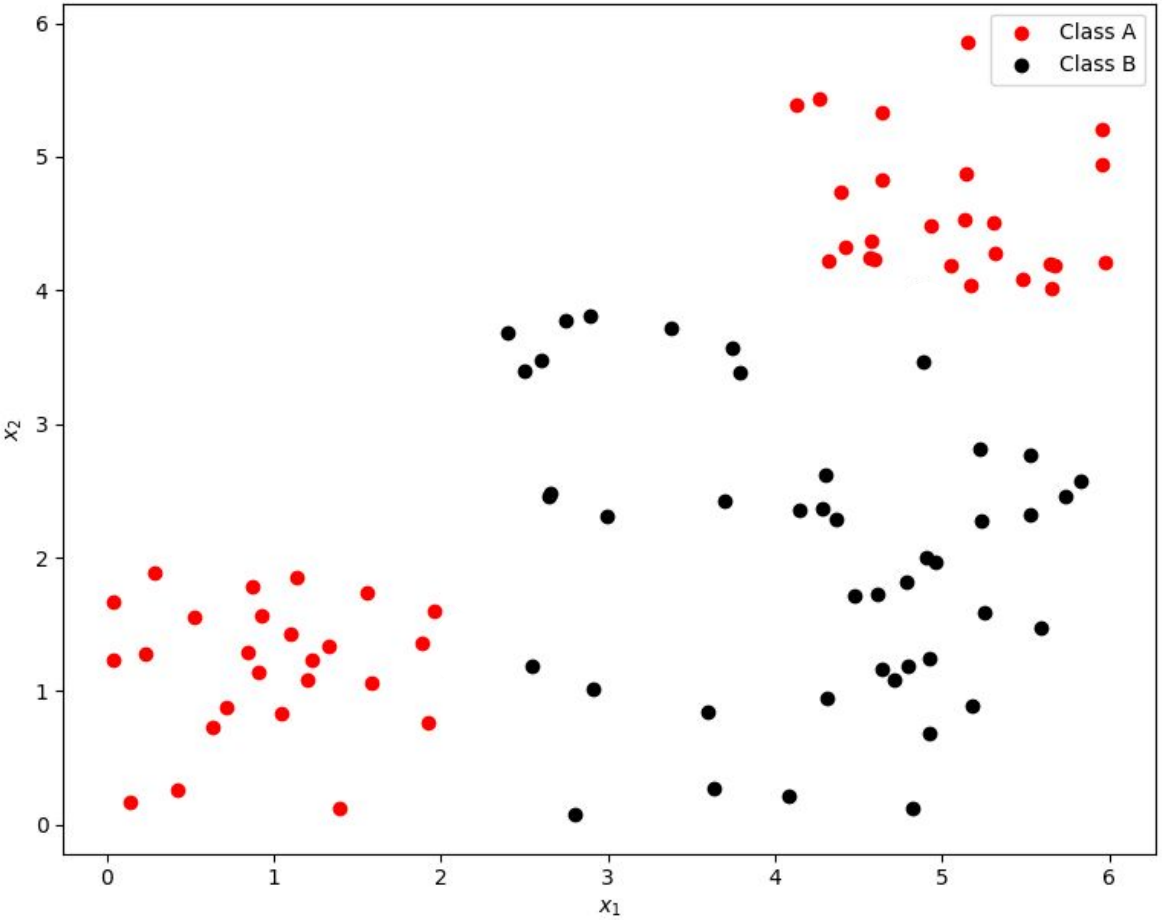
Perceptrón Multicapa

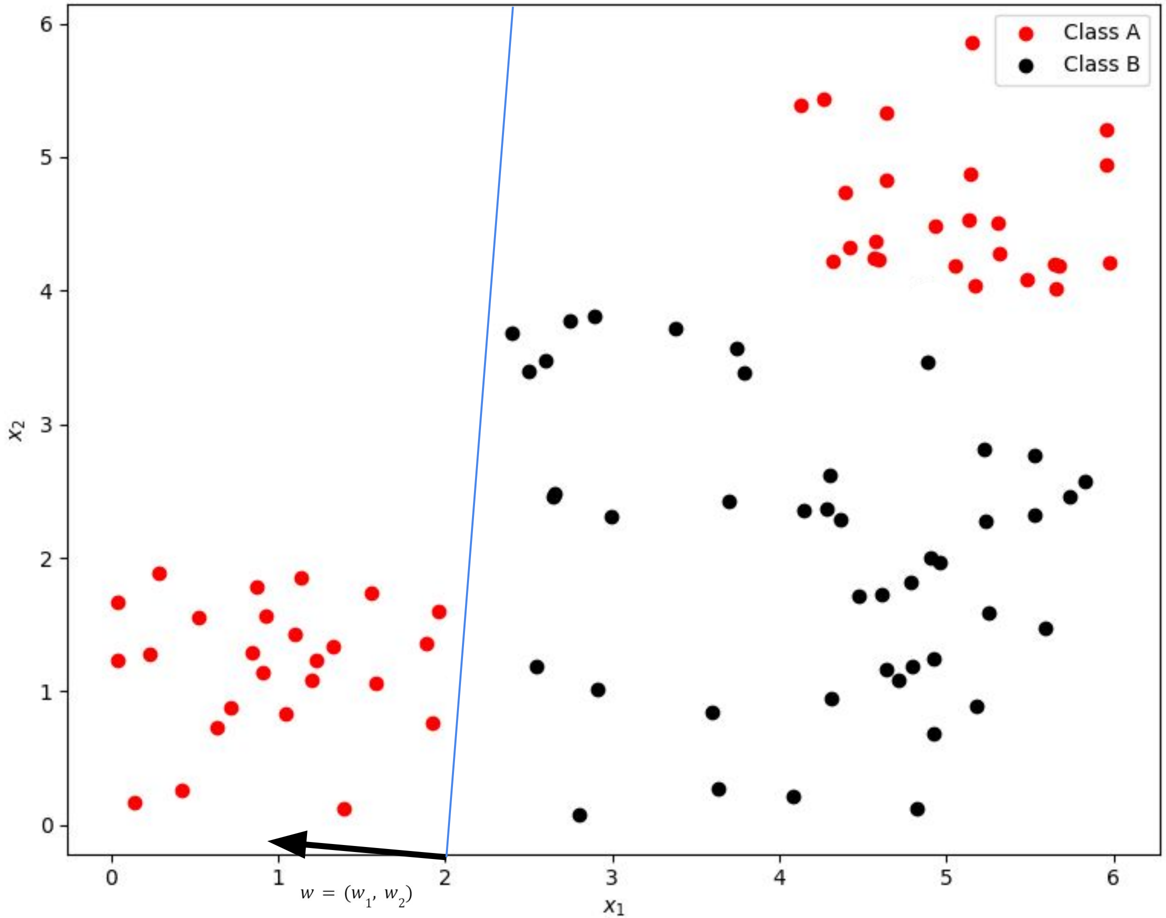
Sistemas de Inteligencia Artificial

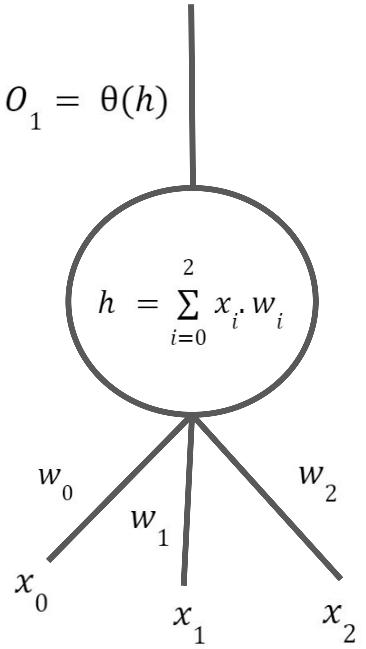
Primer Cuatrimestre 2023

Rodrigo Ramele
Eugenia Piñeiro
Alan Pierri
Santiago Reyes
Marina Fuster
Luciano Bianchi

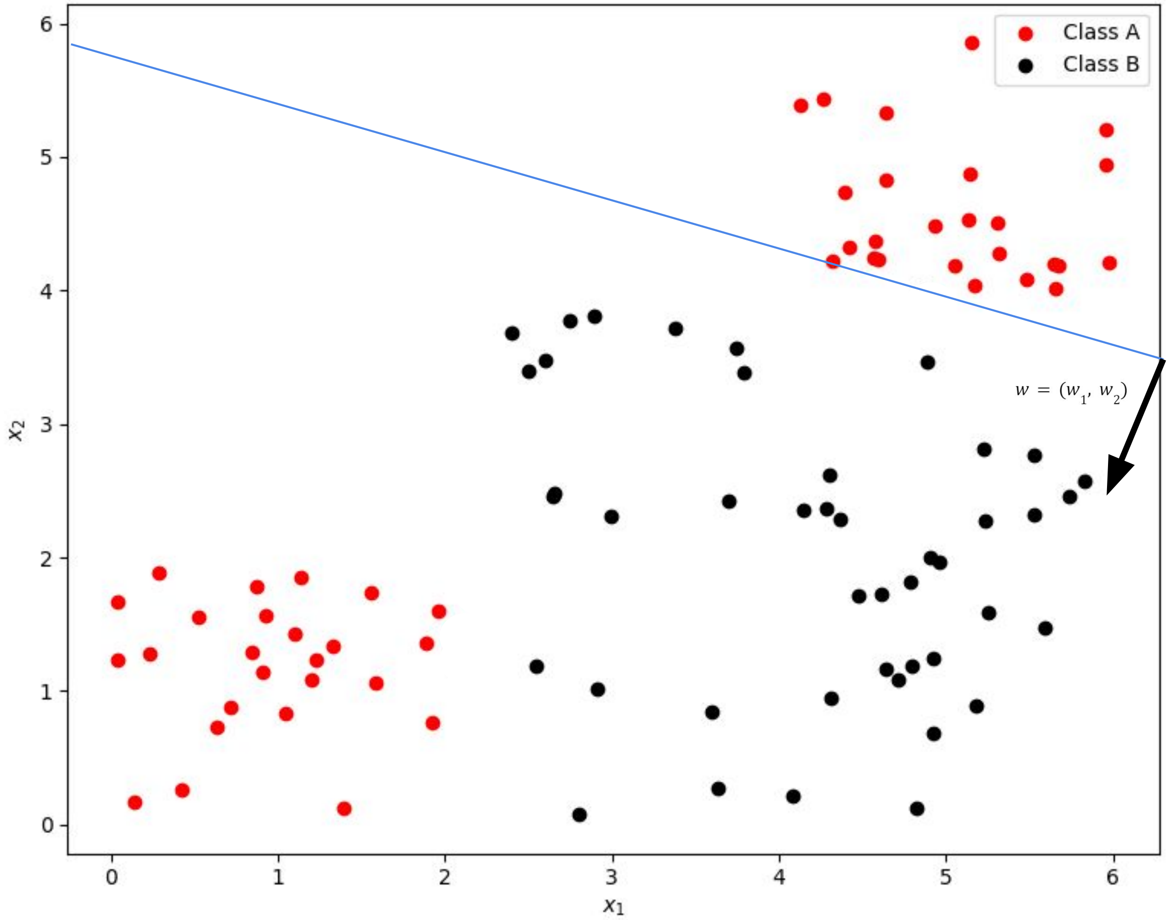

Class A
Class B
x_1
x_2
0
1
2
3
4
5
6
0
1
2
3
4
5
6

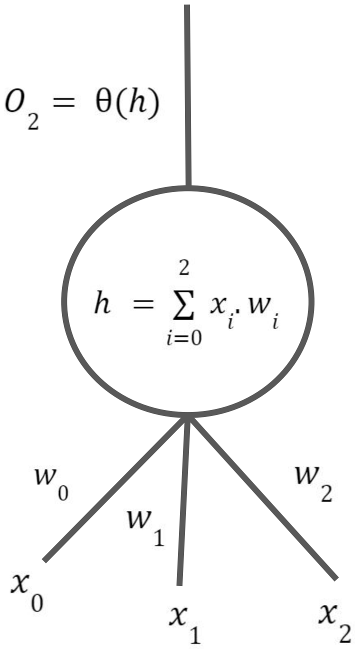
$\theta\ =\ función\ activación\ escalón$



