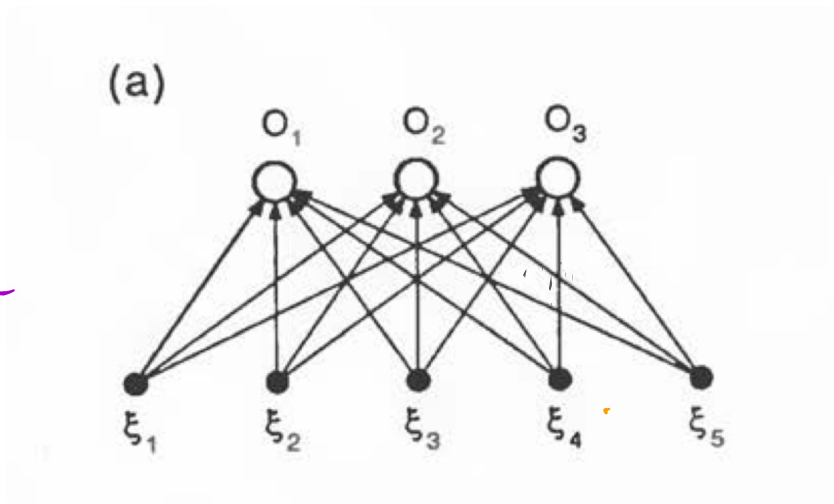
APRENDIZAJE SUPERVIZADO

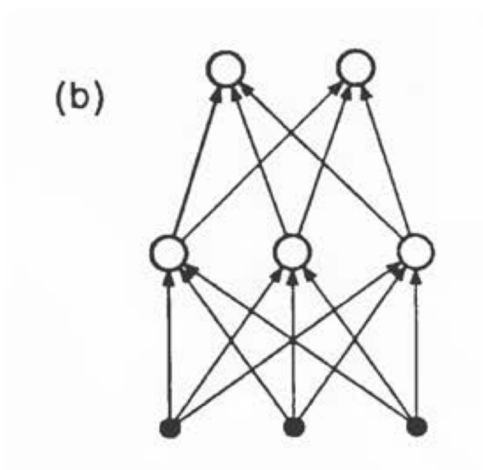
ARQUITECTURAS FEED-FORWARD

1 capa



Perceptron simple

2 capas



salida

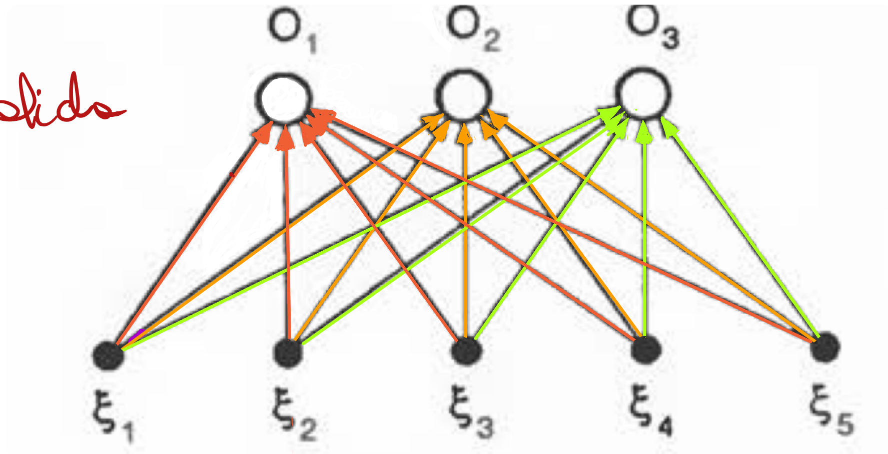
oculta

entrada

Perceptron multi capas

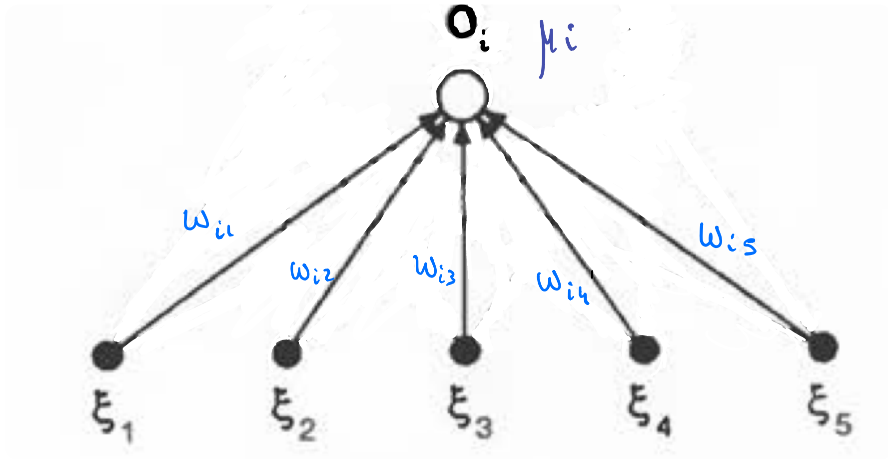
EL PERCEPTRON SIMPLE

capa salida



entrada

Consideremos una arquitectura feed-forward como la de la figura. Observen que tenemos 3 redes desacopladas con una única salida cada una, pues no comparten ninguna sinapsis. El problema entonces puede desacoplarse. Esto no sucede si la red tiene capas ocultas o si es recurrente.



1 neurona de salida O_i

