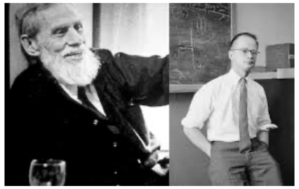
Universidad Nacional de General Sarmiento

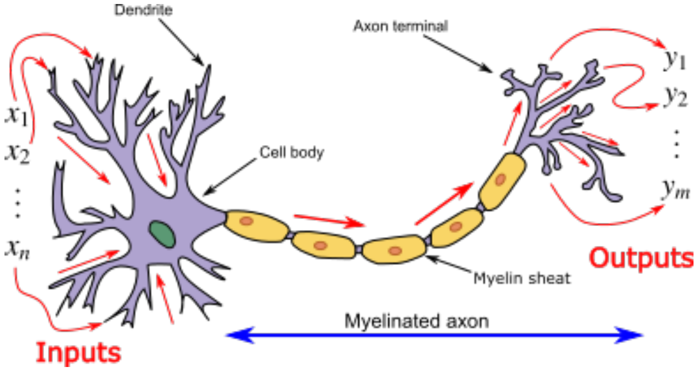
Licenciatura en Sistemas

Taller de Tesina

1943 – Warren McCulloch y Walter Pitts.
Modelo matemático de la neurona

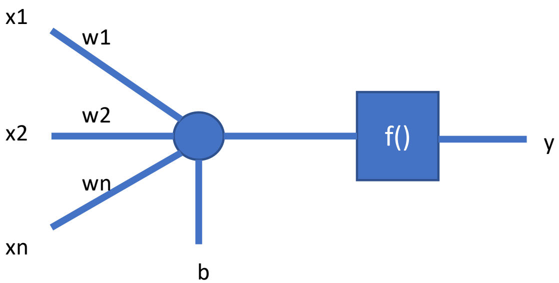


“Because of the all-or-none character of nervous activity, neural events and the relations among them can be treated by means of propositional logic”

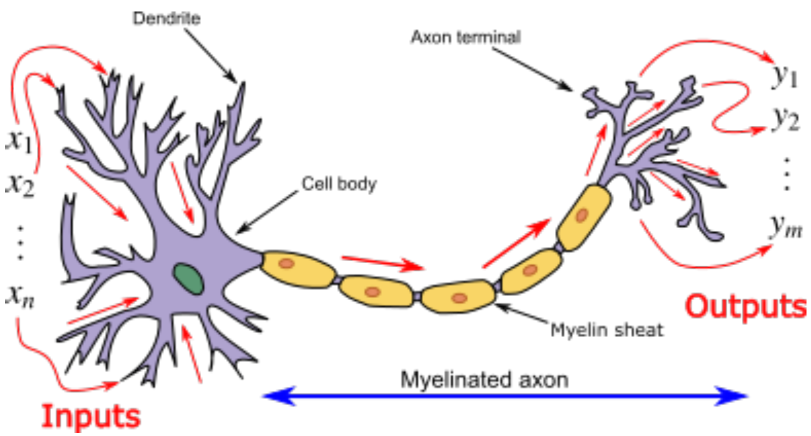


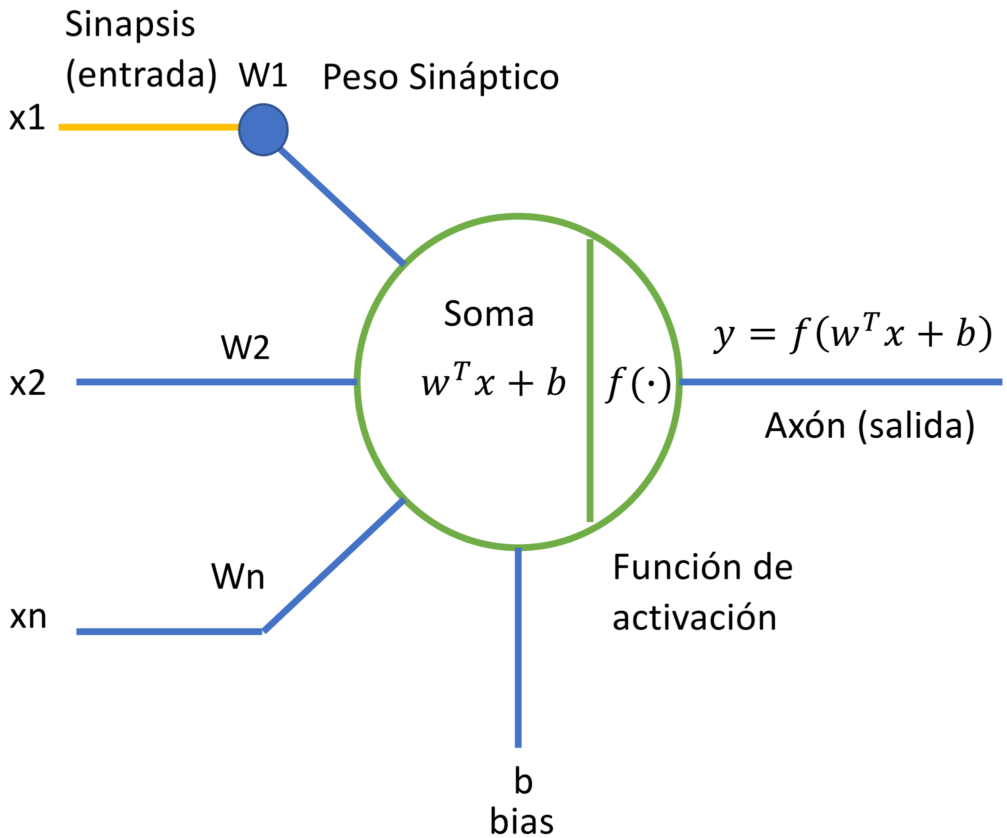
$$y = f\left(\sum_j w_j\, x_j + b\right)$$