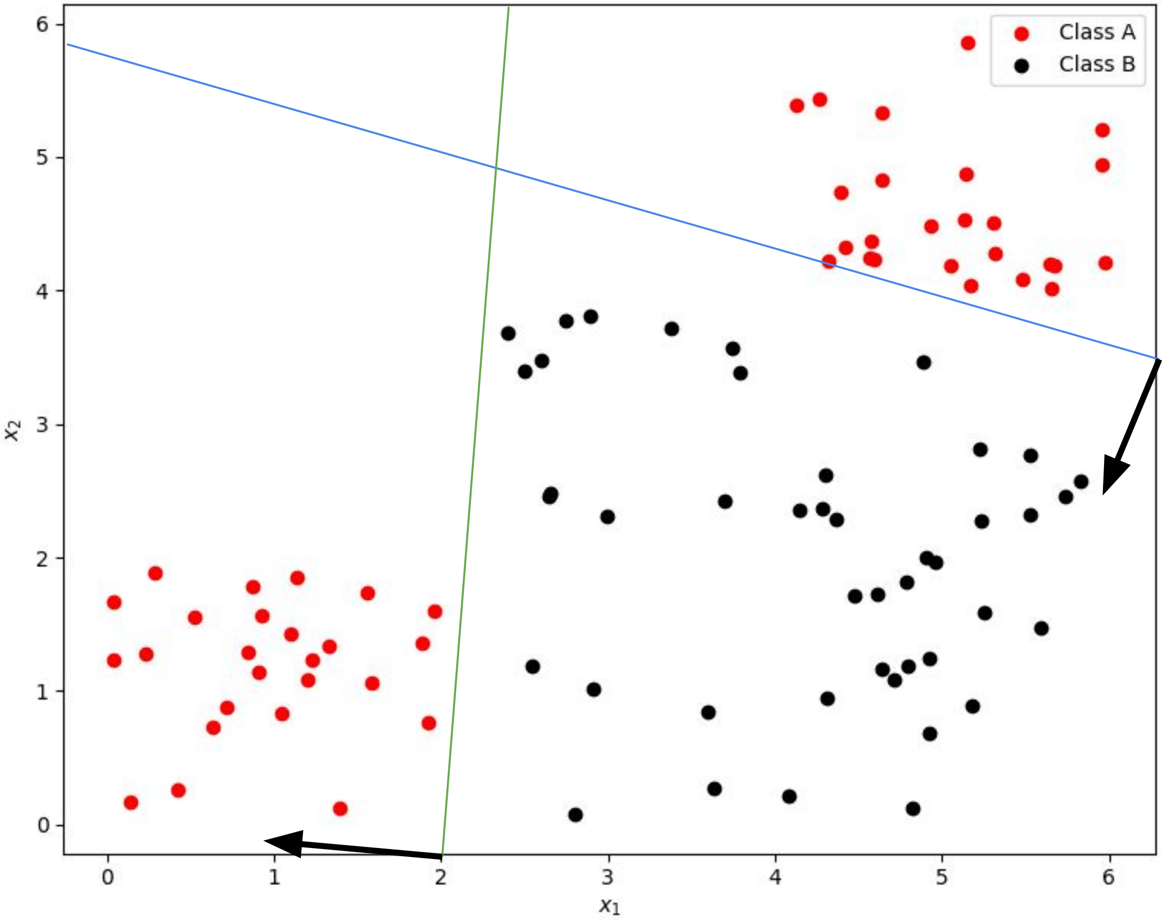


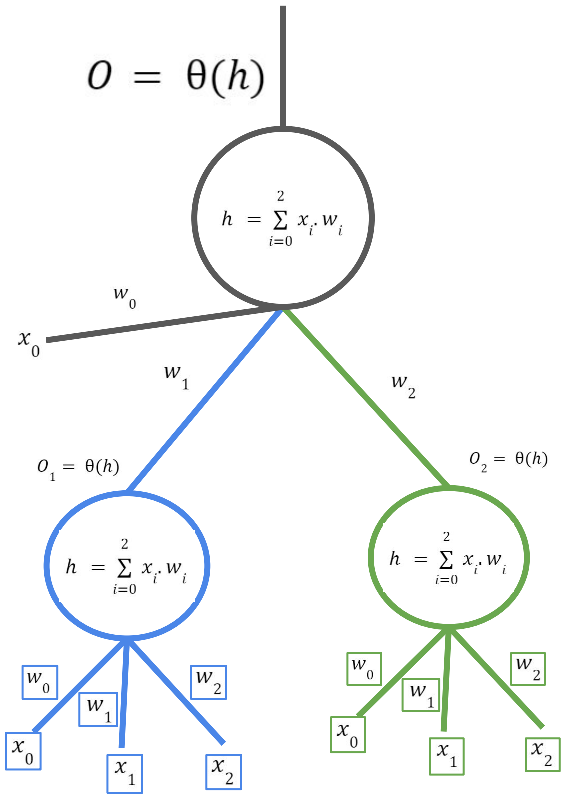
$\theta = función\ activación\ escalón$



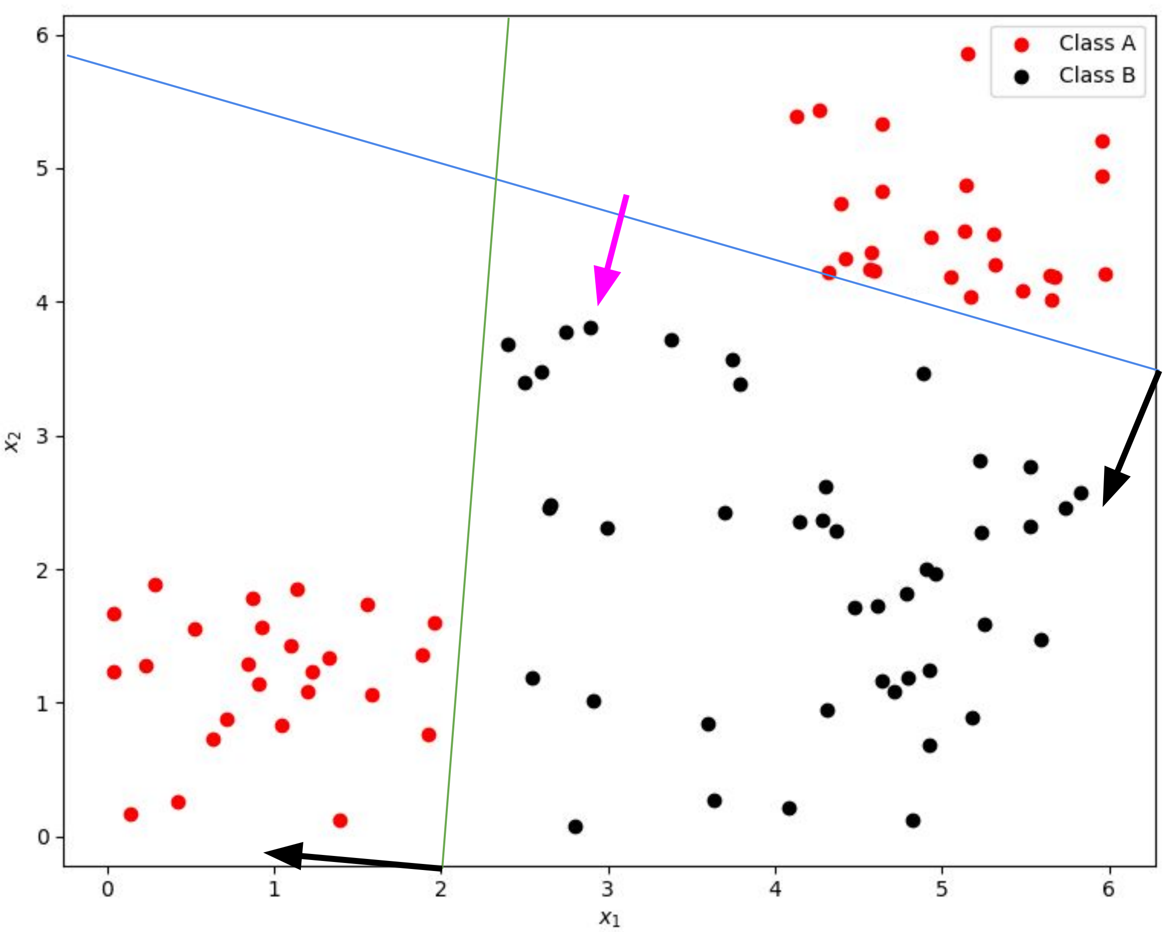


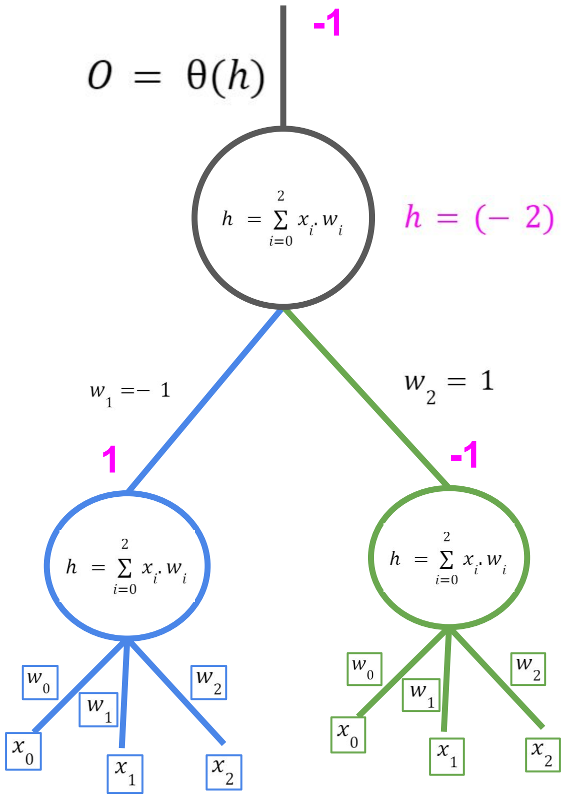
$\theta = función\ activación\ escalón$



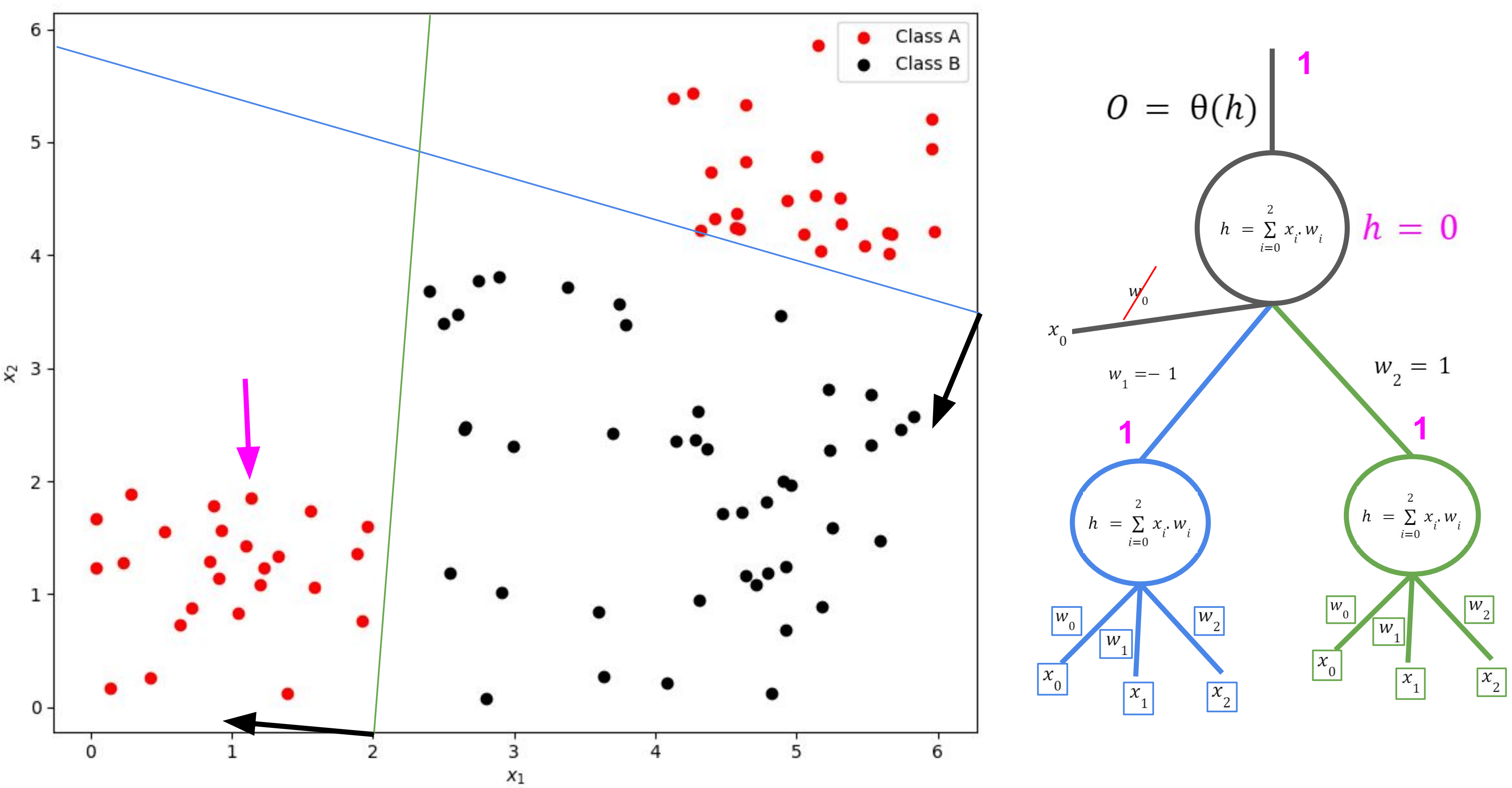


$\theta = función\ activación\ escalón$

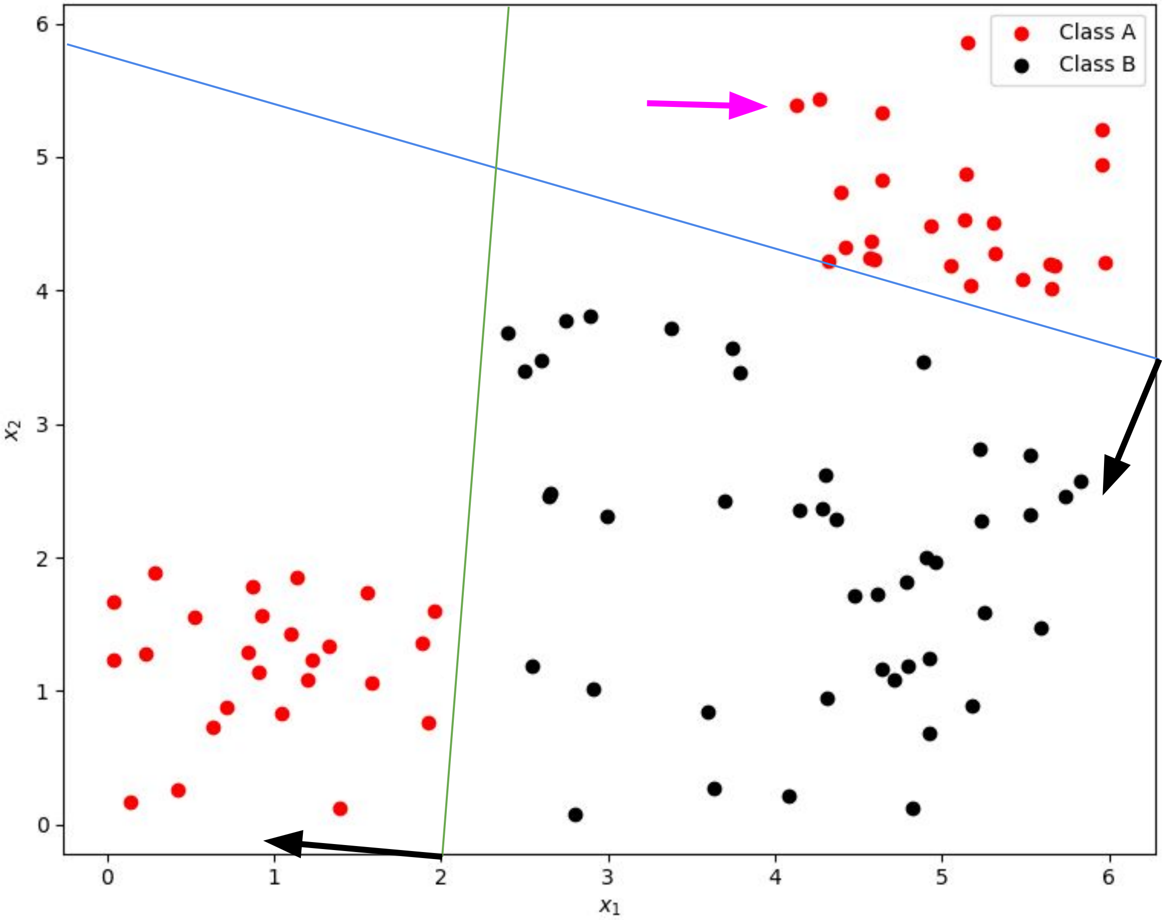


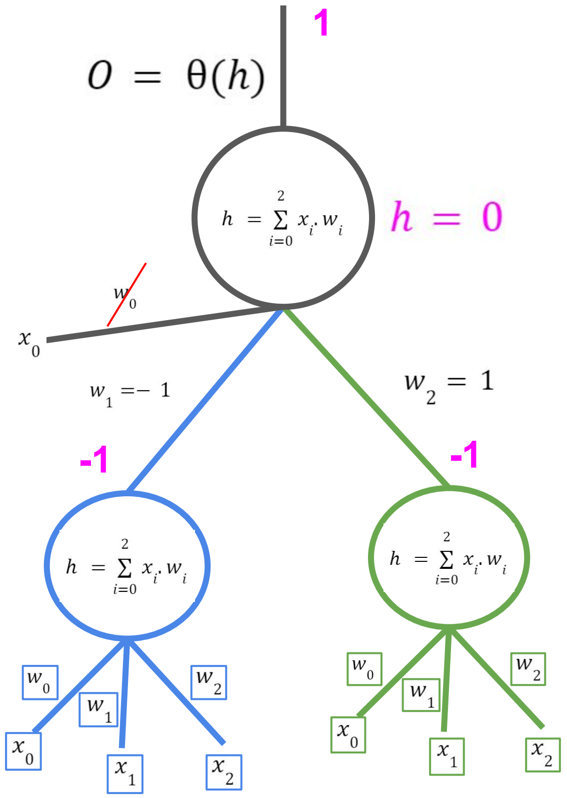


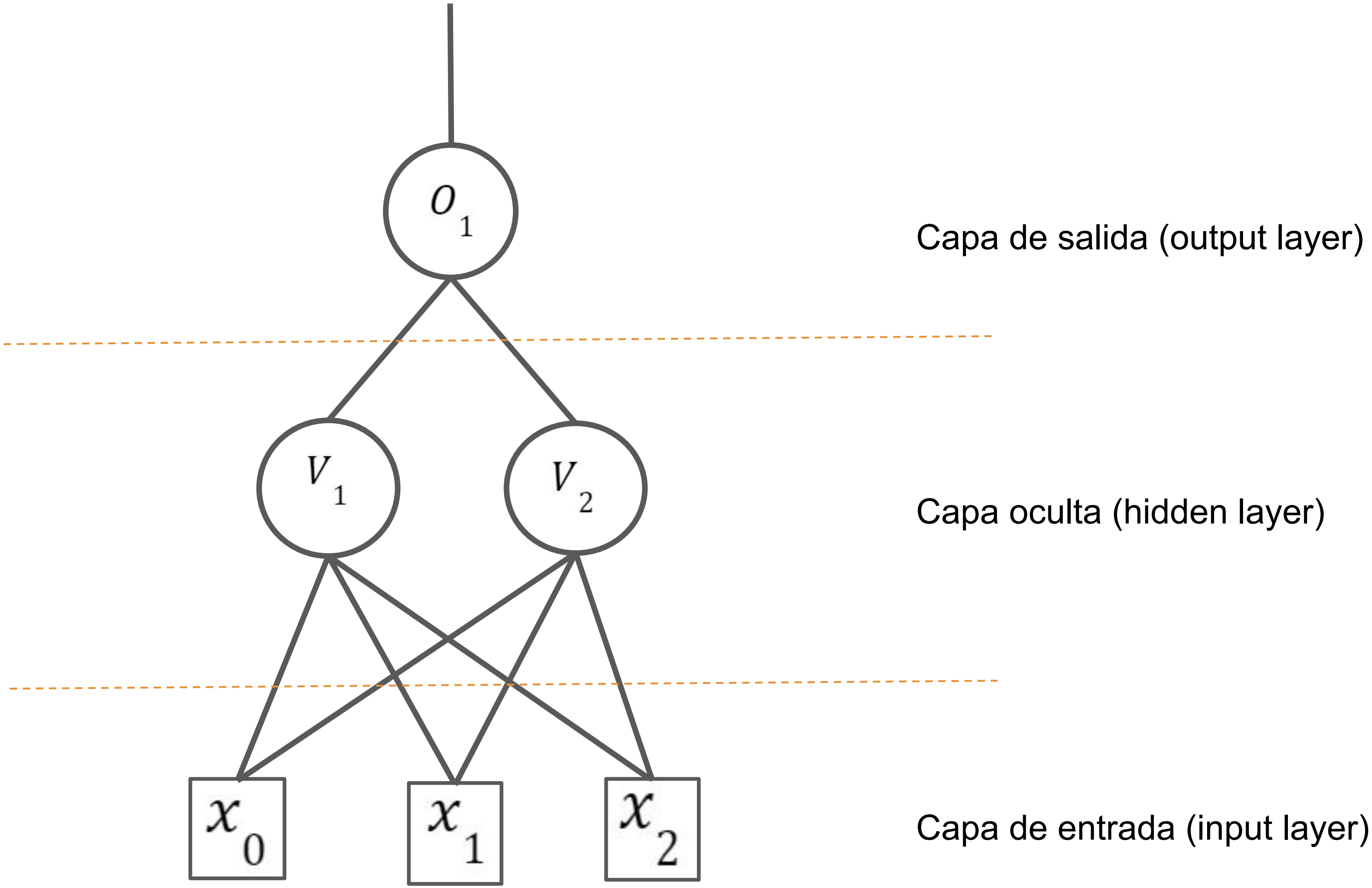
$\theta = función\ activación\ escalón$



$\theta\ =\ función\ activación\ escalón$



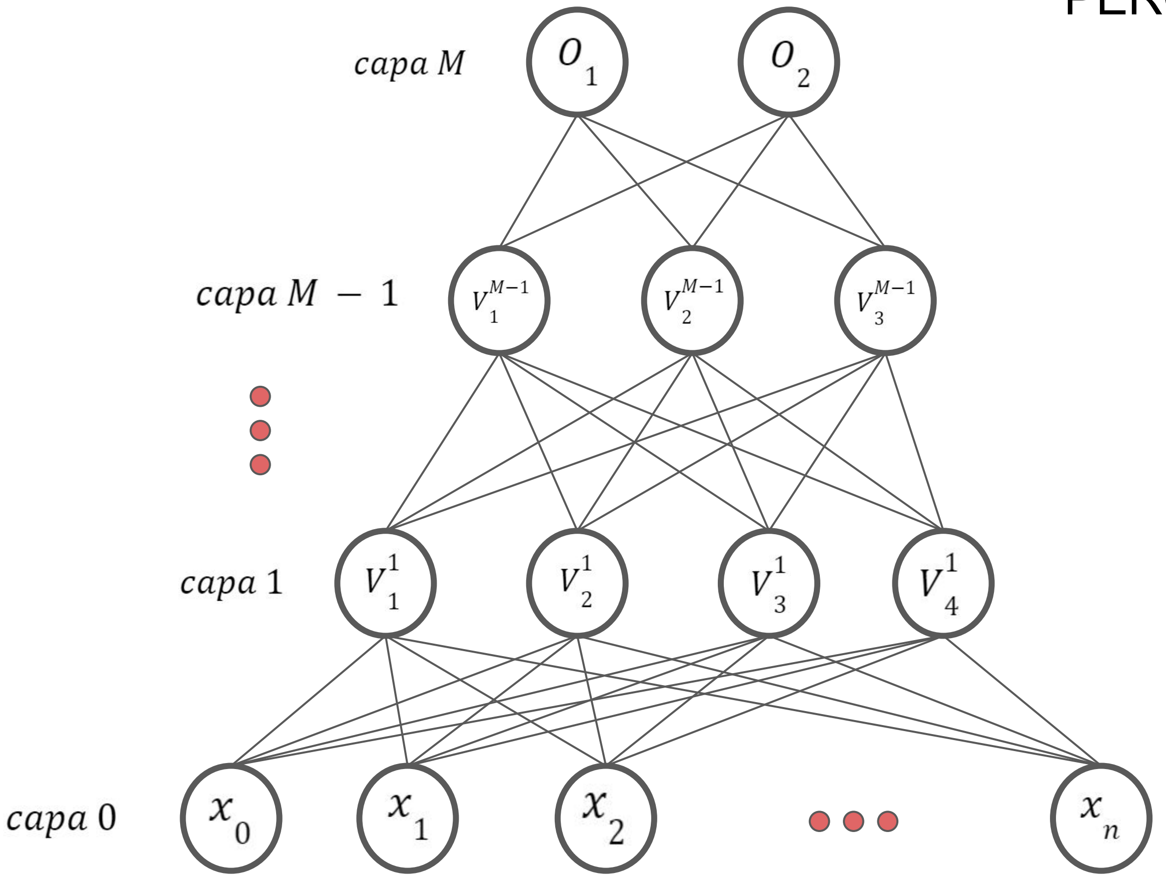




o_1
V_1
V_2
x_0
x_1
x_2
Capa de salida (output layer)
Capa oculta (hidden layer)
Capa de entrada (input layer)

$M = cantidad\ capas$

PERCEPTRÓN MULTICAPA



$p = cantidad\ de\ datos$
