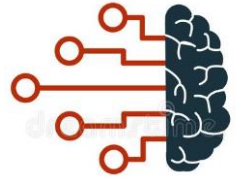


# Aprendizaje Supervisado

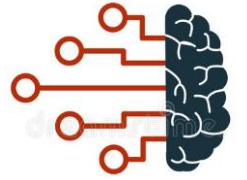
# Deep Learning

# Contenido

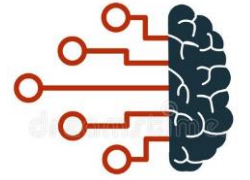


1. Definiciones
2. Función de activación
3. Red Neuronal
4. Tipos de redes neuronales
5. Problemas comunes

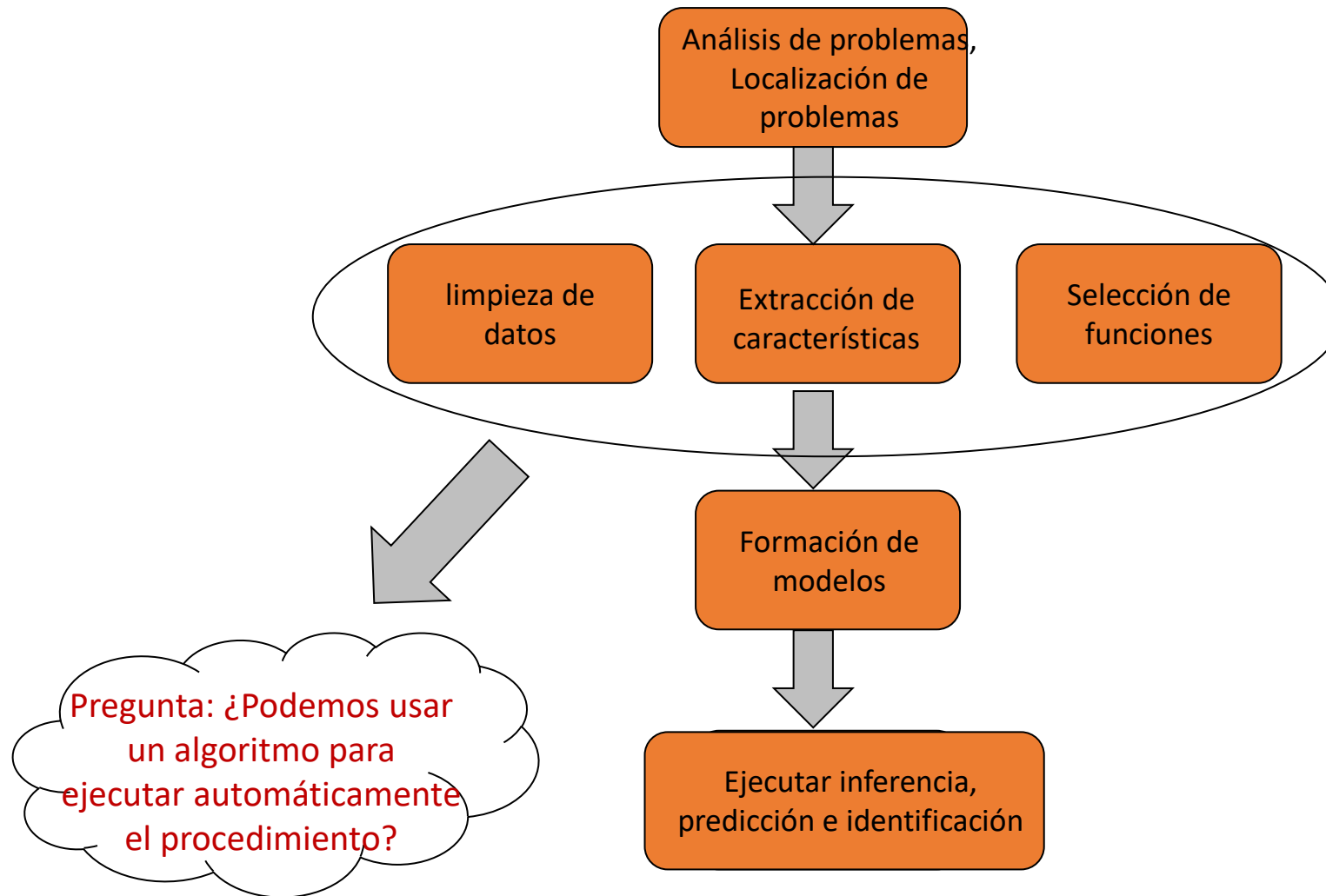
# Definición

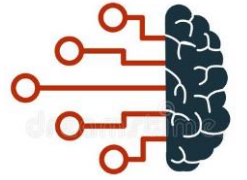


Machine Learning tradicional	Aprendizaje profundo
Bajos requisitos de hardware en el equipo: Dada la cantidad de computación limitada, el equipo no necesita una GPU para computación paralela	Mayores requisitos de hardware en el ordenador: Para ejecutar operaciones matriciales en datos masivos, el ordenador necesita una GPU para realizar computación paralela.