N entradas ξ_k

$(k = 1, 2, \ldots, N)$

El conjunto $\{\xi_1, \xi_2, \ldots, \xi_N\}$ puede pensarse como un vector en $\mathbb{R}^N$

$$\bar{\xi} = (\xi_1, \xi_2, \xi_3, \ldots, \xi_N) \in \mathbb{R}^N$$

Las variables ξ_k pueden tomar valores discretos o continuos (acotados o no).

El rango de valores que puede tomar O_i estará dado por la función de activación $g(x)$

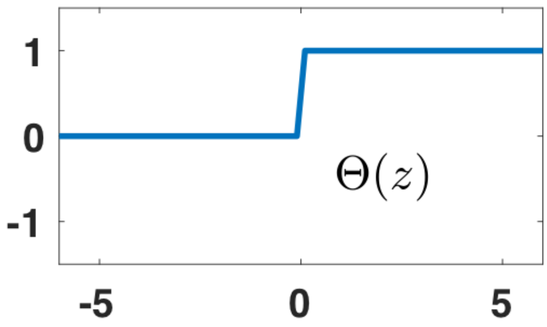
$$O_i = g\left(h_i - \mu_i\right)$$

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

w_{ik} : eficacia sináptica entre entrada k y neurona i

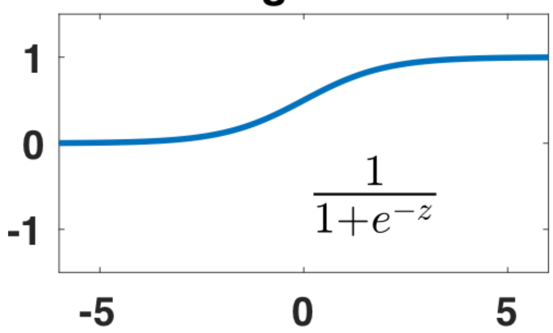
μ_i : umbral de activación de la neurona i