$n = cantidad\ de\ variables$

NOTACIÓN

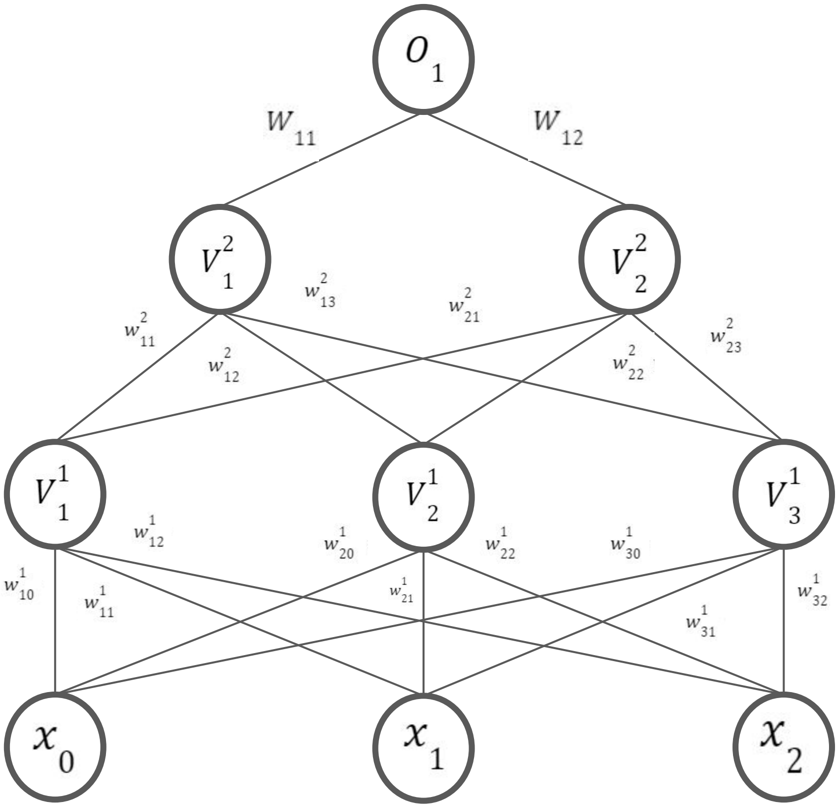
Índice	Descripción
i	índice de la neurona de la capa de salida (o siguiente)
j	índice de la neurona de la capa intermedia
k	índice de la neurona de entrada o de la capa anterior
m	índice de la capa intermedia
p	cantidad datos
μ	dato en particular

$$w^{m}_{jk} = pesos\ sinápticos$$

$$V^{m}_{j} = neurona\ de\ capa\ intermedia$$

$$W_{ij} = pesos\ sináticos\ de\ la\ última\ capa$$

$$O_{i} = neurona\ de\ capa\ de\ salida$$



NOTACIÓN

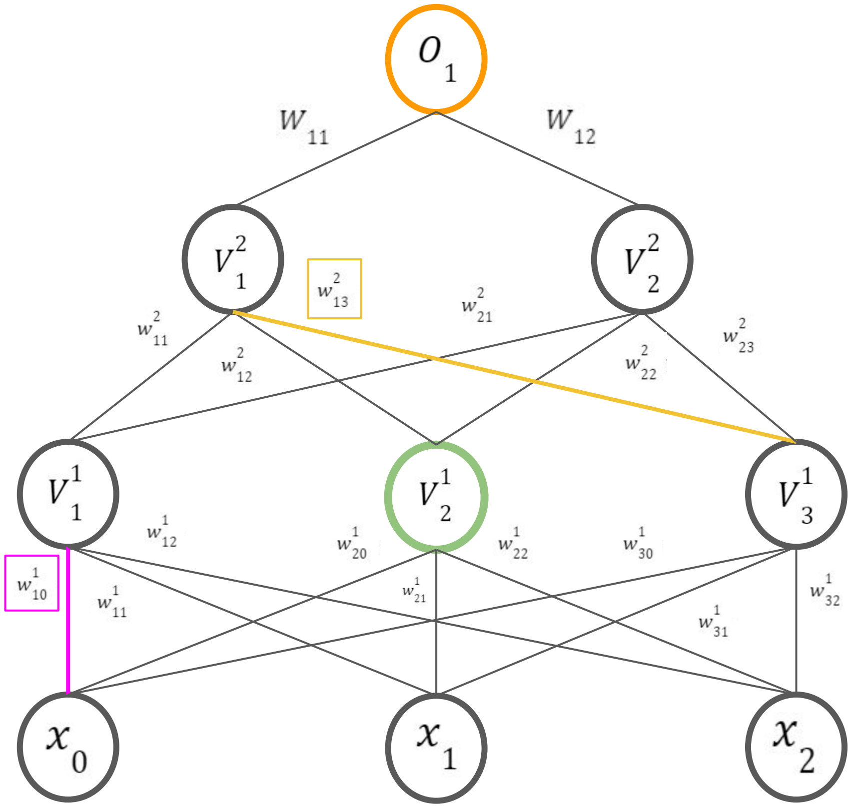
Índice	Descripción
i	índice de la neurona de la capa de salida
j	índice de la neurona de la capa intermedia
k	índice de la neurona de entrada o de la capa anterior
m	índice de la capa intermedia
p	cantidad datos
μ	dato en particular