Aplicable a entrenamiento con una pequeña cantidad de datos y cuyo rendimiento no se puede mejorar continuamente a medida que aumenta la cantidad de datos.	El rendimiento puede ser alto cuando se proporcionan parámetros de peso de alta dimensión y datos de entrenamiento masivos.
Desglose por niveles del problema	Aprendizaje de extremo a extremo (E2E)
Selección manual de funciones	Extracción automática de funciones basada en algoritmos
Funciones fáciles de explicar	Funciones difíciles de explicar



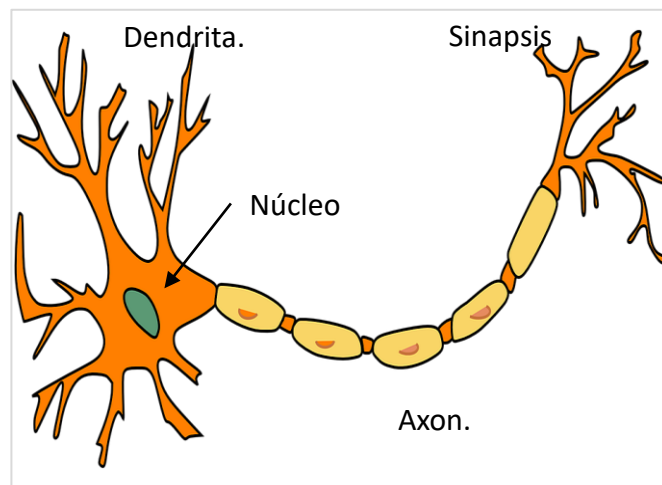
# Aprendizaje automático tradicional



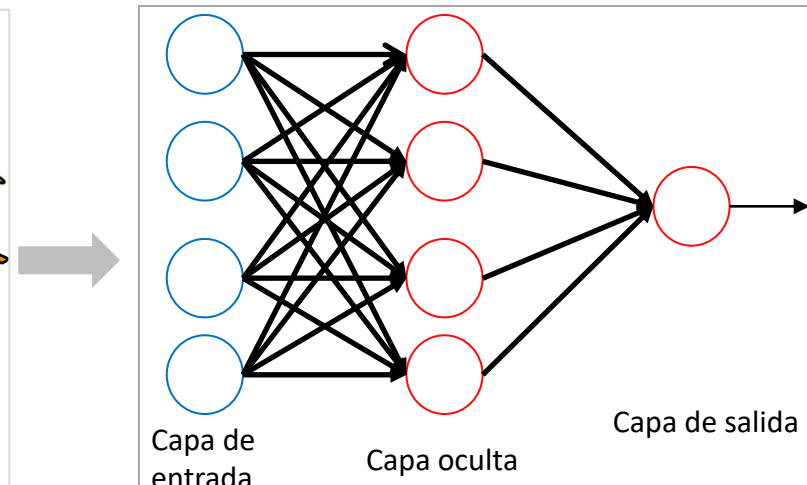


# Aprendizaje profundo

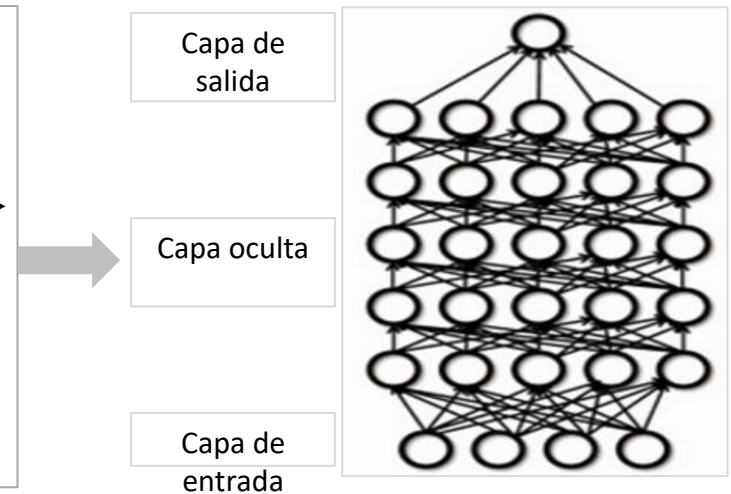
Generalmente, la arquitectura de aprendizaje profundo es una red neural artificial profunda.  
"Profundo" en "aprendizaje " se refiere al número de capas de la red neural