Donde f() es una función de activación:



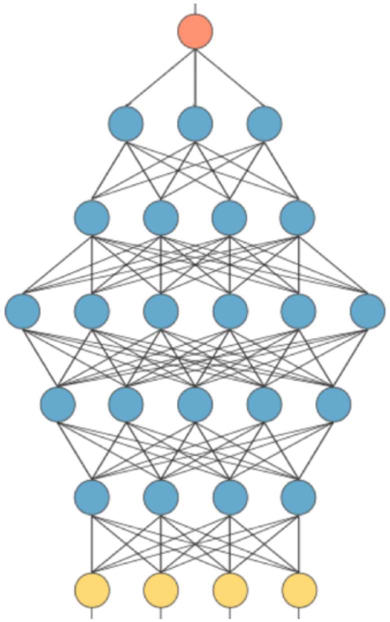
$$\mathrm{f}(v) = \begin{cases} 1\ si\ v > \theta \\ 0\ si\ v \le \theta \end{cases}$$

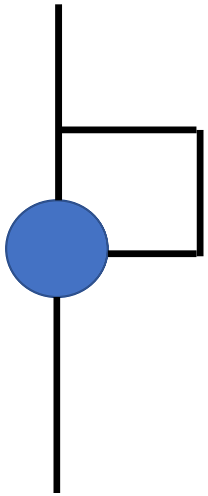


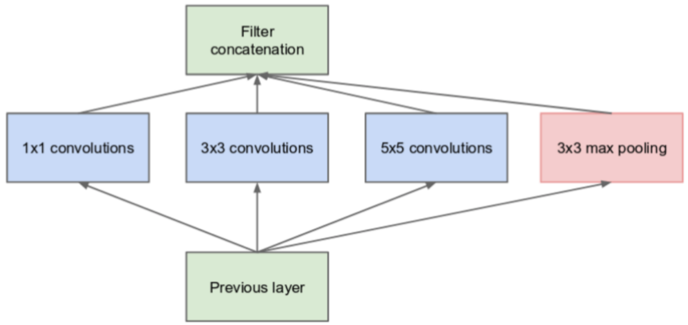


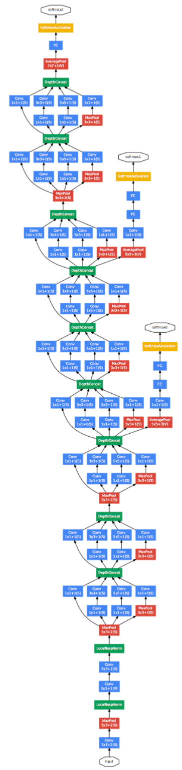
- Es un modelo matemático
- Las sinapsis biológicas son sistemas dinámicos mucho más complejos
- Lo mismo sucede en el soma neuronal.
- La función de activación emula al potencial de acción neuronal, no obstante hay transmisión de información entre neuronas biológicas sin potenciales de acción.

Estructuras de Redes



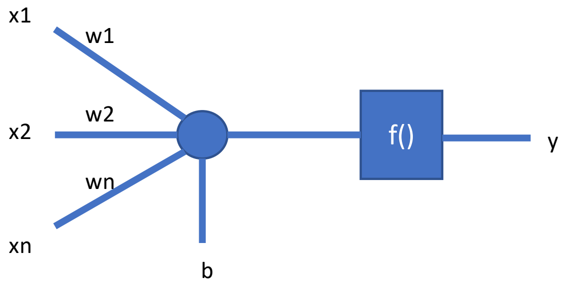






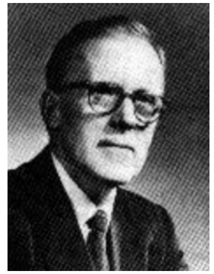
$$y = f(w^T x + b)$$

¿Cómo definir W?