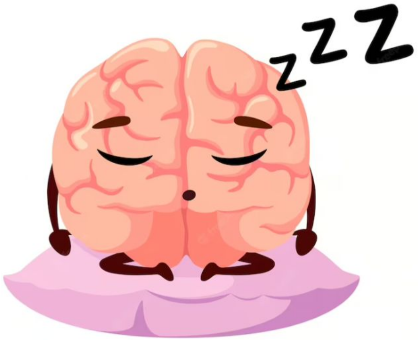
w_{jk}^{m} = *pesos sinápticos*

V_{j}^{m} = *neurona de capa intermedia*

W_{ij} = *pesos sináticos de la última capa*

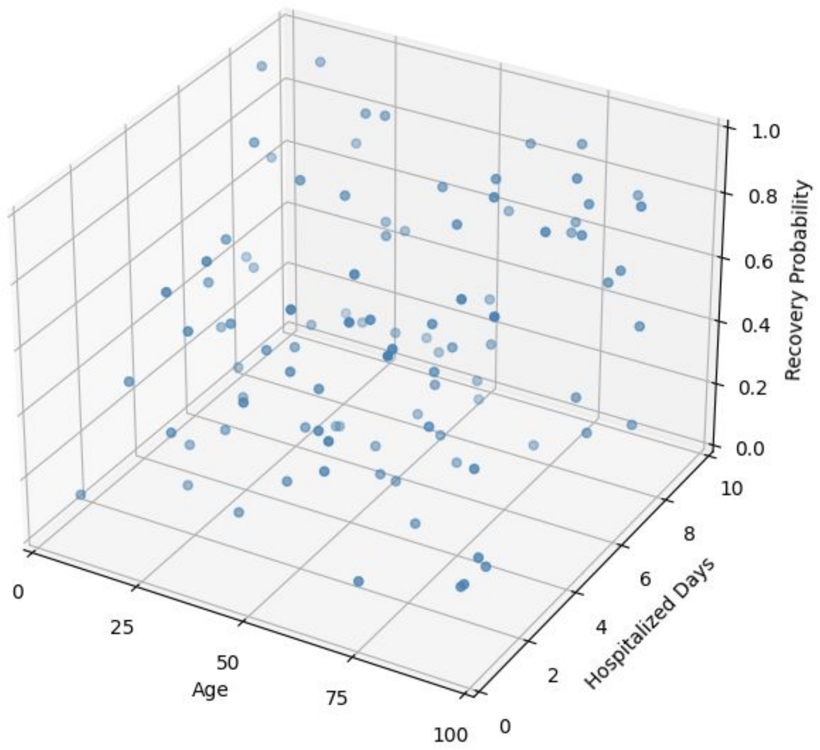
O_{i} = *neurona de capa de salida*

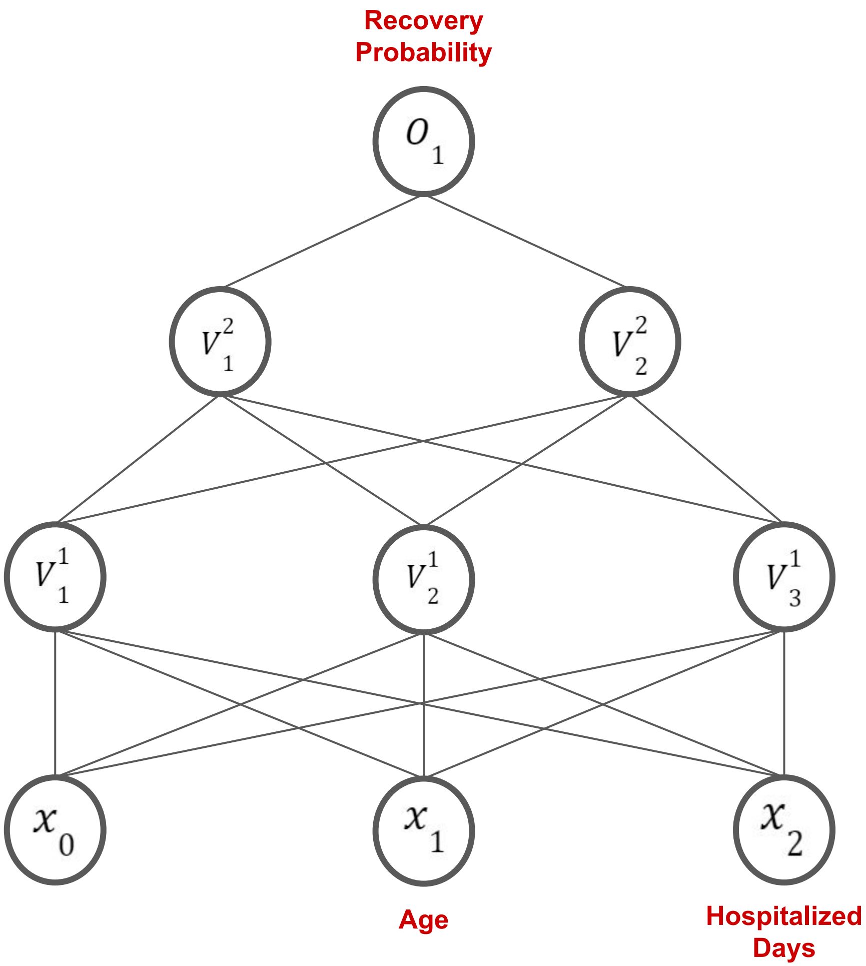



zzZ

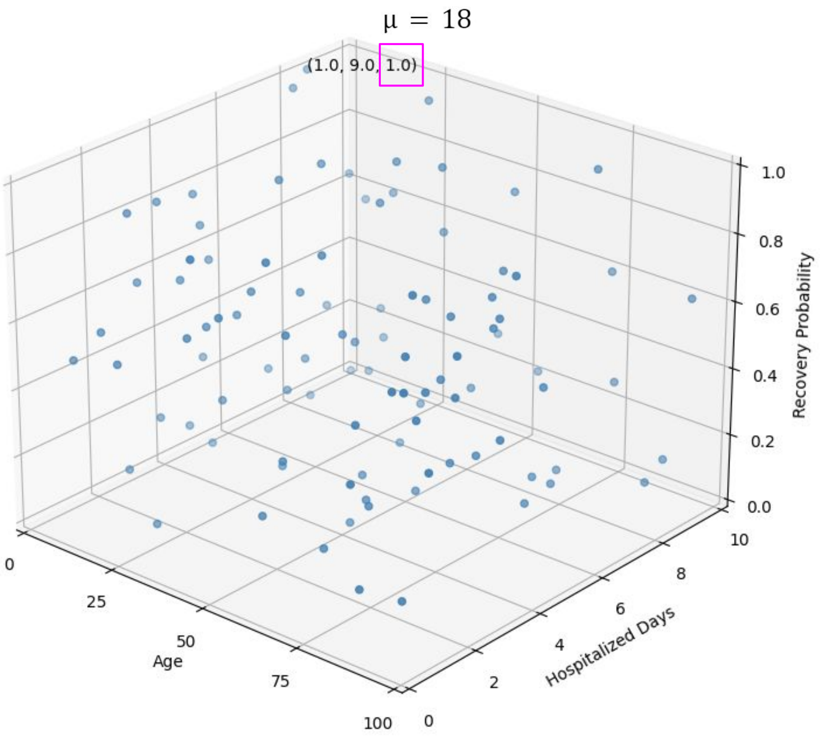
	PERCEPTRÓN SIMPLE	PERCEPTRÓN MULTICAPA
Proceso para obtener la salida de la neurona (“predecir”)	$O = \theta(\sum_{i=0}^{n} x_i . w_i)$	?
Función de error con respecto a la salida esperada	$E(O) = \frac{1}{2}\sum_{\mu=1}^{p} (\zeta^{\mu} - O^{\mu})^2$	?
Proceso de “aprendizaje” para ajustar los pesos	$w^{nuevo} = w^{anterior} + \Delta w$ $\Delta w = -\eta \frac{\partial E}{\partial w}$	?