Red neuronal humana

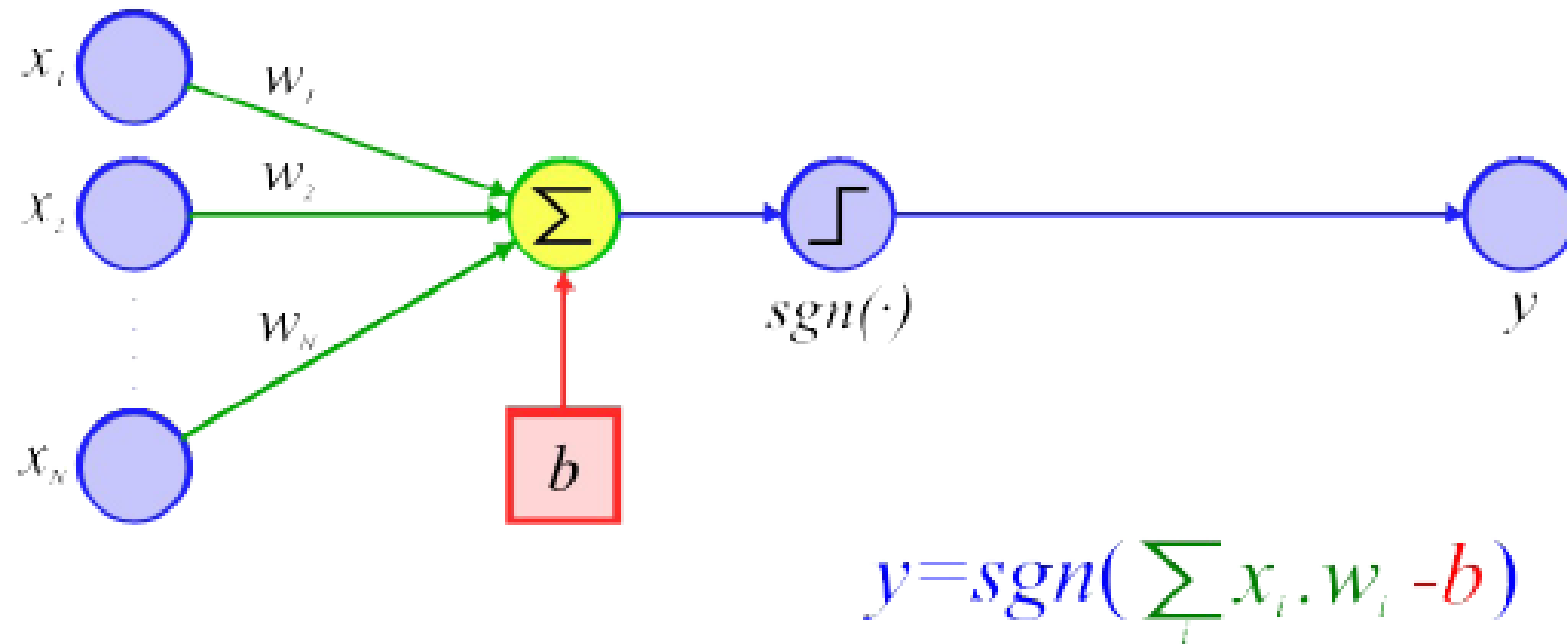
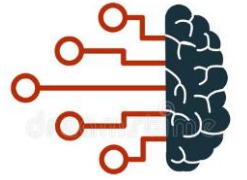