1949 – Donald Hebb : Aprendizaje Asociativo

No Supervisado



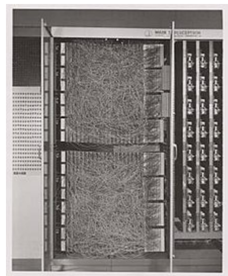
$$w_i = y x_i \qquad w_i(n+1) = w_i(n) + \Delta w_i$$

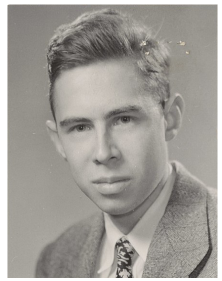
$$\Delta w_i = \eta y x_i$$

1962 – Frank Rosemblatt : Perceptrón.

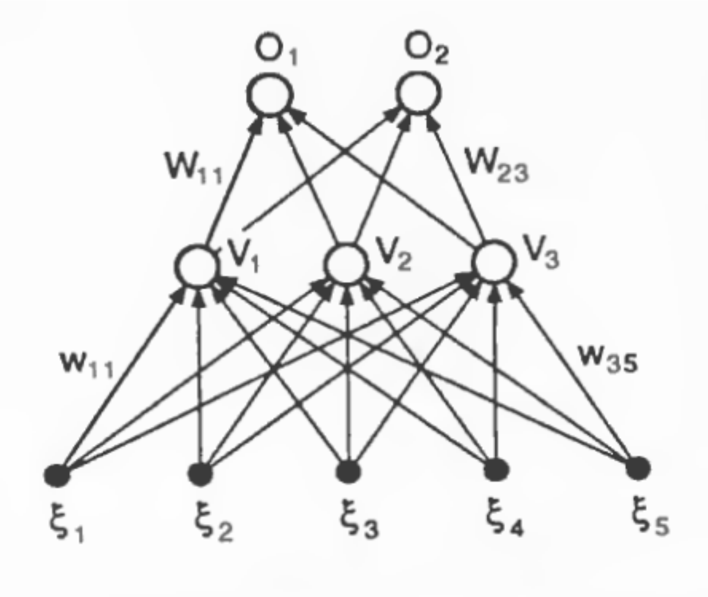
Supervisado

$$w_j(n+1) = w_j(n) + \rho(l_k - y_i) x_{j,k}$$





Forward - Propagation



Datos de entrada y etiquetas de entrenamiento

$$\xi^\mu = [\xi_1, \xi_2, \cdots, \xi_k, \cdots]^\mu \leftrightarrow \lambda^\mu = [\lambda_1, \lambda_2]^\mu$$