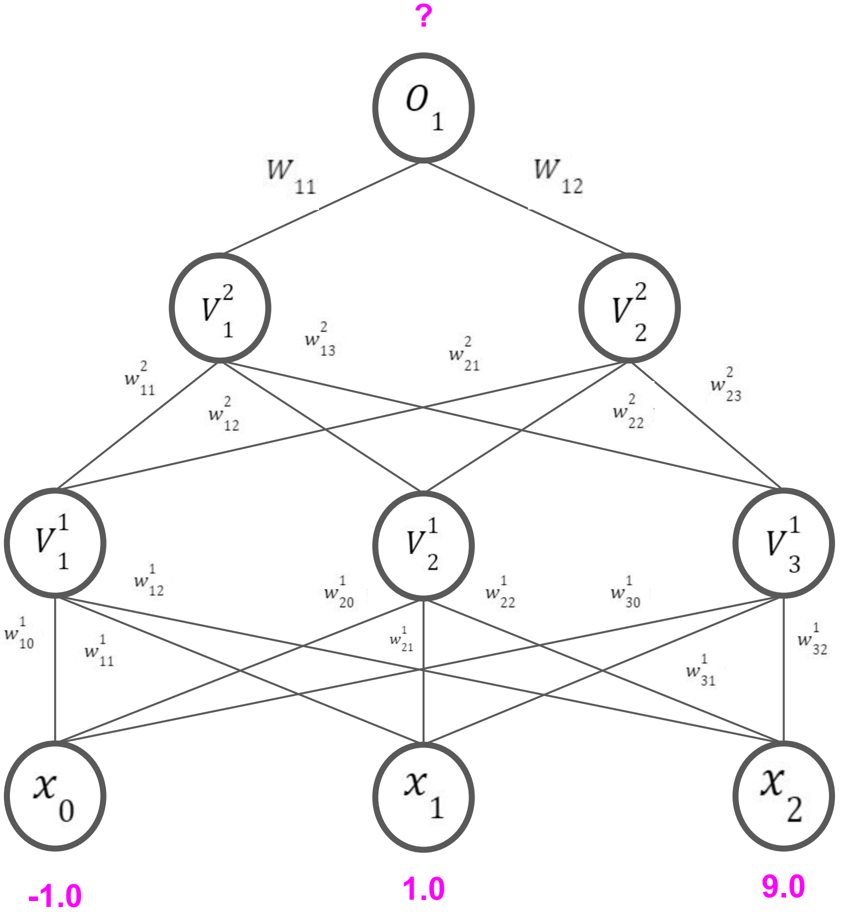
CALCULAR LA SALIDA DE LA NEURONA



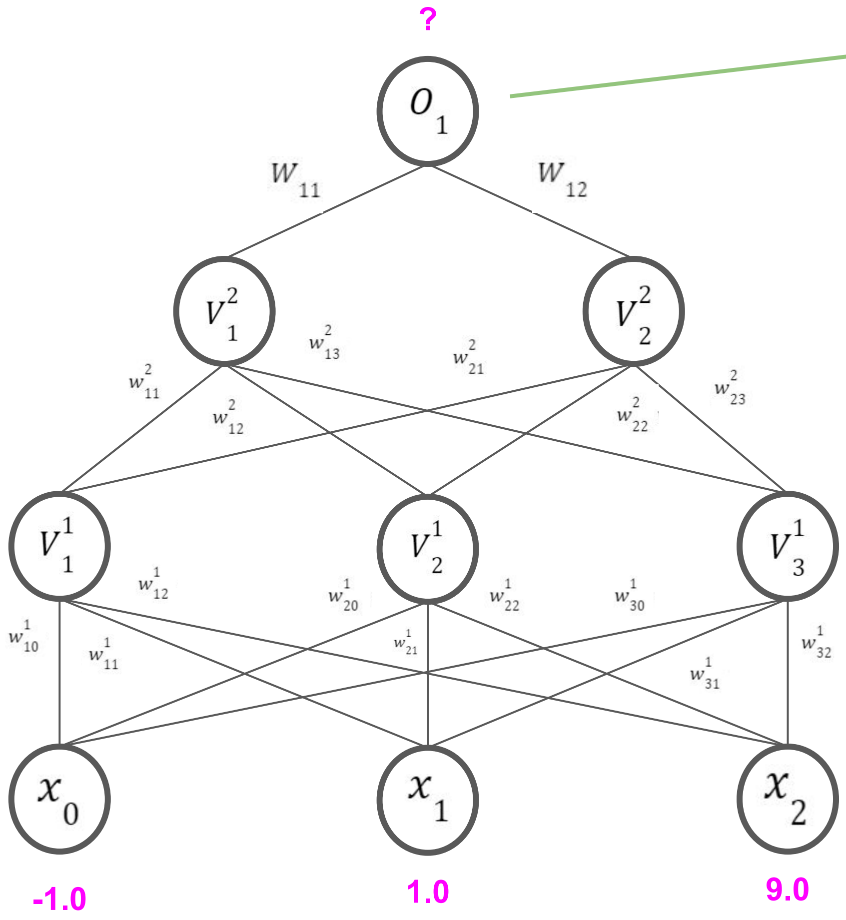


CALCULAR LA SALIDA DE LA NEURONA



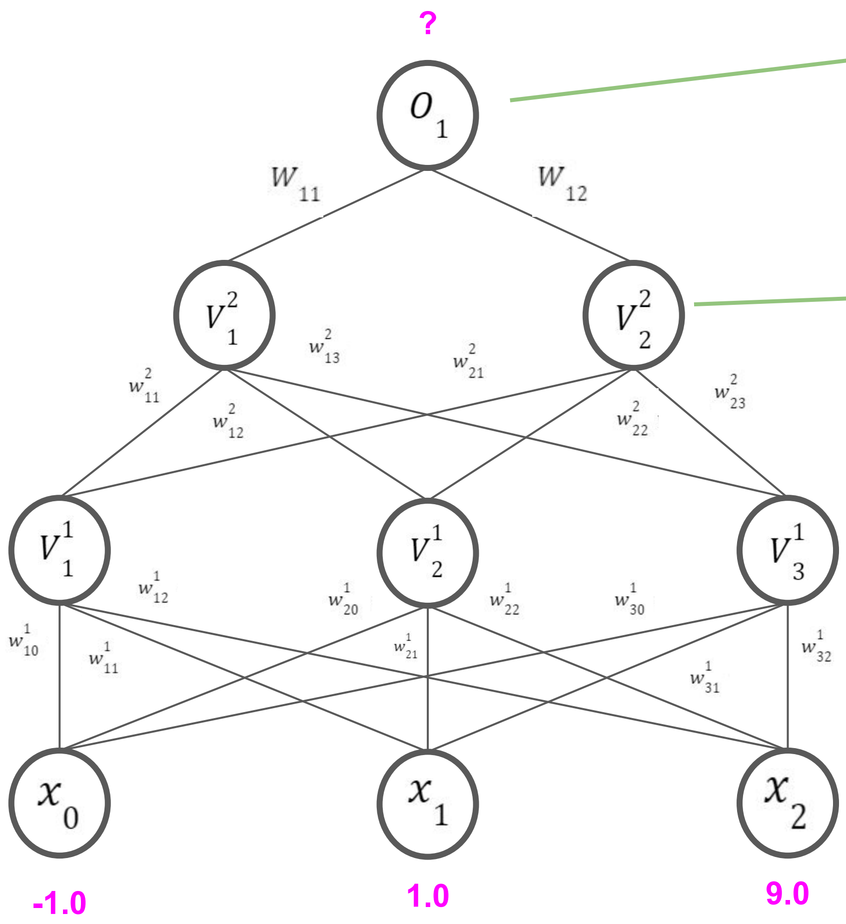


$\mu = 18$



$$O_1 = \theta\left(\sum_{k=1}^{2} V_k^2 \,.\, W_{1k}\right)$$

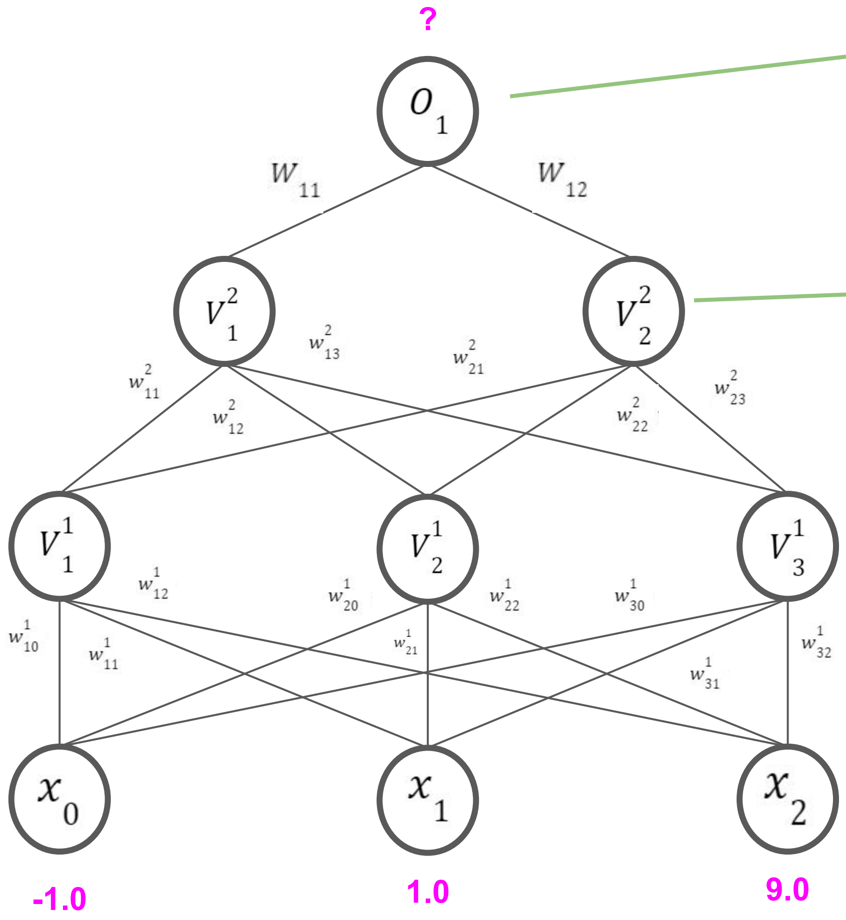
$\mu = 18$



$$O_1 = \theta\left(\sum_{k=1}^{2} V^2_k \,.\, W_{1k}\right)$$

$$V^2_2 = \theta\left(\sum_{k=1}^{3} V^1_k \,.\, w^2_{2k}\right)$$

$\mu = 18$

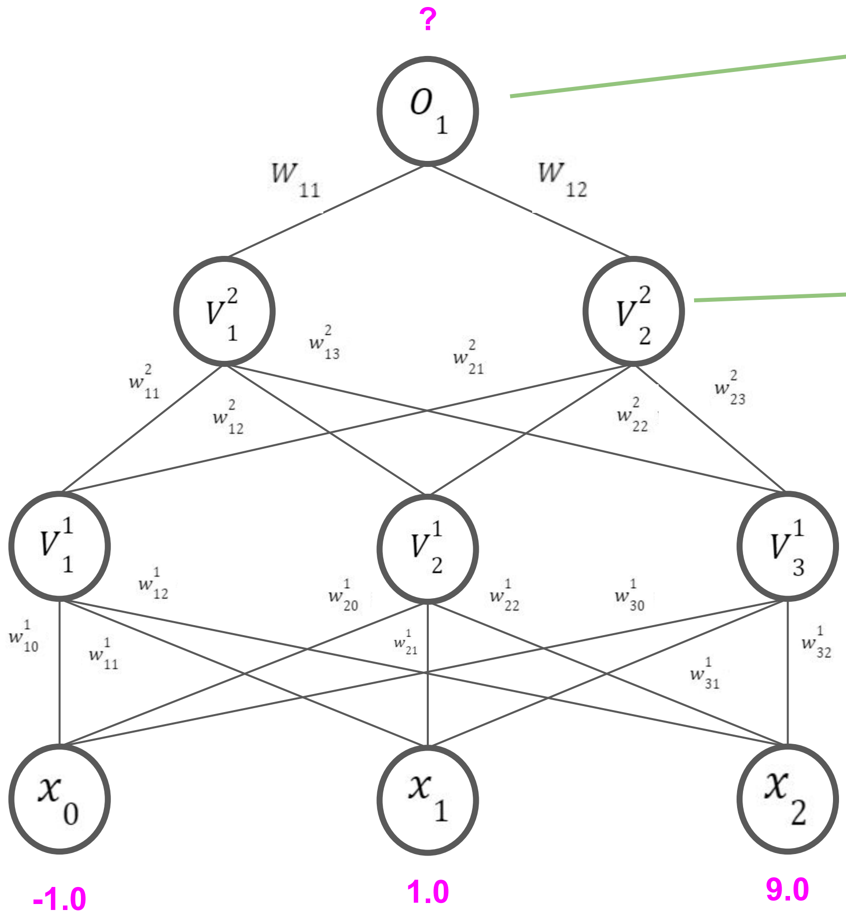


$$O_1 = \theta\left(\sum_{k=1}^{2} V^2_k \,.\, W_{1k}\right)$$

$$V^2_2 = \theta\left(\sum_{k=1}^{3} V^1_k \,.\, w^2_{2k}\right)$$

$$V^1_3 = \theta\left(\sum_{k=0}^{2} x^{18}_k \,.\, w^1_{3k}\right)$$

$\mu = 18$



$$O_1 = \theta\left(\sum_{k=1}^{2} V_k^2 \, . \, W_{1k}\right)$$

$$V_2^2 = \theta\left(\sum_{k=1}^{3} V_k^1 \, . \, w_{2k}^2\right)$$

$$V_3^1 = \theta\left(\sum_{k=0}^{2} x_k^{18} \, . \, w_{3k}^1\right)$$

Propagación
hacia adelante
(forward
propagation)

PROPAGACIÓN HACIA ADELANTE (FORWARD PROPAGATION)

Salida de la neurona:

$$O_i = \theta(\sum_{k=1} V_k^{M-1} . W_{ik})$$

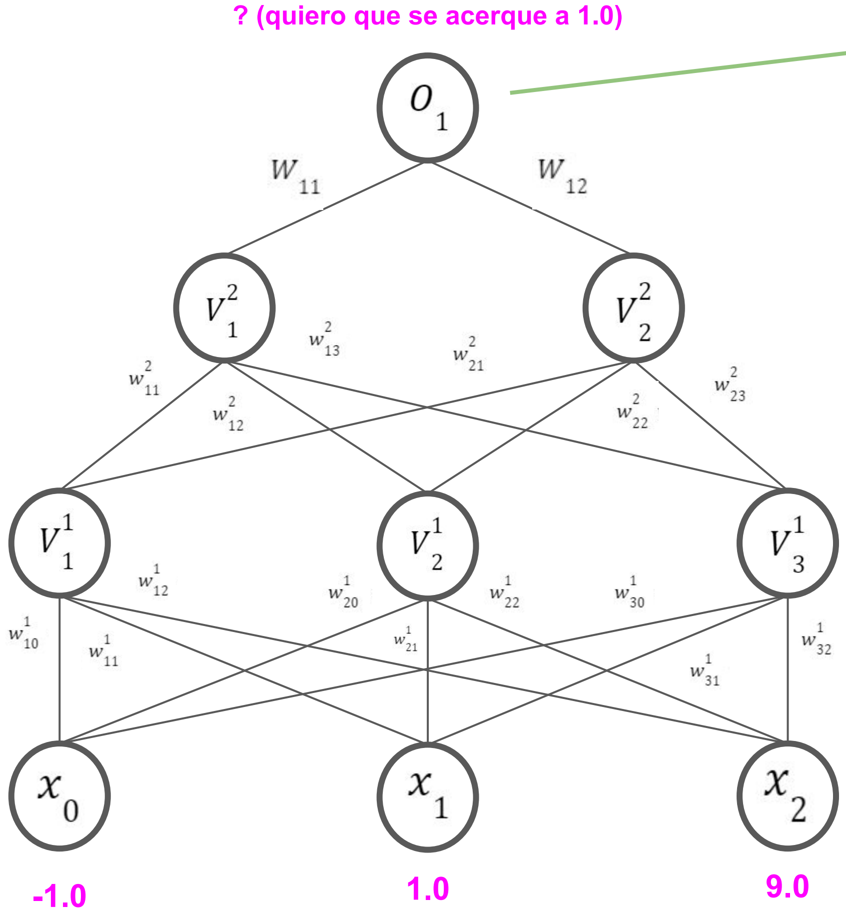
Salida de la neurona de una capa intermedia:

$$V_j^m = \theta(\sum_{k=1} V_k^{m-1} . w_{jk}^m) \qquad m = 2 \ldots M - 1$$

Salida de la neurona de la primera capa intermedia:

$$V_j^1 = \theta(\sum_{k=1} x_k^\mu . w_{jk}^1)$$

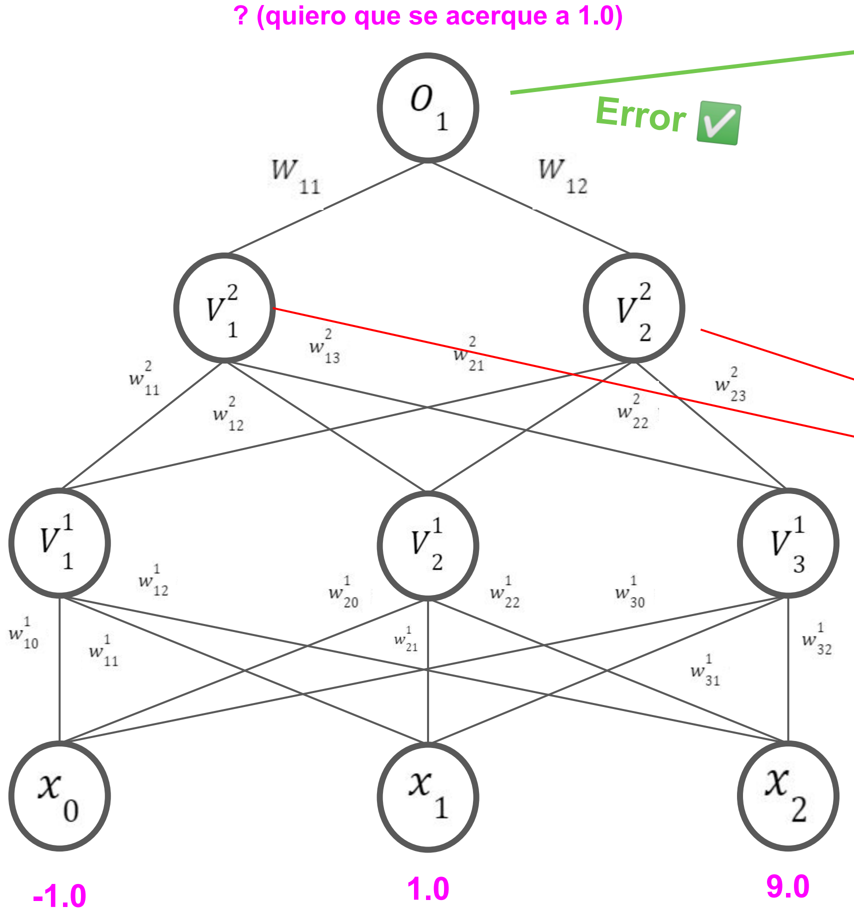
	PERCEPTRÓN SIMPLE	PERCEPTRÓN MULTICAPA
Proceso para obtener la salida de la neurona ("predecir")	$O = \theta(\sum_{i=0}^{n} x_i . w_i)$	$V_j^m = \theta(\sum_{k=1} V_k^{m-1} . w_{jk}^m)$ $m = 1 ... M \ (O_i = V_i^M \ , x_i = V_i^0)$
Función de costo (medir error con respecto a la salida esperada)	$E(O) = \frac{1}{2}(\zeta^\mu - O^\mu)^2$	?
Proceso de "aprendizaje" para ajustar los pesos	$w^{nuevo} = w^{anterior} + \Delta w$ $\Delta w = - \eta \frac{\partial E}{\partial w}$	?



$$O_1 = \theta\left(\sum_{k=1}^{2} V_k^2 \, . W_{1k}\right)$$

$$E(O) = \frac{1}{2}\sum_i \left(\zeta^\mu - O_i^\mu\right)^2$$

	PERCEPTRÓN SIMPLE	PERCEPTRÓN MULTICAPA
Proceso para obtener la salida de la neurona ("predecir")	$O = \theta(\sum_{i=0}^{n} x_i . w_i)$	$V_j^m = \theta(\sum_{k=1} V_k^{m-1} . w_{mk})$ $m = 1 ... M \ (O_i = V_i^M , x_i = V_i^0)$
Función de error con respecto a la salida esperada	$E(O) = \frac{1}{2}(\zeta^\mu - O^\mu)^2$	$E(O) = \frac{1}{2}\sum_i (\zeta^\mu - O_i^\mu)^2$
Proceso de "aprendizaje" para ajustar los pesos	$w^{nuevo} = w^{anterior} + \Delta w$ $\Delta w = - \eta \frac{\partial E}{\partial w}$	?



