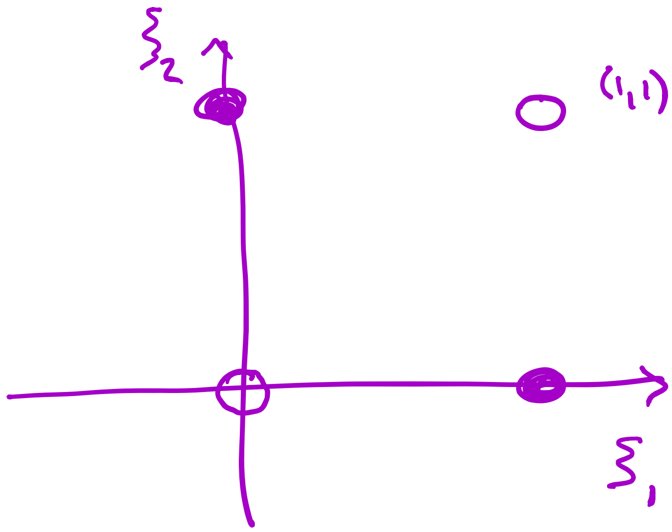
$\vec{W} = (1,1)$



¿Qué demostró Minsky?

XOR

ξ_1	ξ_2	d
1	1	-1
1	0	1
0	1	1
0	0	-1



no es LS

UN ALGORITMO SIMPLE

Supongamos que el problema es L.S.

Buscamos un algoritmo tal que

partiendo de un $\bar{w}^0$ tomado al azar, nos lleve a un $\bar{w}^*$ que resuelva el problema EN UN NÚMERO FINITO DE PASOS

$$\begin{array}{c} \bar{w}^{old} \longrightarrow \bar{w}^{new} \end{array}$$

$$\bar{w}_i^{new} = \bar{w}_i^{old} + \Delta \bar{w}_i$$

Secuencialmente le muestro cada ejemplo $\mu=1$, $\mu=2$, ... $\mu=P$.

η (eta)

Miremos $\mu=1$. Calculamos O_i^μ

Si $O_i^\mu = d_i^\mu$ no hago nada

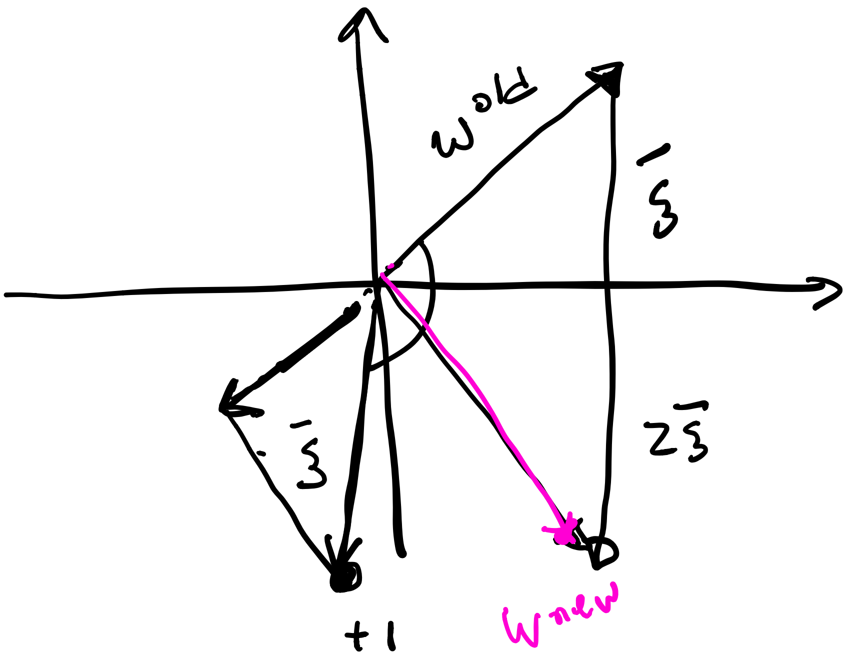
$O_i^\mu \neq d_i^\mu$ $\quad \Delta w_{ik} = 2\eta\, d_i^\mu \xi_k^\mu$

$$\Delta w_{ik} = \begin{cases} 2\eta\, d_i^k \xi_k^\mu & \text{si } O_i^\mu \neq d_i^\mu \\ 0 & \text{si } O_i^\mu = d_i^\mu \end{cases}$$