Salida Nivel oculto

$$V_j = f\left(\sum_k w_{jk}\xi_k\right)$$

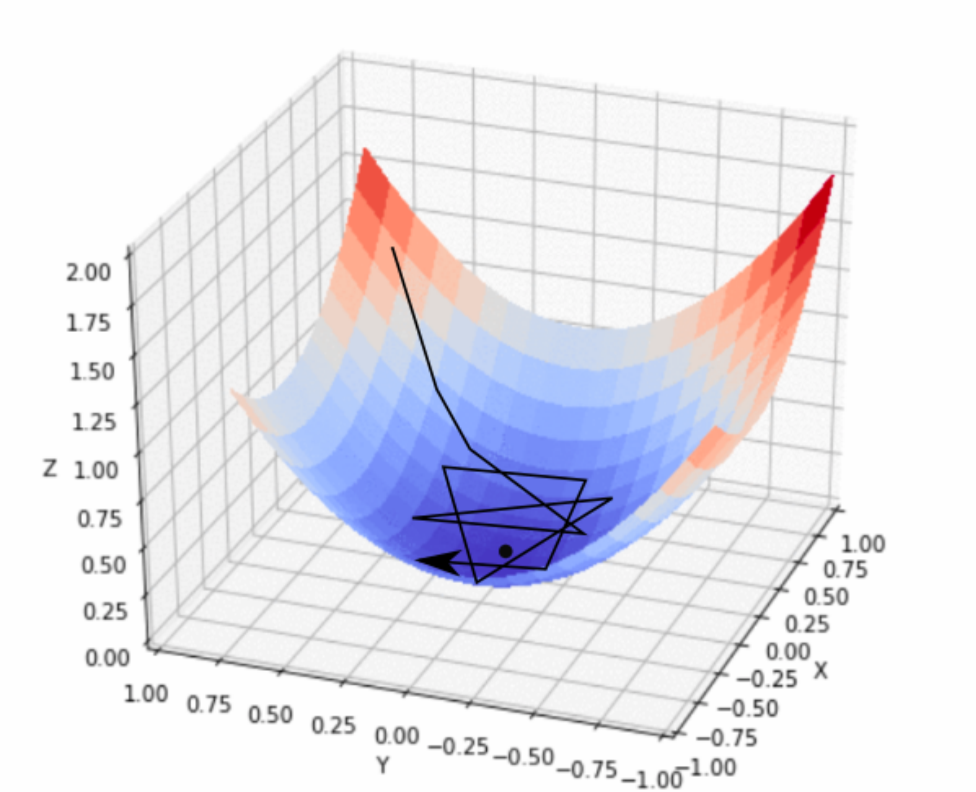
Salida de la Red

$$O_i = f\left(\sum_j W_{ij} f\left(\sum_k w_{jk}\xi_k\right)\right)$$

Error cuadrático

$$E(w) = \frac{1}{2}\sum_{\mu i}\left\|\lambda_i^\mu - O_i^\mu\right\|^2$$

Gradiente descendente



$$E(w) = \frac{1}{2}\sum_{i\mu}\left\|\lambda_i^{\mu} - o_i^{\mu}\right\|^2$$

$$\Delta W_{ij} = -\eta\frac{\partial E}{\partial W_{ij}} \qquad \Delta w_{jk} = -\eta\frac{\partial E}{\partial w_{jk}}$$

$$W_{ij}(n+1) = W_{ij}(n) + \Delta W_{ij}$$

$$w_{jk}(n+1) = w_{jk}(n) + \Delta w_{jk}$$

$$E(w) = \frac{1}{2}\sum_{i\mu}\left\|\lambda_i^{\mu} - O_i^{\mu}\right\|^2 \qquad O_i = f\left(\sum_j W_{ij}V_j\right) \qquad V_j = f\left(\sum_k w_{jk}\xi_k\right)$$

$$\Delta W_{ij} = -\eta\frac{\partial E}{\partial W_{ij}} \qquad \Delta w_{jk} = -\eta\frac{\partial E}{\partial w_{jk}} = -\eta\sum_{\mu}\frac{\partial E}{\partial V_j}\frac{\partial V_j}{\partial w_{jk}}$$

$$\Delta W_{ij} = \eta\sum_{\mu}(\lambda_i^{\mu} - O_i^{\mu})f'(\cdot)V_j \qquad \Delta w_{jk} = \eta\sum_{\mu}\sum_{i}(\lambda_i^{\mu} - O_i^{\mu})f'(\cdot_i)W_{ij}f'(\cdot_j)\xi_j$$

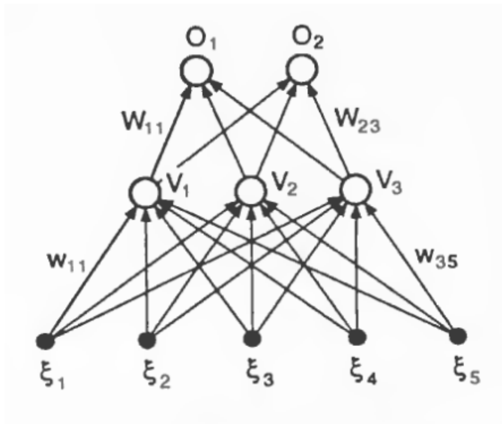
$$E(w) = \frac{1}{2}\sum_{i\mu} \left\| \lambda_i^\mu - O_i^\mu \right\|^2 \qquad O_i = f\left(\sum_j W_{ij} V_j\right) \qquad V_j = f\left(\textstyle\sum_k w_{jk}\xi_k\right)$$

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}} \qquad \Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} = -\eta \sum_\mu \boxed{\frac{\partial E}{\partial V_j}} \frac{\partial V_j}{\partial w_{jk}}$$

$$\Delta W_{ij} = \eta \sum_\mu (\lambda_i^\mu - O_i^\mu) f'(\cdot) V_j \longrightarrow \Delta w_{jk} = \eta \sum_\mu \boxed{\sum_i (\lambda_i^\mu - O_i^\mu) f'(\cdot_i) W_{ij}} f'(\cdot_j)\xi_j$$

$$\Delta W_{ij} = \eta \sum_\mu (\lambda_i^\mu - O_i^\mu) f'(\cdot_i) V_j$$

$$\Delta w_{jk} = \eta \sum_\mu f'(\cdot_j)\xi_j \sum_i (\lambda_i^\mu - O_i^\mu) f'(\cdot_i) W_{ij}$$