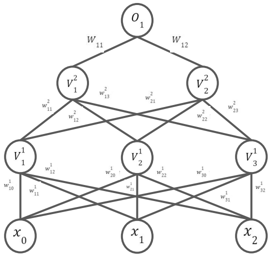
Error ✅

$$O_1 = \theta\left(\sum_{k=1}^{2} V_k^2 \,.\, W_{1k}\right)$$

$$E(O) = \frac{1}{2}\sum_i \left(\zeta^\mu - O_i^\mu\right)^2$$

Error ❓

RUMELHART, HINTON, WILLIAMS (1986)



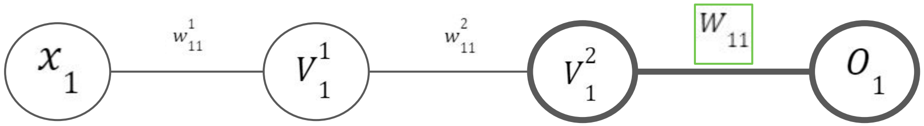
$$\Delta w = \begin{bmatrix} 0.1 \\ -3.4 \\ 2 \\ 0.6 \\ 0.34 \\ -1.2 \\ -0.01 \\ 0.86 \\ -1.9 \\ -2.3 \\ 0.7 \\ 0.02 \\ \dots \end{bmatrix}$$

- Trabajan en un nuevo algoritmo para actualizar los pesos de una red neuronal llamado **retropropagación** (***back-propagation***)

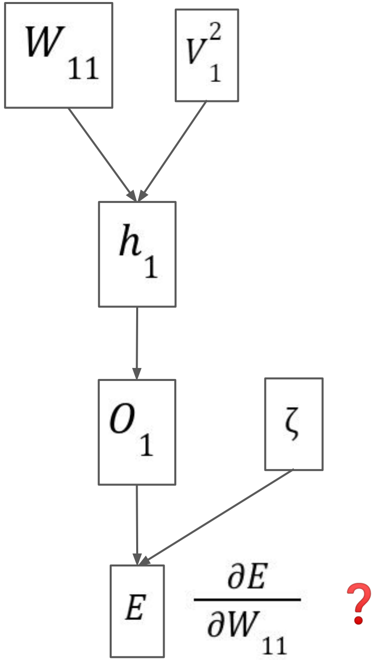
- Para calcular la actualización de los pesos utilizaremos el algoritmo del gradiente descendente y la regla de la cadena para la diferenciación.

Paper: https://www.nature.com/articles/323533a0

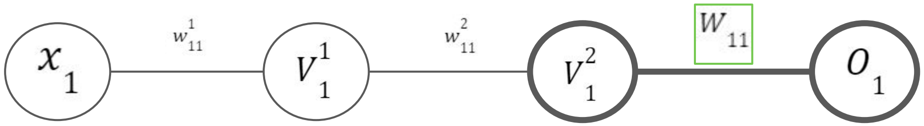
RETROPROPAGACIÓN: EXPLICACIÓN CONCEPTUAL



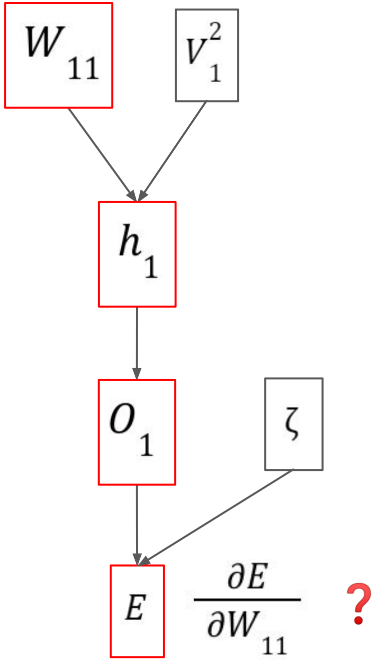
$\zeta = 1.0$



RETROPROPAGACIÓN: EXPLICACIÓN CONCEPTUAL



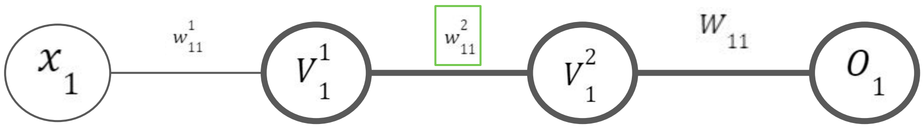
$\zeta = 1.0$



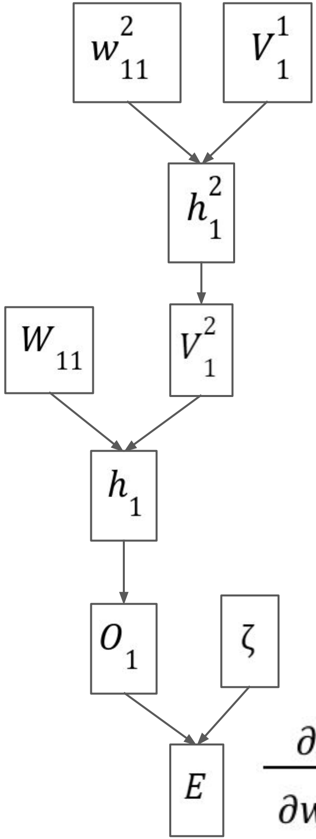
Aplicamos la regla de la cadena:

$$\frac{\partial E}{\partial W_{11}} = \frac{\partial E}{\partial O_1} \frac{\partial O_1}{\partial h_1} \frac{\partial h_1}{\partial W_{11}}$$

RETROPROPAGACIÓN: EXPLICACIÓN CONCEPTUAL

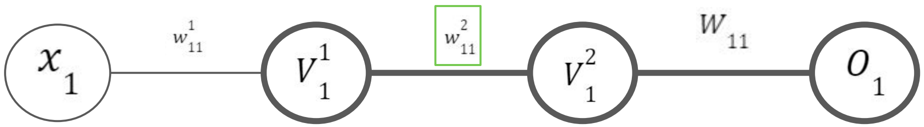


$$\zeta = 1.0$$



$$\frac{\partial E}{\partial w^{2}_{11}} \ ?$$

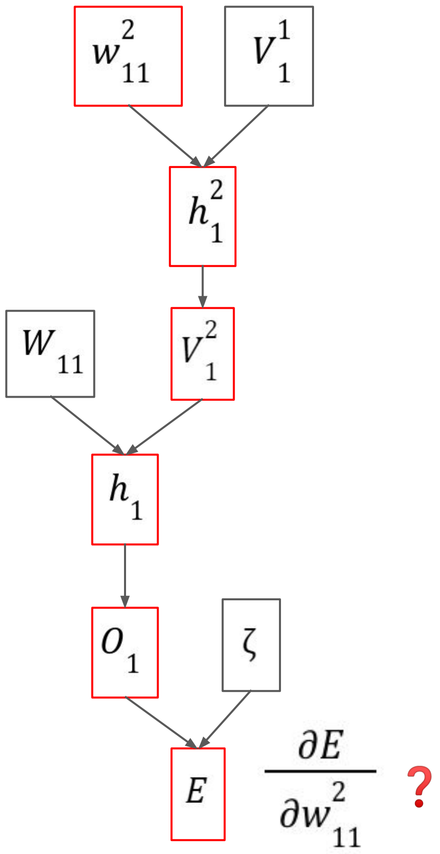
RETROPROPAGACIÓN: EXPLICACIÓN CONCEPTUAL



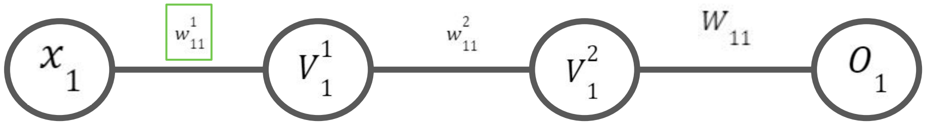
$\zeta = 1.0$

Aplicamos la regla de la cadena:

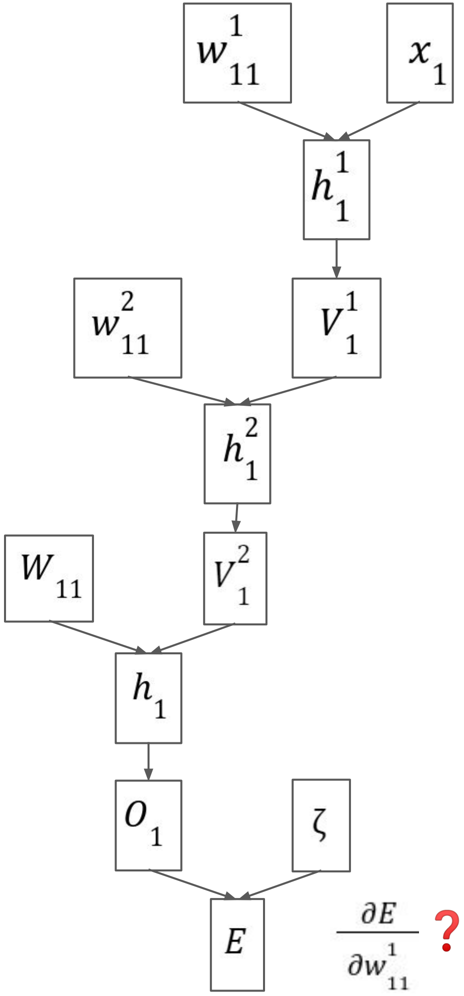
$$\frac{\partial E}{\partial w^2_{11}} = \frac{\partial E}{\partial O_1}\frac{\partial O_1}{\partial h_1}\frac{\partial h_1}{\partial V^2_1}\frac{\partial V^2_1}{\partial h^2_1}\frac{\partial h^2_1}{\partial w^2_{11}}$$



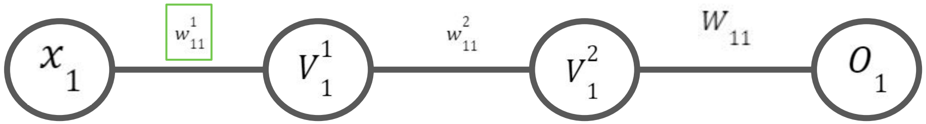
RETROPROPAGACIÓN: EXPLICACIÓN CONCEPTUAL



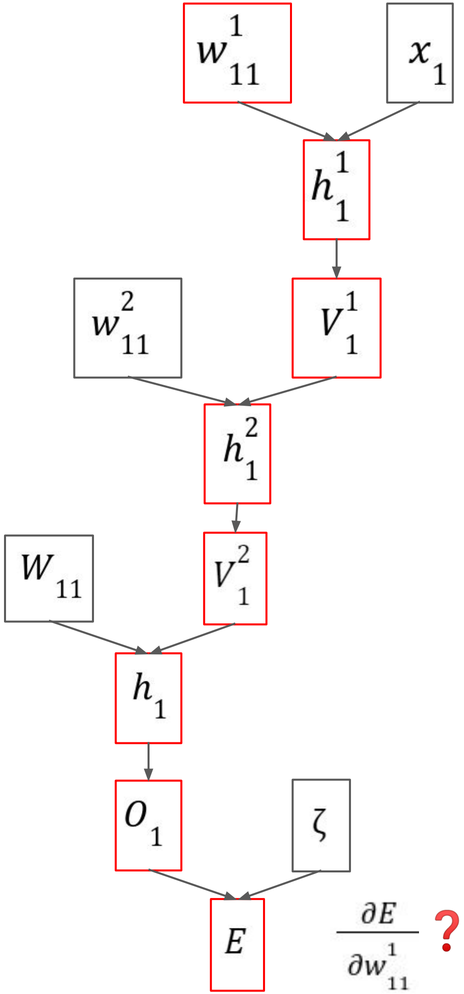
$$\zeta = 1.0$$



RETROPROPAGACIÓN: EXPLICACIÓN CONCEPTUAL

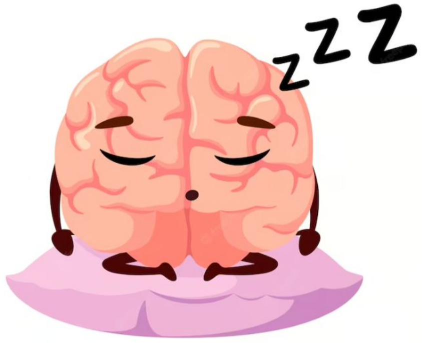


$\zeta = 1.0$



Aplicamos la regla de la cadena:

$$\frac{\partial E}{\partial w^1_{11}} = \frac{\partial E}{\partial O_1}\frac{\partial O_1}{\partial h_1}\frac{\partial h_1}{\partial V^2_1}\frac{\partial V^2_1}{\partial h^2_1}\frac{\partial h^2_1}{\partial V^2_1}\frac{\partial V^2_1}{\partial h^1_1}\frac{\partial h^1_1}{\partial w^1_{11}}$$