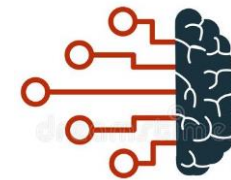
Perceptrón



Red neural profunda

# Neurona Artificial





# La función de transferencia

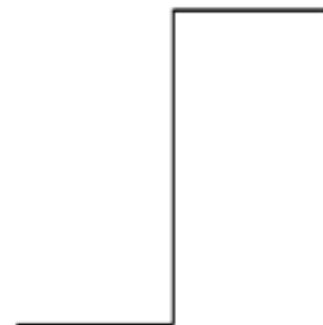
Lineal

$$s(x) = x$$



Escalón

$$s(x) = \text{sgn}(x)$$



Sigmoidea  
Binaria

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

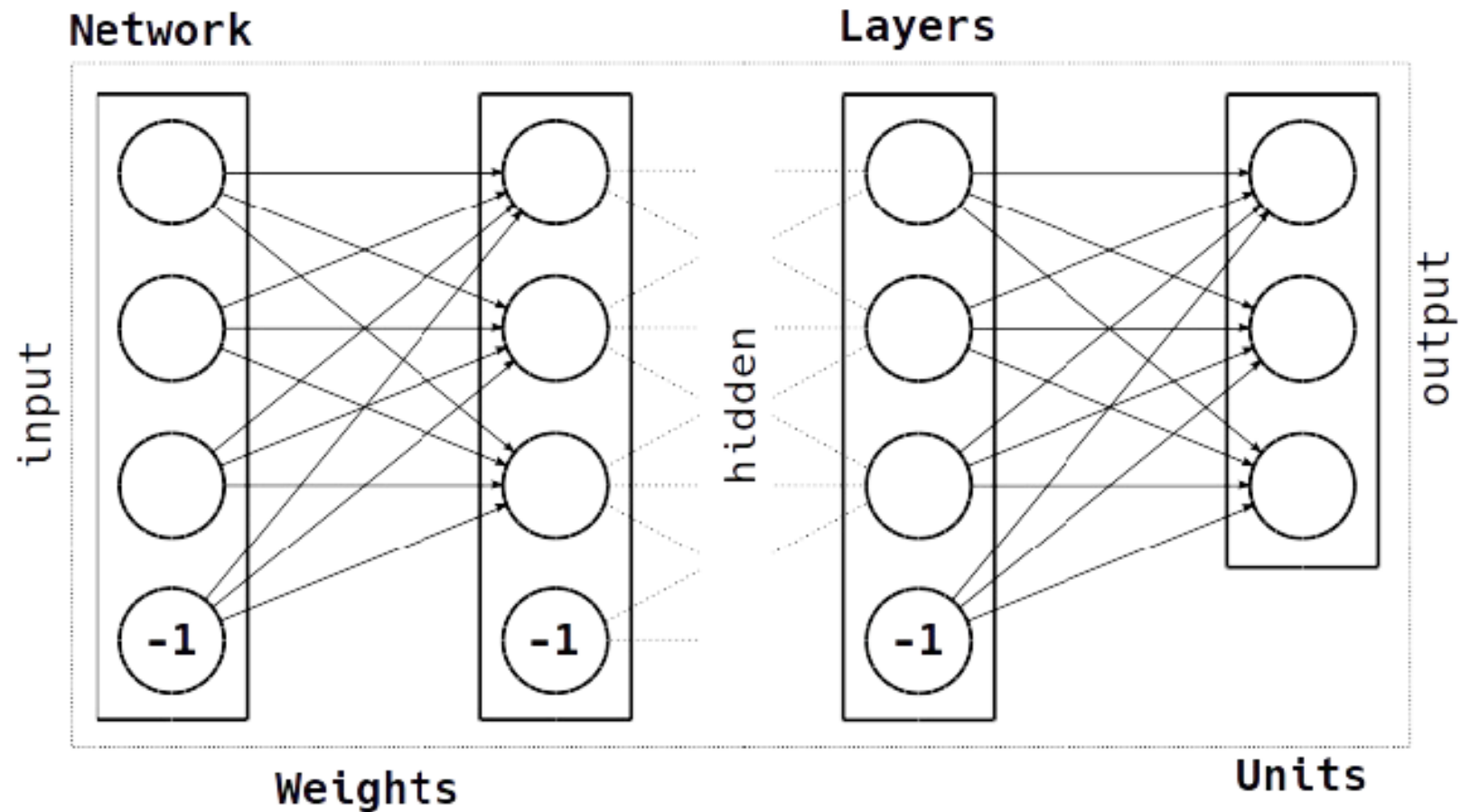
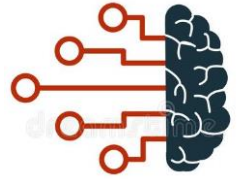


Sigmoidea  
Bipolar

$$s(x) = \tanh(x)$$

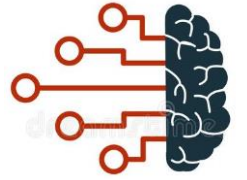


# Red Neuronal





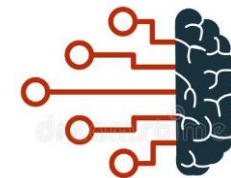
# Postulado de Hebb



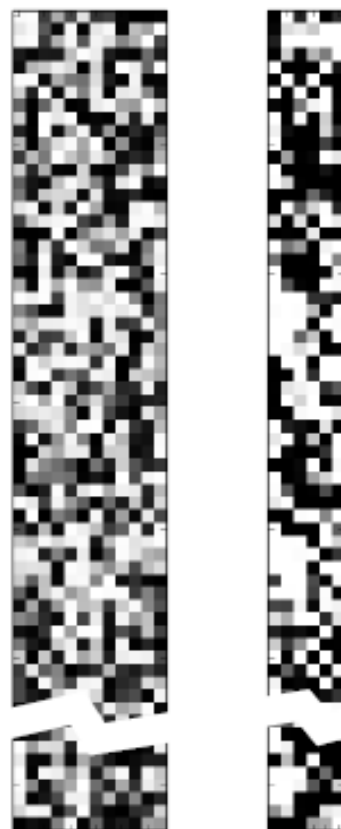
“Cuando el axón de una célula A está lo suficientemente cerca como para excitar a una célula B y repetidamente toma parte en la activación, ocurren procesos de crecimiento en una o ambas células de manera que tanto la eficiencia de la célula A, como la capacidad de excitación de la célula B son aumentadas.”