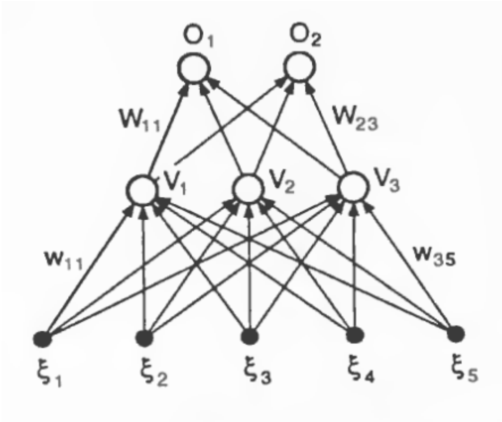
$$E(w) = \frac{1}{2}\sum_{i\mu}\left\|\lambda_i^\mu - O_i^\mu\right\|^2 \qquad O_i = f\left(\sum_j W_{ij}V_j\right) \qquad V_j = f\left(\sum_k w_{jk}\xi_k\right)$$

$$\Delta W_{ij} = -\eta\frac{\partial E}{\partial W_{ij}} \qquad \Delta w_{jk} = -\eta\frac{\partial E}{\partial w_{jk}} = -\eta\sum_\mu \frac{\partial E}{\partial V_j}\frac{\partial V_j}{\partial w_{jk}}$$

$$\Delta W_{ij} = \eta\sum_\mu (\lambda_i^\mu - O_i^\mu) f'(\cdot)V_j \qquad \Delta w_{jk} = \eta\sum_\mu\sum_i (\lambda_i^\mu - O_i^\mu) f'(\cdot_i) W_{ij} f'(\cdot_j)\xi_j$$

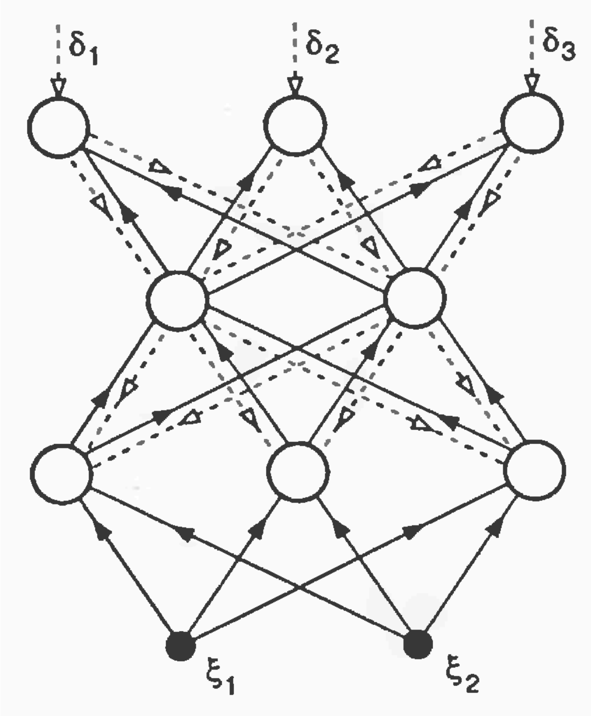
$$\Delta W_{ij} = \eta\sum_\mu \boxed{(\lambda_i^\mu - O_i^\mu) f'(\cdot_i)}V_j$$

$$\Delta w_{jk} = \eta\sum_\mu f'(\cdot_j)\xi_j\sum_i \boxed{(\lambda_i^\mu - O_i^\mu) f'(\cdot_i)}W_{ij}$$



$$\Delta W_{ij} = \eta \sum_{\mu} \boxed{(\lambda_i^{\mu} - O_i^{\mu}) f'(\cdot_i)} V_j$$

$$\Delta w_{jk} = \eta \sum_{\mu} f'(\cdot_j) \xi_k \sum_{i} \boxed{(\lambda_i^{\mu} - O_i^{\mu}) f'(\cdot_i)} W_{ij}$$



$$\delta_i = (\lambda_i^{\mu} - O_i^{\mu}) f'(\cdot_i)$$

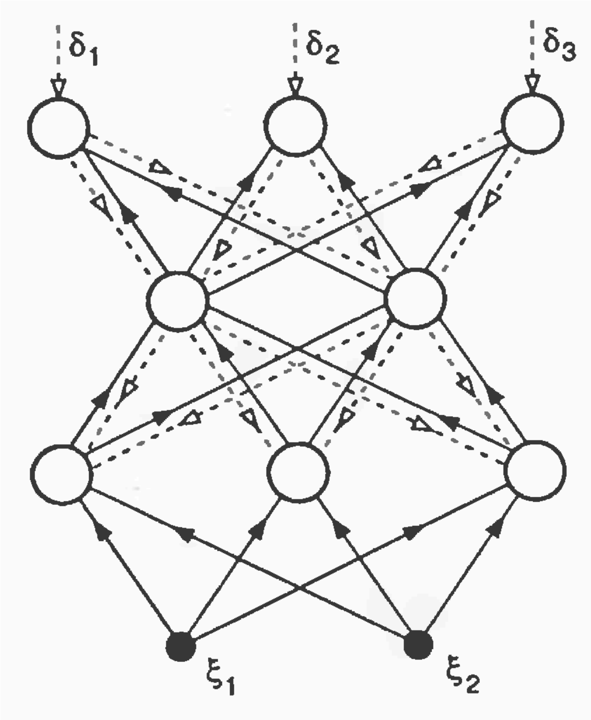
$$\delta_j = f'(\cdot_j) \sum_{i} \delta_i W_{ij}$$