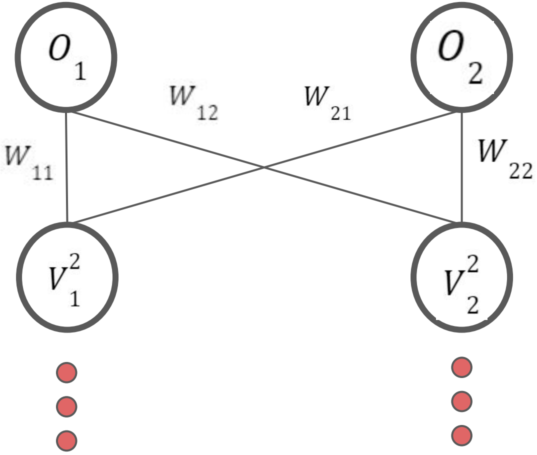


RETROPROPAGACIÓN: CAPA DE SALIDA

$$E(O) = \frac{1}{2}\sum_{i} (\zeta^{\mu} - O_{i}^{\mu})^{2}$$

$$O_{i} = \theta(h_{i})$$

$$h_{i} = \sum_{j=1} V_{j}^{M-1} \, . W_{ij}$$



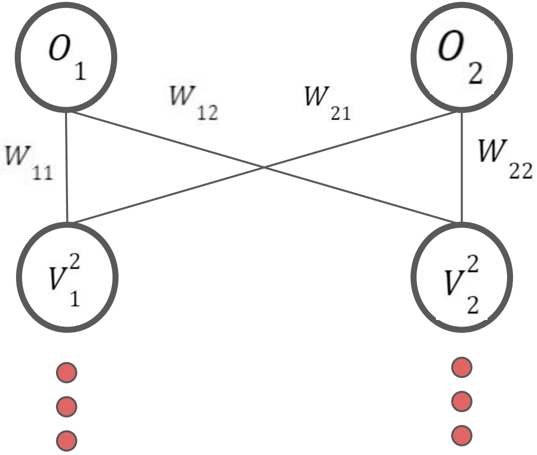
$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial O_{i}} \frac{\partial O_{i}}{\partial h_{i}} \frac{\partial h_{i}}{\partial W_{ij}}$$

RETROPROPAGACIÓN: CAPA DE SALIDA

$$E(O) = \frac{1}{2}\sum_i (\zeta^{\mu} - O_i^{\mu})^2$$

$$O_i = \theta(h_i)$$

$$h_i = \sum_{j=1} V_j^{M-1} . W_{ij}$$



$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial W_{ij}} = \boxed{\frac{\partial E}{\partial O_i}} \boxed{\frac{\partial O_i}{\partial h_i}} \boxed{\frac{\partial h_i}{\partial W_{ij}}}$$

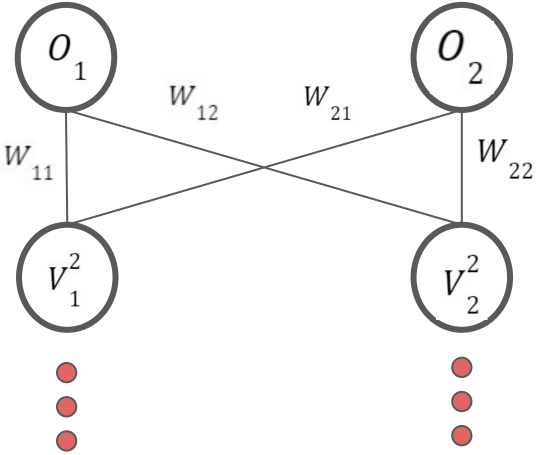
$$\frac{\partial E}{\partial W_{ij}} = (\zeta_i - O_i)(-1)\,\theta'(h_i)(1)\,V_j^{M-1}$$

RETROPROPAGACIÓN: CAPA DE SALIDA

$$E(O) = \frac{1}{2}\sum_i (\zeta^{\mu} - O_i^{\mu})^2$$

$$O_i = \theta(h_i)$$

$$h_i = \sum_{j=1} V_j^{M-1} . W_{ij}$$



$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial W_{ij}} = \boxed{\frac{\partial E}{\partial O_i}} \boxed{\frac{\partial O_i}{\partial h_i}} \boxed{\frac{\partial h_i}{\partial W_{ij}}}$$

$$\frac{\partial E}{\partial W_{ij}} = (\zeta_i - O_i)(-1)\,\theta'(h_i)(1)\,V_j^{M-1}$$

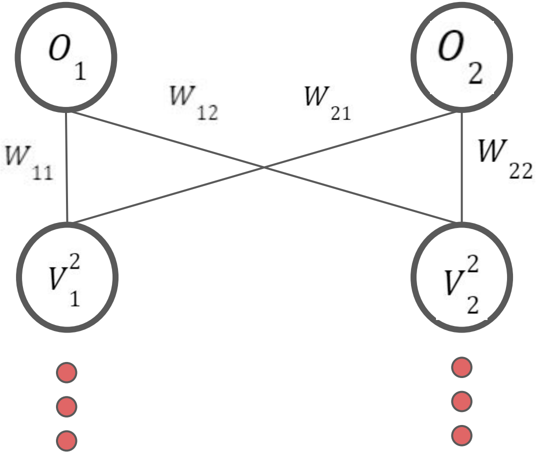
$$\frac{\partial E}{\partial W_{ij}} = -\delta_i V_j^{M-1} \qquad \delta_i = (\zeta_i - O_i)\theta'(h_i)$$

RETROPROPAGACIÓN: CAPA DE SALIDA

$$E(O) = \frac{1}{2}\sum_i (\zeta^{\mu} - O_i^{\mu})^2$$

$$O_i = \theta(h_i)$$

$$h_i = \sum_{j=1} V_j^{M-1} . W_{ij}$$



$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial W_{ij}} = \boxed{\frac{\partial E}{\partial O_i}} \boxed{\frac{\partial O_i}{\partial h_i}} \boxed{\frac{\partial h_i}{\partial W_{ij}}}$$

$$\frac{\partial E}{\partial W_{ij}} = (\zeta_i - O_i)(-1)\,\theta'(h_i)(1)\,V_j^{M-1}$$

$$\frac{\partial E}{\partial W_{ij}} = -\delta_i V_j^{M-1} \qquad \delta_i = (\zeta_i - O_i)\theta'(h_i)$$

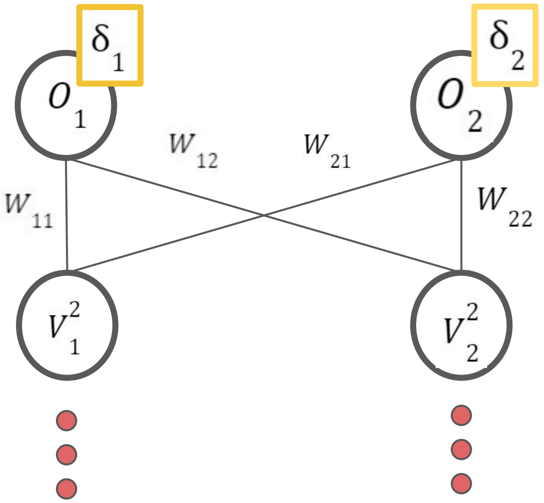
$$\Delta W_{ij} = \eta \delta_i V_j^{M-1}$$

$$E(O) = \frac{1}{2}\sum_i (\zeta^\mu - O_i^\mu)^2$$

$$O_i = \theta(h_i)$$

$$h_i = \sum_{j=1} V_j^{M-1} . W_{ij}$$

RETROPROPAGACIÓN: CAPA DE SALIDA



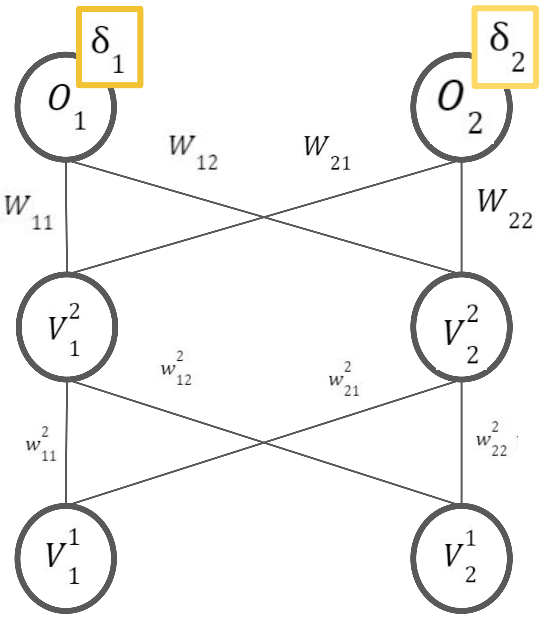
$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial W_{ij}} = \boxed{\frac{\partial E}{\partial O_i}} \boxed{\frac{\partial O_i}{\partial h_i}} \boxed{\frac{\partial h_i}{\partial W_{ij}}}$$

$$\frac{\partial E}{\partial W_{ij}} = (\zeta_i - O_i)(-1)\,\theta'(h_i)(1)\,V_j^{M-1}$$

$$\frac{\partial E}{\partial W_{ij}} = -\delta_i V_j^{M-1} \qquad \delta_i = (\zeta_i - O_i)\theta'(h_i)$$

$$\Delta W_{ij} = \eta \delta_i V_j^{M-1}$$

RETROPROPAGACIÓN: CAPA OCULTA (*M-1*)



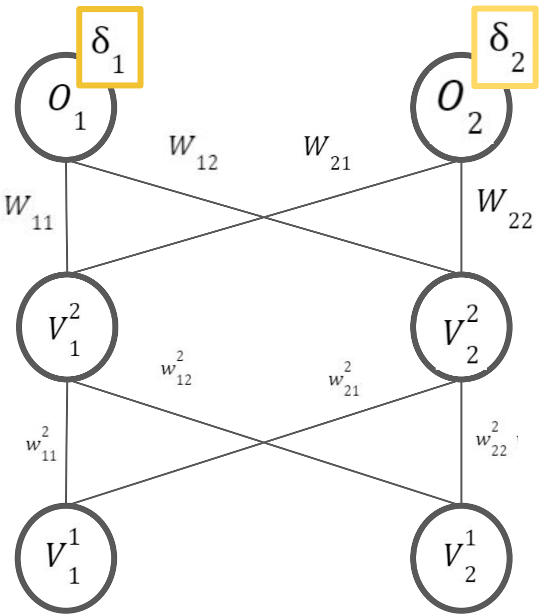
$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial w^m_{jk}} = \frac{\partial E}{\partial O} \frac{\partial O}{\partial h} \frac{\partial h}{\partial V^m_j} \frac{\partial V^m_j}{\partial h^m_j} \frac{\partial h^m_j}{\partial w^m_{jk}}$$

RETROPROPAGACIÓN: CAPA OCULTA (*M-1*)

$$E(O) = \frac{1}{2}\sum_i (\zeta^\mu - O_i^\mu)^2$$

$$O_i = \theta(h_i)$$

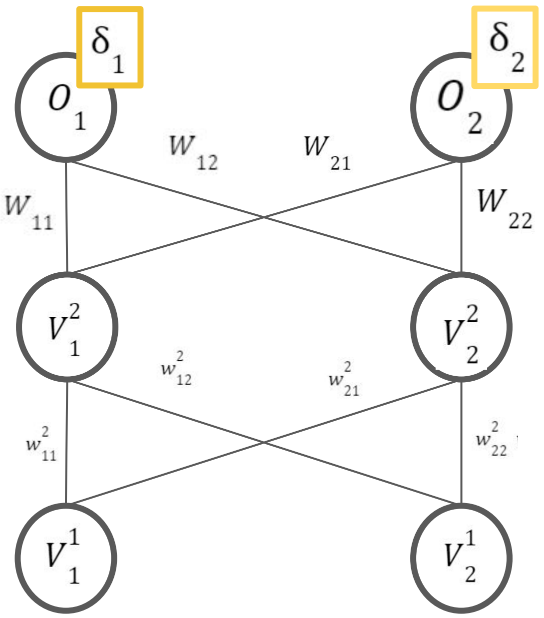
$$h_i = \sum_{j=1} V_j^{M-1} . W_{ij}$$



$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial w_{jk}^m} = \boxed{\frac{\partial E}{\partial O}} \boxed{\frac{\partial O}{\partial h}} \boxed{\frac{\partial h}{\partial V_j^m}} \frac{\partial V_j^m}{\partial h_j^m} \frac{\partial h_j^m}{\partial w_{jk}^m}$$

$$\frac{\partial E}{\partial w_{jk}^m} = (-1)\sum_i (\zeta_i - O_i)\,\theta'(h_i)(1)\,W_{ij}\,\frac{\partial V_j^m}{\partial h_j^m}\,\frac{\partial h_j^m}{\partial w_{jk}^m}$$

RETROPROPAGACIÓN: CAPA OCULTA (*M-1*)



$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial w^m_{jk}} = \boxed{\frac{\partial E}{\partial O}}\, \boxed{\frac{\partial O}{\partial h}}\, \boxed{\frac{\partial h}{\partial V^m_j}}\, \frac{\partial V^m_j}{\partial h^m_j}\, \frac{\partial h^m_j}{\partial w^m_{jk}}$$

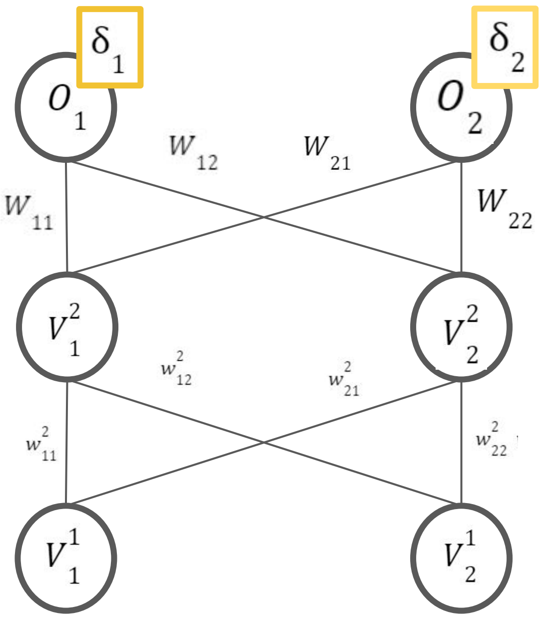
$$\frac{\partial E}{\partial w^m_{jk}} = (-1) \sum_i \underbrace{(\zeta_i - O_i)\, \theta'(h_i)(1)}_{\delta_i}\, W_{ij}\, \frac{\partial V^m_j}{\partial h^m_j}\, \frac{\partial h^m_j}{\partial w^m_{jk}}$$

$$\frac{\partial E}{\partial w^m_{jk}} = -\sum_i \delta_i W_{ij}\, \frac{\partial V^m_j}{\partial h^m_j}\, \frac{\partial h^m_j}{\partial w^m_{jk}}$$

$$V_j^m = \theta(h_j^m)$$

$$h_j^m = \sum_{k=1} V_k^{m-1} \, . w_{jk}^m$$

RETROPROPAGACIÓN: CAPA OCULTA (*M-1*)