Perceptron



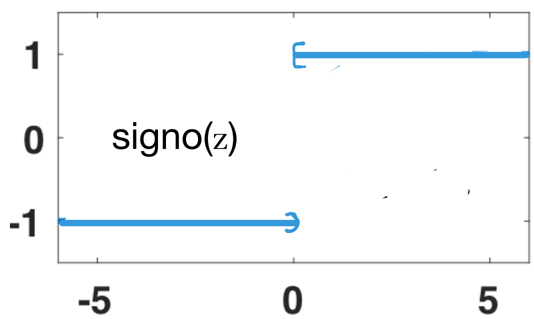
$O_i \in \{0, 1\}$

Sigmoid



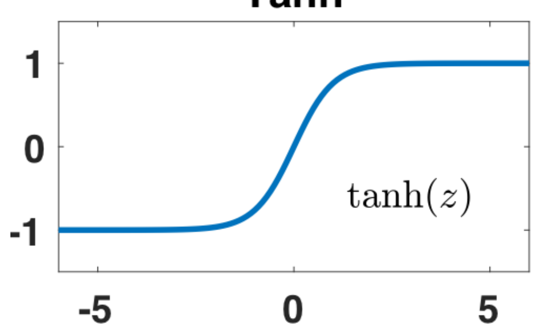
$0 < O_i < 1$

Ising

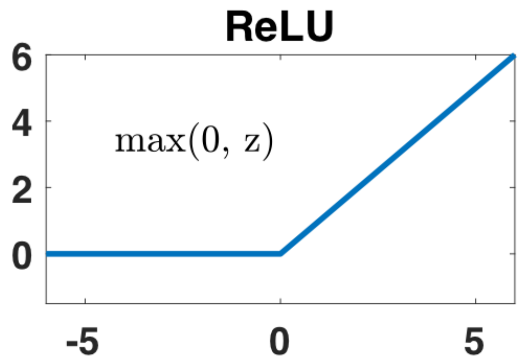


$O_i \in \{-1, +1\}$

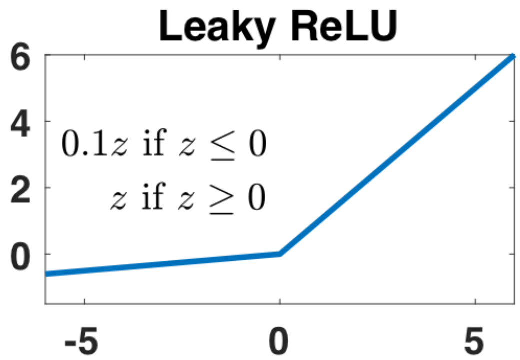
Tanh



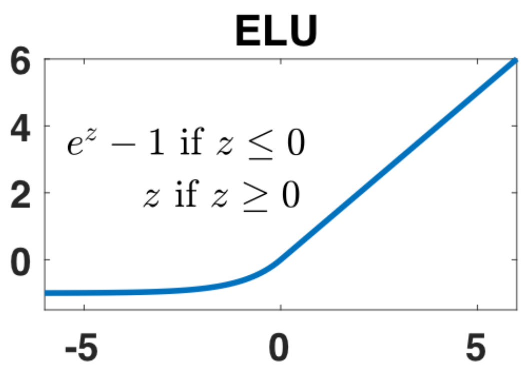
$-1 < O_i < +1$



$0 \leq \theta_i < +\infty$



$-\infty < \theta_i < +\infty$



$-1 < \theta_i < +\infty$

Aprendizaje supervisado

Supondremos que disponemos de p entradas denotadas

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

con $\mu = 1, 2, 3, \ldots, p$, a las cuales le conocemos la salida correcta, denotada $\mathcal{O}_i^{\mu}$.

$$\vec{\xi}^{\mu} \longrightarrow \zeta_i^{\mu} \in \mathbb{R}$$

El desafío consiste en encontrar el conjunto de acoplamientos $w_{i1}, w_{i2}, \ldots, w_{iN}$ que hagan correctamente esta asociación. también μ_i

Noten que los acoplamientos definen un vector

$$\vec{W}_i = (w_{i1}, w_{i2}, \ldots, w_{iN})$$

y que, al igual que $\bar{\xi}$, pertenece a $\mathbb{R}^N$.

O sea, $\bar{\xi} \in \mathbb{R}^N$ y $\bar{w}_i \in \mathbb{R}^N$.

(suponemos M neuronas de salida O_i, $i=1,2,...M$)

Resumiendo, tenemos muchas cosas ya definidas:

$\bar{\xi}$: vector de entrada (N componentes)

$\bar{\xi}^\mu$: p entradas particulares (N×p componentes)

ζ_i^μ : salida correcta asociada a ξ_i^μ

O_i^μ : salida real asociada a ξ_i^μ

μ_i : umbral de activación asociado a la neurona i (salida).

$\bar{w}_i$: conjunto de acoplamientos sinápticos. (M×N)

$$\vec{\xi},\ \vec{\xi}^{\mu},\ \vec{w}_i \in \mathbb{R}^N$$

$$\zeta_i^{\mu},\ O_i^{\mu},\ \mu_i \in \mathbb{R}$$

con $\mu = 1, 2, \ldots, p$.

Como deshacernos del molesto μ_i

Vamos a imaginar que en la capa de entrada tenemos una neurona extra a la cual le asignamos el subíndice $k=0$, y además le asignamos el valor fijo $\xi_0 = -1$, cualquiera sea la entrada.

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

$$= \sum_{k=1}^{N} w_{ik}\xi_k + \mu_i \xi_0$$

$$h_i - \mu_i = \sum_{k=0}^{N} W_{ik}\, \xi_k$$

si hacemos $W_{i0} = \mu_i$.

Con esto nos podemos olvidar del trato diferenciado que requiere μ_i.

Ahora, para nosotros APRENDER significa encontrar un conjunto de acoplamientos

$$\overline{W} = (W_{i0}, W_{i1}, \dots, W_{iN})$$

tales que, ante cualquier entrada ξ^μ del conjunto de entrenamiento, la red de la salida correcta:

$$O_i^\mu = g(h_i^\mu) = g\left(\sum_{k=0}^{N} W_{ik}\, \xi_k^\mu\right) = \zeta_i^\mu$$

$$O_i^\mu = \zeta_i^\mu$$ (la salida es la deseada)

No siempre podemos aprender "perfectamente", y encontraremos maneras eficientes de medir cuan cerca estamos de una solución "perfecta".