The Organization of Behavior  
Donald Hebb. (1949)

# Parámetros



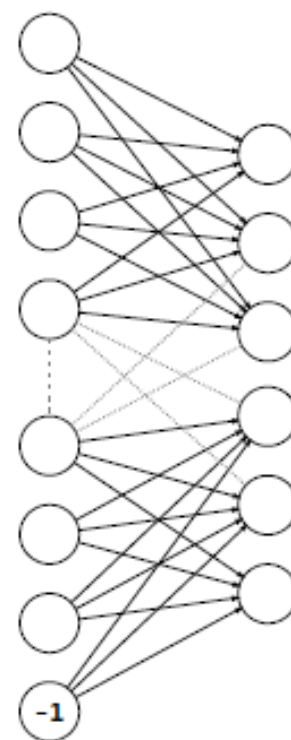
dataset



$X$

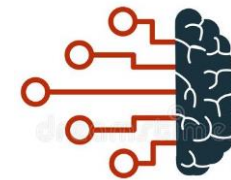
$Z$

network

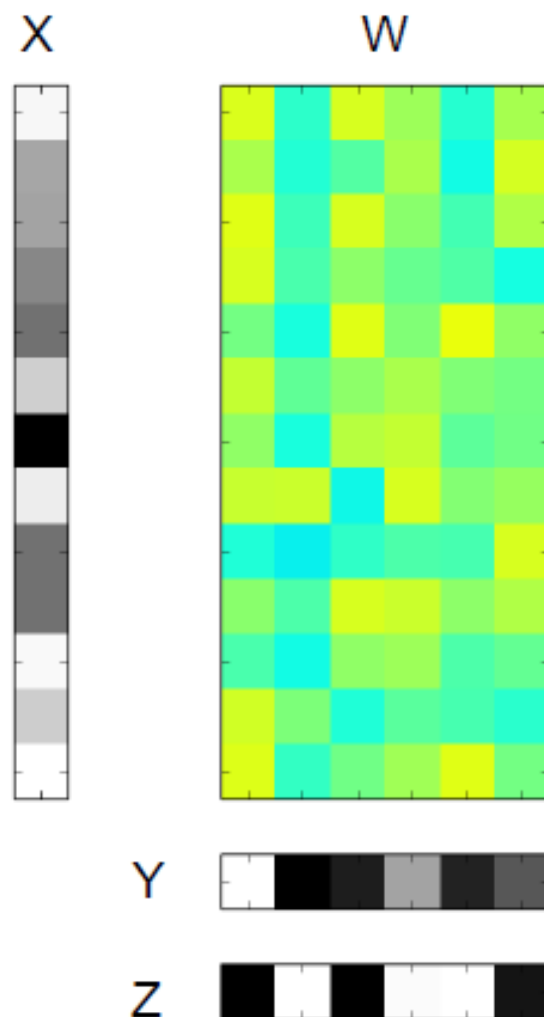


$Y_0$

$Y_1$



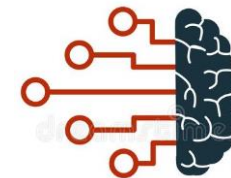
# Función de Costo



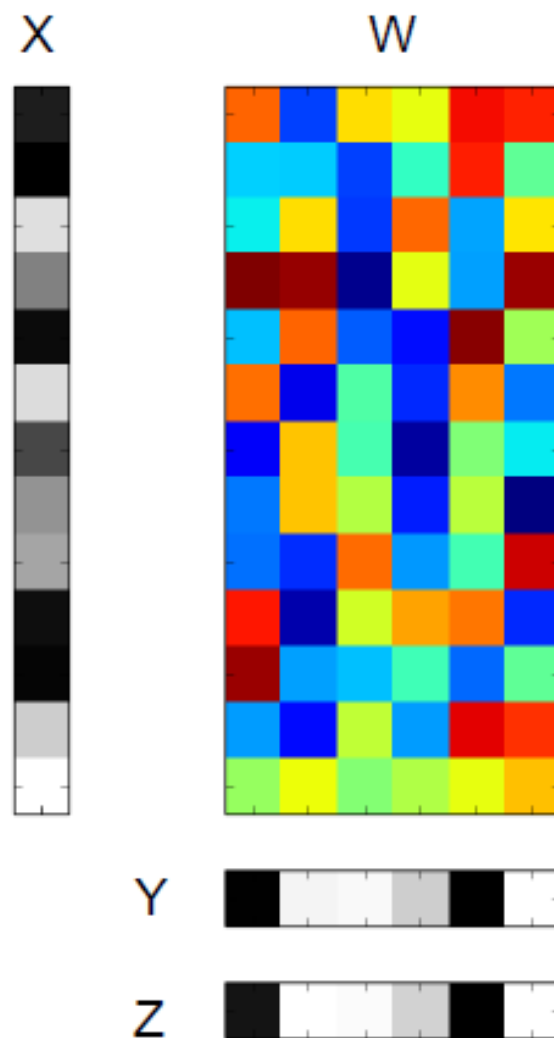
$$Y = s(X \cdot W)$$

$$E = \sum \|Z - Y\|^2$$

$$E = \sum \|Z - s(X \cdot W)\|^2$$