$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial w_{jk}^m} = \frac{\partial E}{\partial O} \frac{\partial O}{\partial h} \frac{\partial h}{\partial V_j^m} \frac{\partial V_j^m}{\partial h_j^m} \frac{\partial h_j^m}{\partial w_{jk}^m}$$

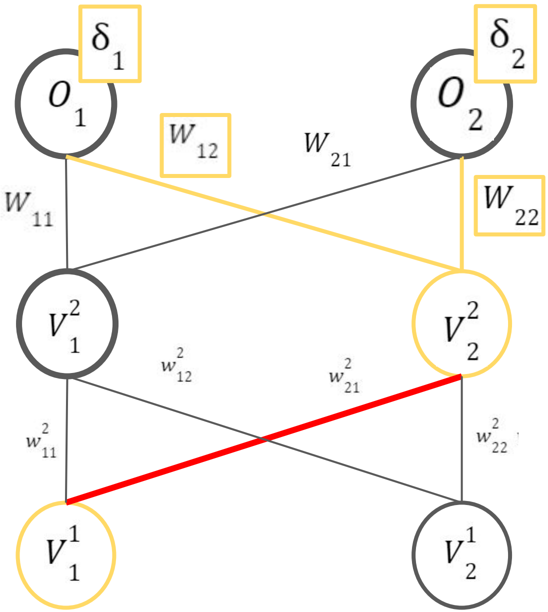
$$\frac{\partial E}{\partial w_{jk}^m} = -\sum_i \delta_i W_{ij} \boxed{\frac{\partial V_j^m}{\partial h_j^m}} \boxed{\frac{\partial h_j^m}{\partial w_{jk}^m}}$$

$$\frac{\partial E}{\partial w_{jk}^m} = -\sum_i \delta_i W_{ij} \; \theta'(h_j^m)(1) \; V_k^{m-1}$$

RETROPROPAGACIÓN: CAPA OCULTA (*M-1*)

$$V_j^m = \theta(h_j^m)$$

$$h_j^m = \sum_{k=1} V_k^{m-1} \, . \, w_{jk}^m$$



$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial w_{jk}^m} = \frac{\partial E}{\partial O}\,\frac{\partial O}{\partial h}\,\frac{\partial h}{\partial V_j^m}\,\frac{\partial V_j^m}{\partial h_j^m}\,\frac{\partial h_j^m}{\partial w_{jk}^m}$$

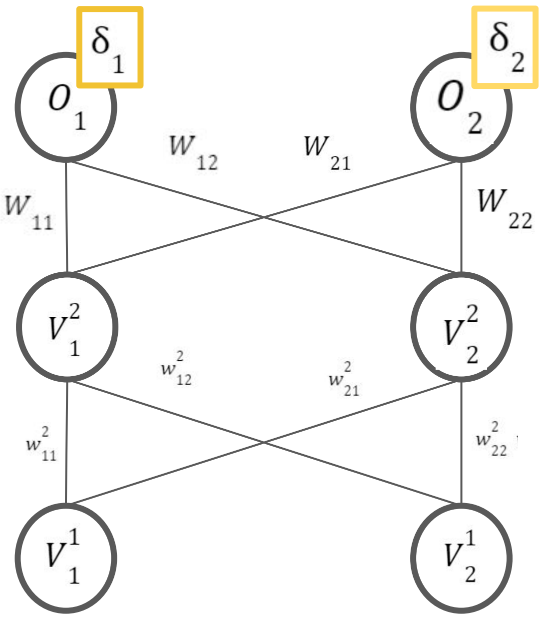
$$\frac{\partial E}{\partial w_{jk}^m} = -\sum_i \delta_i W_{ij} \boxed{\frac{\partial V_j^m}{\partial h_j^m}} \boxed{\frac{\partial h_j^m}{\partial w_{jk}^m}}$$

$$\frac{\partial E}{\partial w_{jk}^m} = -\sum_i \delta_i W_{ij} \; \theta'(h_j^m)(1) \; V_k^{m-1}$$

$$V_j^m = \theta(h_j^m)$$

$$h_j^m = \sum_{k=1} V_k^{m-1} . w_{jk}^m$$

RETROPROPAGACIÓN: CAPA OCULTA (*M-1*)



$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial w_{jk}^m} = \frac{\partial E}{\partial O} \frac{\partial O}{\partial h} \frac{\partial h}{\partial V_j^m} \frac{\partial V_j^m}{\partial h_j^m} \frac{\partial h_j^m}{\partial w_{jk}^m}$$

$$\frac{\partial E}{\partial w_{jk}^m} = -\sum_i \delta_i W_{ij} \boxed{\frac{\partial V_j^m}{\partial h_j^m}} \boxed{\frac{\partial h_j^m}{\partial w_{jk}^m}}$$

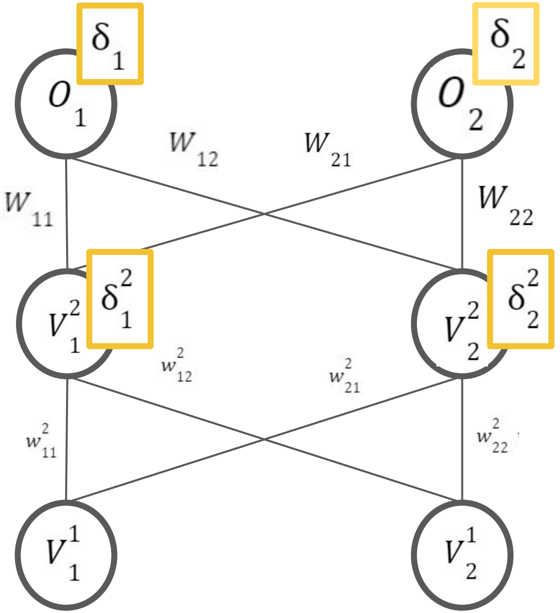
$$\frac{\partial E}{\partial w_{jk}^m} = -\underbrace{\sum_i \delta_i W_{ij} \, \theta'(h_j^m)(1)}_{\delta_j^m} V_k^{m-1}$$

$$\Delta w = \eta \delta_j^m V_k^{m-1}$$

$$V_j^m = \theta(h_j^m)$$

$$h_j^m = \sum_{k=1} V_k^{m-1} \cdot w_{jk}^m$$

RETROPROPAGACIÓN: CAPA OCULTA (*M-1*)



$$\Delta w = -\eta \boxed{\frac{\partial E}{\partial w}} \longrightarrow \frac{\partial E}{\partial w_{jk}^m} = \frac{\partial E}{\partial O} \frac{\partial O}{\partial h} \frac{\partial h}{\partial V_j^m} \frac{\partial V_j^m}{\partial h_j^m} \frac{\partial h_j^m}{\partial w_{jk}^m}$$

$$\frac{\partial E}{\partial w_{jk}^m} = -\sum_i \delta_i W_{ij} \boxed{\frac{\partial V_j^m}{\partial h_j^m}} \boxed{\frac{\partial h_j^m}{\partial w_{jk}^m}}$$

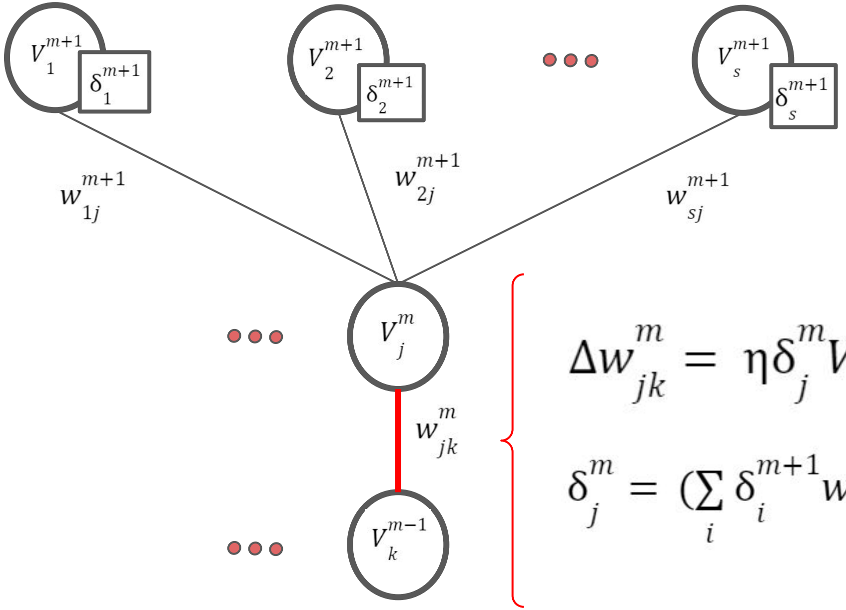
$$\frac{\partial E}{\partial w_{jk}^m} = -\underbrace{\sum_i \delta_i W_{ij}\, \theta'(h_j^m)(1)}_{\delta_j^m} V_k^{m-1}$$

$$\Delta w = \eta \delta_j^m V_k^{m-1}$$

$$V_j^m = \theta(h_j^m)$$

$$h_j^m = \sum_{k=1} V_k^{m-1} \,.\, w_{jk}^m$$

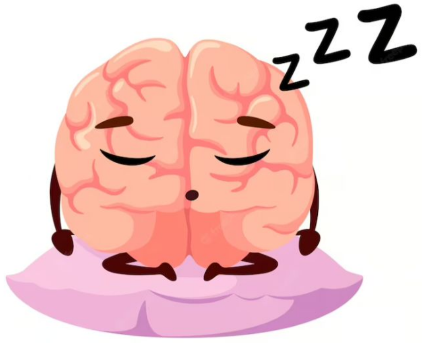
RETROPROPAGACIÓN: CAPA OCULTA



$$\Delta w_{jk}^m = \eta \delta_j^m V_k^{m-1}$$

$$\delta_j^m = \left(\sum_i \delta_i^{m+1} w_{ij}^{m+1}\right)\theta'(h_j^m)$$

	PERCEPTRÓN SIMPLE	PERCEPTRÓN MULTICAPA
Proceso para obtener la salida de la neurona ("predecir")	$O = \theta(\sum_{i=0}^{n} x_i . w_i)$	$V_j^m = \theta(\sum_{k=1} V_k^{m-1} . w_{mk})$ $m = 1 \ldots M \ (O_i = V_i^M ,\ x_i = V_i^0)$
Función de error con respecto a la salida esperada	$E(O) = \frac{1}{2}(\zeta^\mu - O^\mu)^2$	$E(O) = \frac{1}{2}\sum_i (\zeta^\mu - O_i^\mu)^2$
Proceso de "aprendizaje" para ajustar los pesos	$w^{nuevo} = w^{anterior} + \Delta w$ $\Delta w = -\eta\frac{\partial E}{\partial w}$	$\Delta W_{ij} = \eta\delta_i V_j$ $\quad \delta_i = (\zeta_i - O_i)\theta'(h_i)$ $\Delta w_{jk}^m = \eta\delta_j^m V_k^{m-1}$ $\quad \delta_j^m = (\sum_i \delta_i^{m+1} w_{ij}^{m+1})\theta'(h_j^m)$



ALGORITMO DEL PERCEPTRÓN MULTICAPA

1. Definir la arquitectura del perceptrón multicapa
    a. Capas y cantidad de neuronas (*no tenemos un método para conseguir la ideal*)
    b. Funciones de activación (*puede ser la misma para todas las neuronas o no*)

2. Inicializar los pesos sinápticos en valores aleatorios pequeños o cero

3. Definir: método de optimización, tasa de aprendizaje, épocas.

4. Para cada elemento del conjunto de datos

    a. Calcular la salida de la neurona (*feedforward*)

    b. Actualizar los pesos sinápticos (*backpropagation*)

5. Calcular el error del perceptrón para verificar si se alcanzó convergencia (MSE)
    a. Si el perceptrón alcanzó convergencia, finalizar.

6. Repetir 3 y 4 hasta alcanzar convergencia o hasta finalizar la cantidad de épocas

ALGORITMO DEL PERCEPTRÓN MULTICAPA

Para cada elemento del conjunto de datos

a. Calcular la salida de la neurona (*feedforward*) (***M*** es la cantidad de capas)

$$V^0 = x^\mu$$

$$V_j^m = \theta\left(\sum_{k=1} V_k^{m-1} \,.\, w_{mk}\right) \qquad m = 1 \ldots M - 1$$

$$O_i = \theta\left(\sum_j V_j^{M-1} W_{ij}\right)$$

ALGORITMO DEL PERCEPTRÓN MULTICAPA

Para cada elemento del conjunto de datos

b. Actualizar los pesos sinápticos (*backpropagation*)

Calculamos los deltas para cada capa

$$\delta_i = (\zeta_i - O_i)\theta'(h_i) \qquad \textit{capa salida}$$

$$\delta_j^m = (\sum_i \delta_i^{m+1} w_{ij}^{m+1})\theta'(h_j^m) \qquad \textit{capa oculta}$$

Calculamos la actualización correspondiente para los pesos

$$\Delta W_{ij} = \eta \delta_i V_j^{M-1} \qquad \textit{capa salida}$$

$$\Delta w_{jk}^m = \eta \delta_j^m V_k^{m-1} \qquad \textit{capa oculta}$$

RESUMEN

- El perceptrón multicapa me permite modelar transformaciones complejas (expande el comportamiento de los perceptrones simples)

- La arquitectura del perceptŕon multicapa debe realizarse manualmente. A priori, no tenemos una receta que nos permita definir “la mejor arquitectura”.

- El algoritmo de retropropagación provee un mecanismo para hallar las actualizaciones de los pesos que nos permiten minimizar la función de costo.

ALGUNAS VARIACIONES POSIBLES

Variable	Posibles Valores
Función de activación (Θ)	Escalón, Identidad, Sigmoidea
Tasa de aprendizaje (η)	~0.1
Actualización de los pesos	Batch/Online
Error del perceptrón (función de costo)	Accuracy, suma valores absolutos, MSE
Épocas	
Método de optimización	Gradiente Descendente
Parámetros de función de activación	Θ de tanh o de log
Técnica para separar en entrenamiento y testeo	Ejemplo: 80%-20%
Arquitectura de la red	