$$\Delta W_{ij} = \eta \sum_{\mu} \delta_i V_j$$

$$\Delta w_{jk} = \eta \sum_{\mu} \delta_j \xi_k$$

Back - Propagation

1986 – David Rumelhart: Back-Propagation aplicado a redes de neuronas artificiales



$$\delta_i = (\lambda_i^\mu - O_i^\mu) f'(\cdot_i)$$

$$\delta_j = f'(\cdot_j) \sum_i \delta_i W_{ij}$$

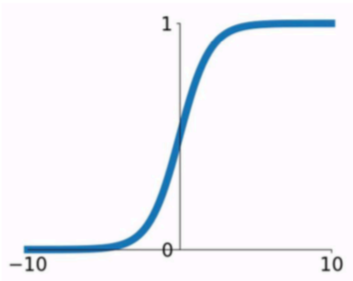
$$\Delta W_{ij} = \eta \sum_\mu \delta_i V_j$$

$$\Delta w_{jk} = \eta \sum_\mu \delta_j \xi_k$$

Activation functions $f: y = f(w^T x + b)$

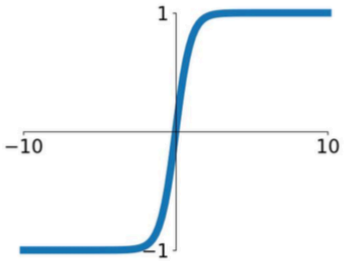
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



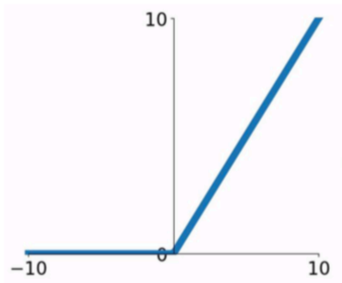
tanh

$$\tanh(x)$$



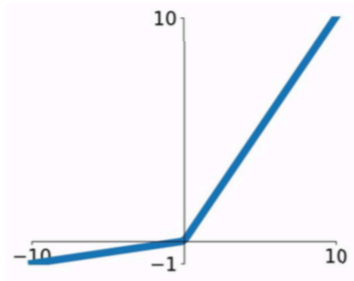
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

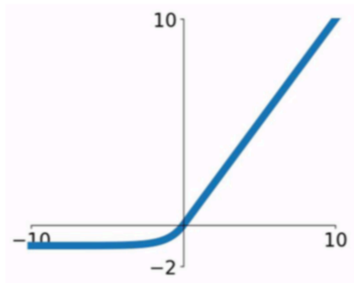


Maxout

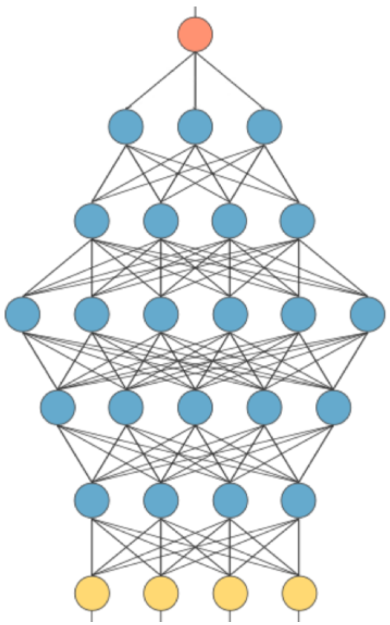
$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

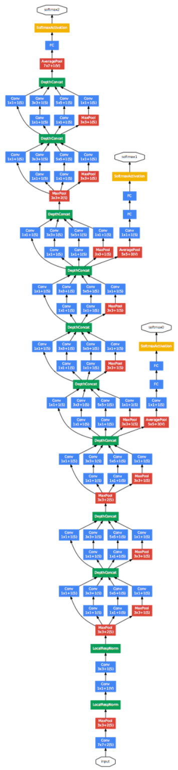
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Backpropagation y …