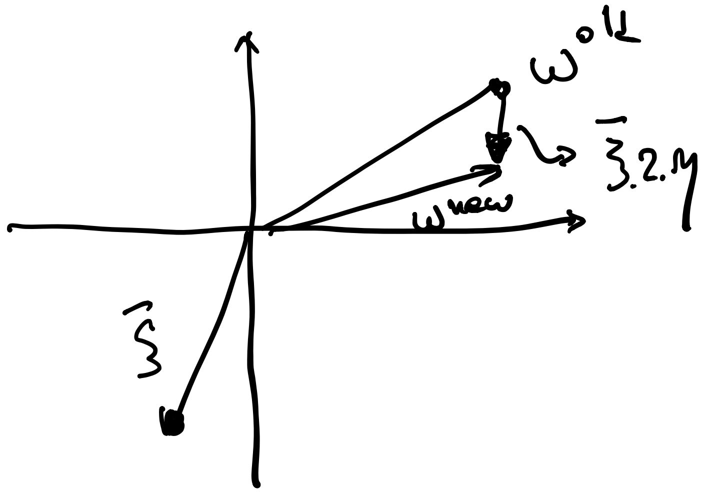
$$\Delta w_{ik} = \eta \left(1 - d_i^\mu O_i^\mu\right) d_i^\mu \xi_k^\mu$$

$$= \eta \left(d_i^\mu - (d_i^\mu)^2 O_i^\mu\right) \xi_k^\mu$$

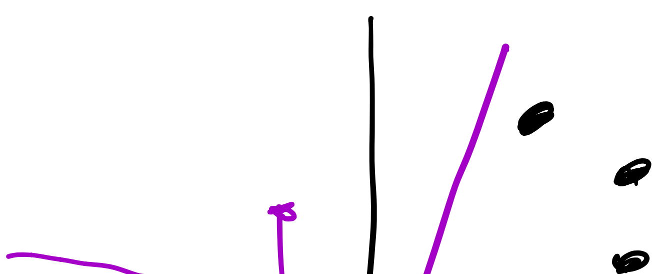
$$= \eta \left(d_i^\mu - O_i^\mu\right) \xi_k^\mu$$

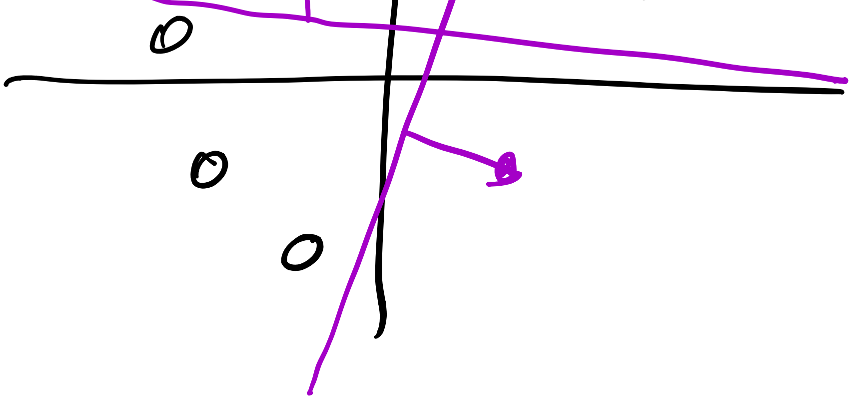


Si $\eta = 1$



Si $\eta = 0{,}01$





$$D(\bar{w}) = \min_{\mu} \bar{w}\,\bar{\xi}^{\mu} / |\bar{w}|$$

El mejor $\bar{w}$ es el que maximiza D

$$D_{max} = \max_{\bar{w}} D(\bar{w})$$