# La Regla Delta



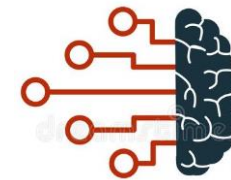
$$Y = s(X \cdot W)$$

$$E = \sum \|Z - Y\|^2$$

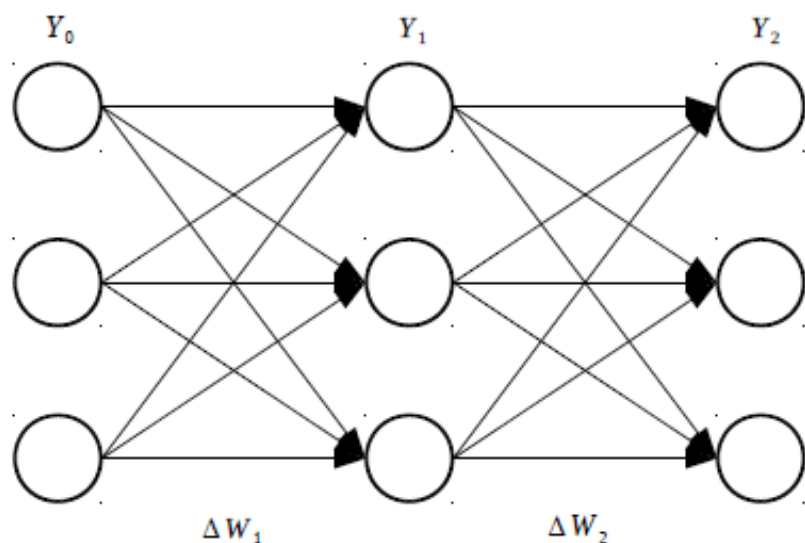
$$E = \sum \|Z - s(X \cdot W)\|^2$$

$$W^{[t+1]} = W^{[t]} + \Delta W^{[t]}$$

$$\Delta w_{ij} = \eta \cdot x_i \cdot (z_j - y_j)$$

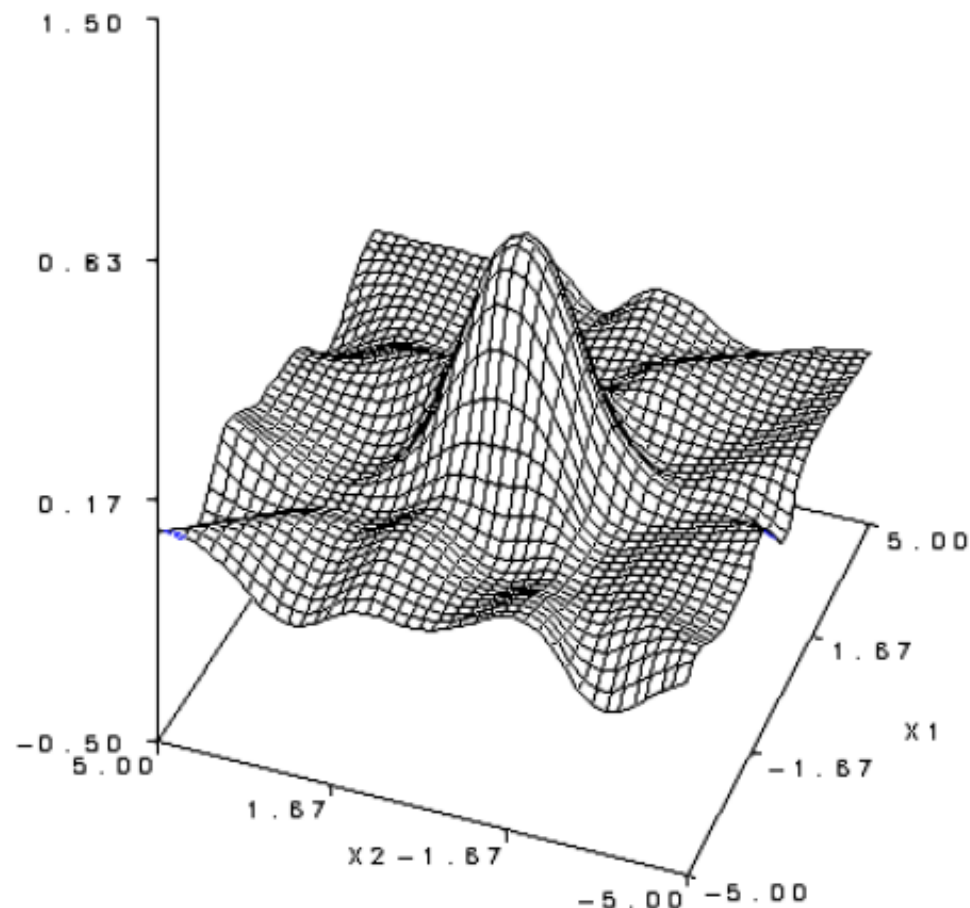


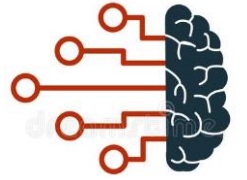
# El Perceptrón Multicapa



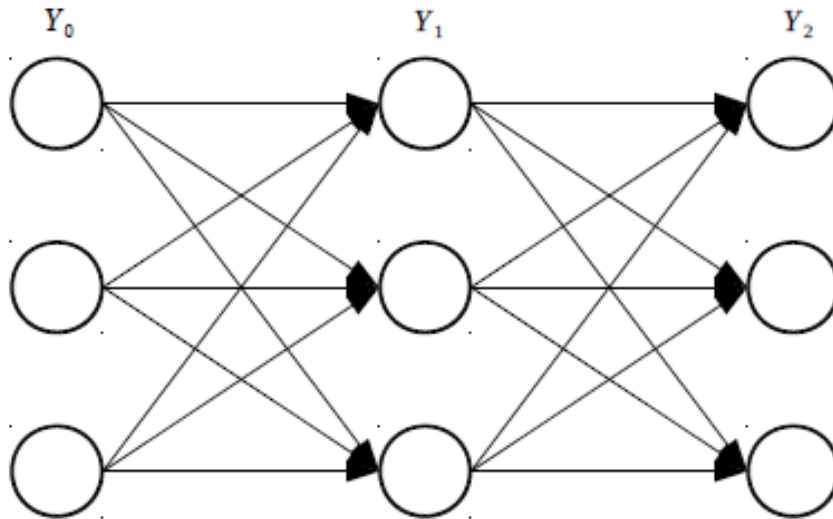
$$\Delta W_2 = \eta \cdot Y_1^T \cdot (Z - Y_2)$$

$$\Delta W_1 = \eta \cdot Y_0^T \cdot (? - Y_1)$$





# Retropropagación del error



$$Y_0 = X$$

$$Y_1 = s_1(Y_0 \cdot W_1)$$

$$Y_2 = s_2(Y_1 \cdot W_2)$$

$$D_2 = (Z - Y_2) \cdot Y_2'$$

$$\Delta W_2 = \eta \cdot (Y_1^T \cdot D_2)$$

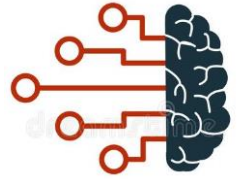
$$\Delta w_{ij} = -\eta \cdot \frac{\partial E}{\partial w_{ij}}$$