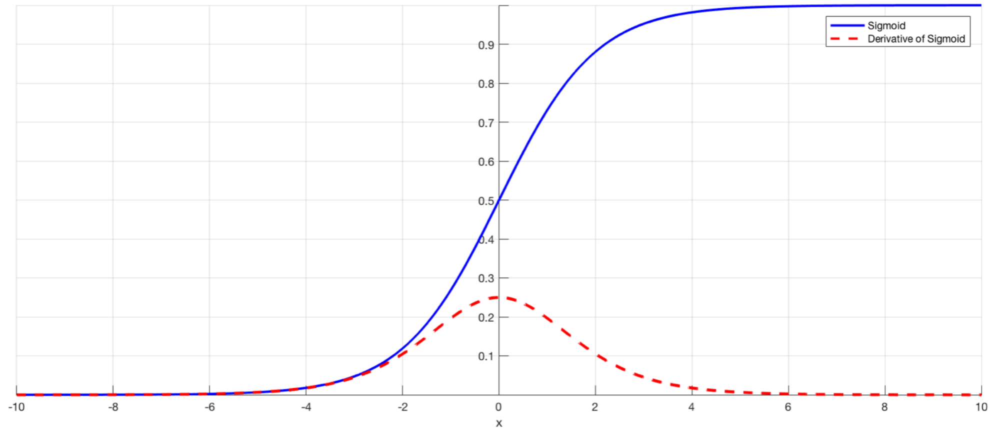
6 layers



30 layers

SIGMOIDEA

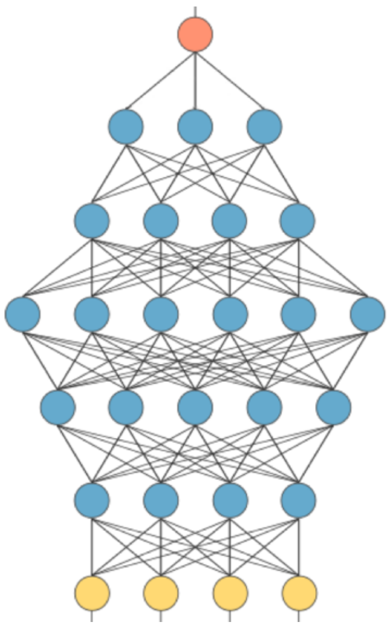
$$f(x) = \frac{1}{1 + e^{-x}} \qquad f'(x) = f(x) * (1 - f(x))$$



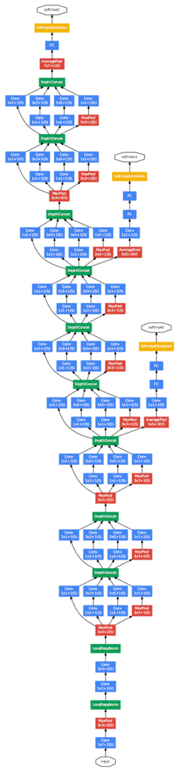
$$\delta_i = \left(\lambda_i^{\mu} - O_i^{\mu}\right) f'(\cdot_i)$$

$$\delta_j = f'(\cdot_j) \sum_i \delta_i W_{ij} = \sum_i \underline{f'(\cdot_j)\, f'(\cdot_i)}\, \left(\lambda_i^{\mu} - O_i^{\mu}\right) W_{ij}$$

Backpropagation y … ReLU – ELU las más usadas en Aprendizaje Profundo



6 layers

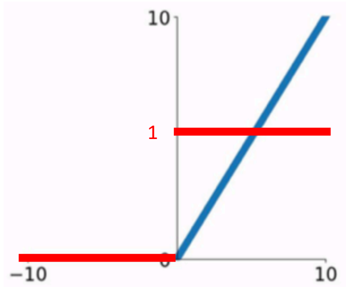


30 layers

$$f_{ReLU}(x) = \max(0, x) \qquad f'_{RELU}(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$$

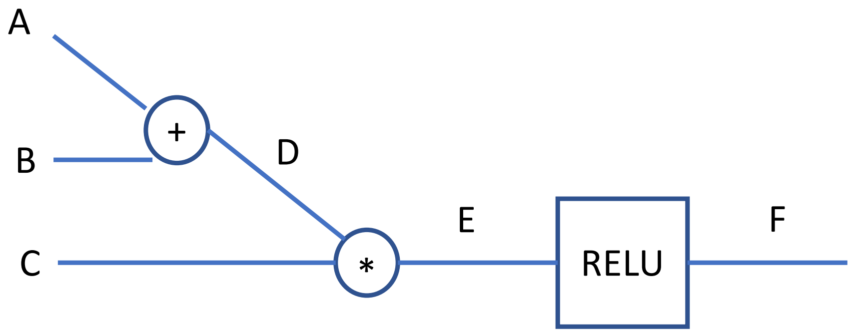
ReLU

$\max(0, x)$



$$\delta_i = (\lambda_i^\mu - O_i^\mu) f'(\cdot_i)$$

$$\delta_j = f'(\cdot_j) \sum_i \delta_i W_{ij} = \sum_i \underline{f'(\cdot_j)\, f'(\cdot_i)}\, (\lambda_i^\mu - O_i^\mu) W_{ij}$$

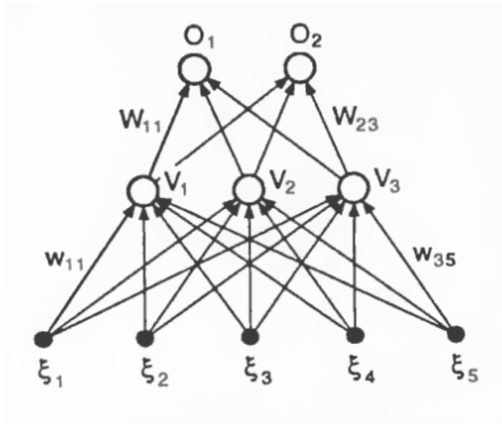


Calcular:

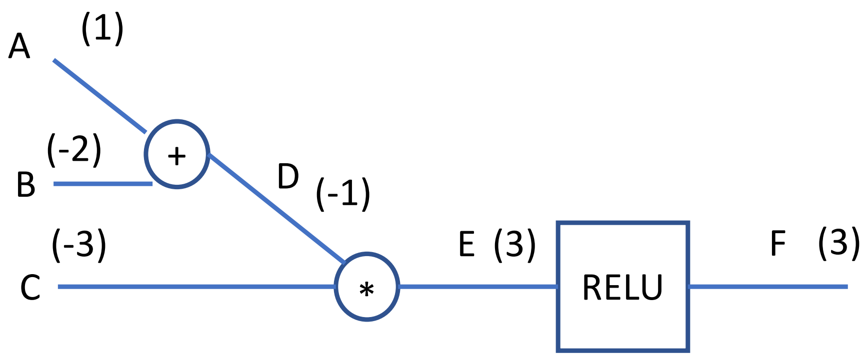
$$\frac{\partial F}{\partial A} \qquad \frac{\partial F}{\partial B} \qquad \frac{\partial F}{\partial C}$$

A=1 B=-2 C = -3

$$\Delta W_{ij} = \eta \sum_{\mu} \delta_i V_j \qquad \Delta w_{jk} = \eta \sum_{\mu} \delta_j \xi_k$$



$$\delta_i = (\lambda_i^{\mu} - O_i^{\mu}) f'(\cdot_i) \qquad \delta_j = f'(\cdot_j) \sum_i \delta_i W_{ij}$$



Calcular:

$$\frac{\partial F}{\partial A} \qquad \frac{\partial F}{\partial B} \qquad \frac{\partial F}{\partial C}$$

A=1 B=-2 C = -3

$$\frac{\partial F}{\partial E} = 1$$

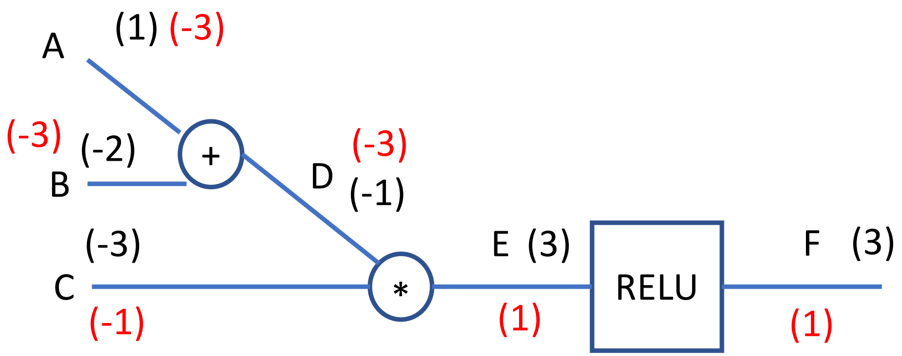
$$\frac{\partial E}{\partial C} = D \qquad \frac{\partial E}{\partial D} = C$$

$$\frac{\partial D}{\partial A} = 1 \qquad \frac{\partial D}{\partial B} = 1$$

$$\frac{\partial F}{\partial A} = \frac{\partial F}{\partial E}\frac{\partial E}{\partial D}\frac{\partial D}{\partial A}$$

$$\frac{\partial F}{\partial B} = \frac{\partial F}{\partial E}\frac{\partial E}{\partial D}\frac{\partial D}{\partial B}$$

$$\frac{\partial F}{\partial C} = \frac{\partial F}{\partial E}\frac{\partial E}{\partial C}$$



Calcular:

$$\frac{\partial F}{\partial A} \qquad \frac{\partial F}{\partial B} \qquad \frac{\partial F}{\partial C}$$

A=1 B=-2 C = -3

$$\frac{\partial F}{\partial E} = 1$$

$$\frac{\partial E}{\partial C} = D \qquad \frac{\partial E}{\partial D} = C$$

$$\frac{\partial D}{\partial A} = 1 \qquad \frac{\partial D}{\partial B} = 1$$

$$\frac{\partial F}{\partial A} = \frac{\partial F}{\partial E}\frac{\partial E}{\partial D}\frac{\partial D}{\partial A} = \text{C} = -3$$

$$\frac{\partial F}{\partial B} = \frac{\partial F}{\partial E}\frac{\partial E}{\partial D}\frac{\partial D}{\partial B} = \text{C} = -3$$

$$\frac{\partial F}{\partial C} = \frac{\partial F}{\partial E}\frac{\partial E}{\partial C} = \text{D} = -1$$

Gracias