$$Y_l' = s'(Y_{l-1} \cdot W_l)$$

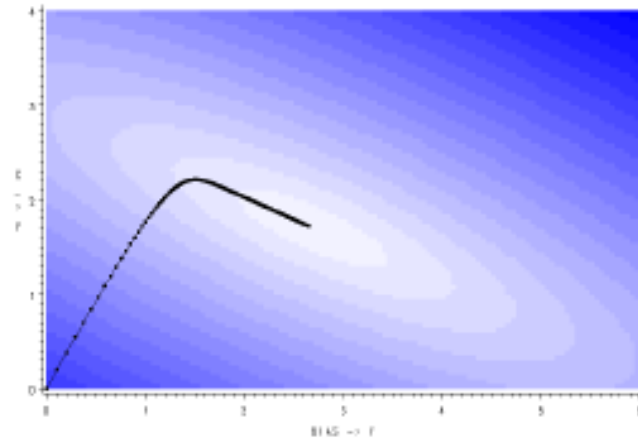
$$D_1 = (D_2 \cdot W_2^T) \cdot Y_1'$$

$$\Delta W_1 = \eta \cdot (Y_0^T \cdot D_1)$$

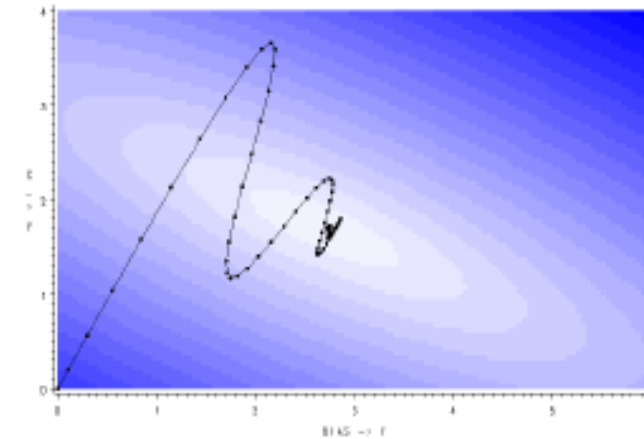
# Mejoras a Backpropagation



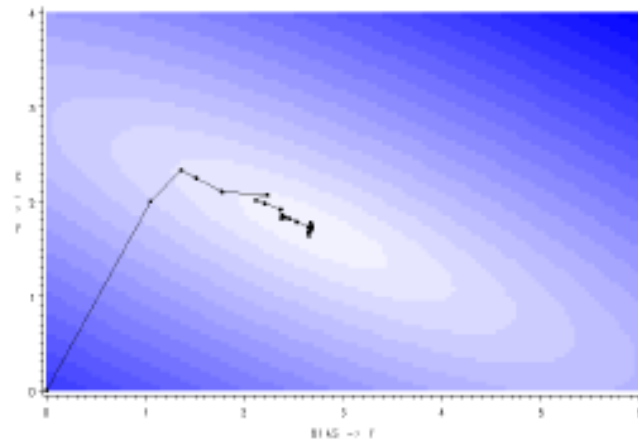
Fair Condition: Backprop Learn=.01 Momentum=0, 593 Iterations



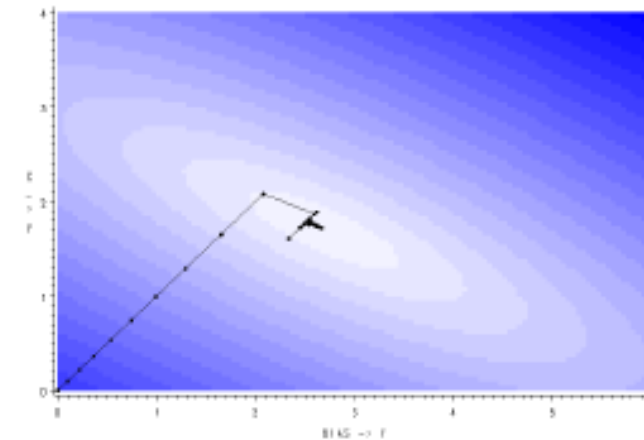
Fair Condition: Backprop Learn=.01 Momentum=.9, 80 Iterations



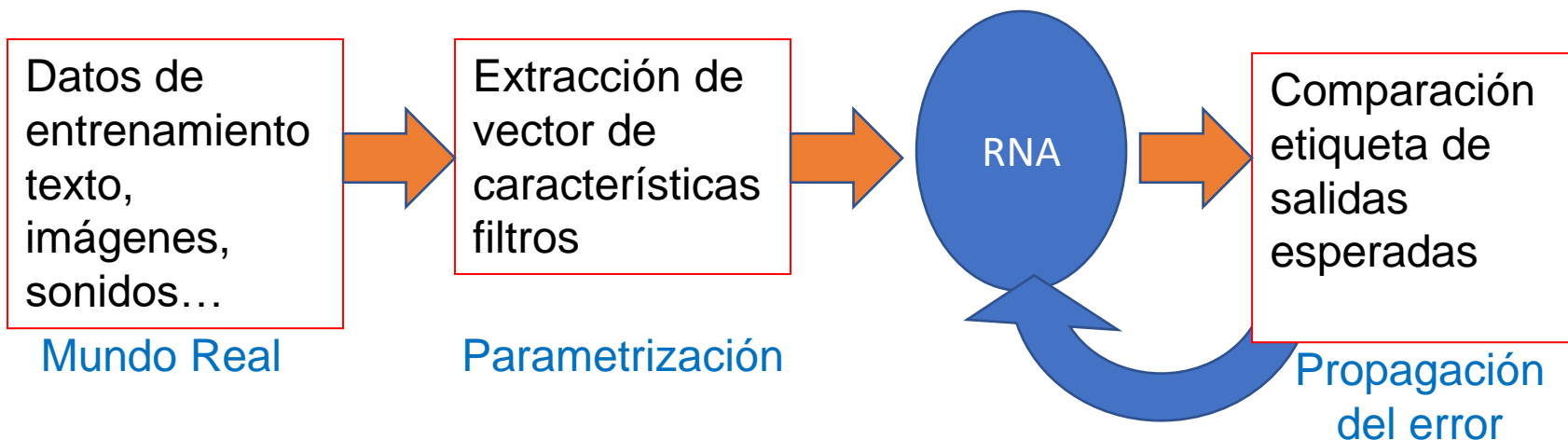
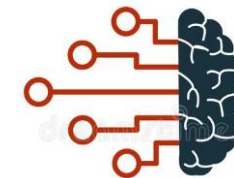
Fair Condition: Quickprop, 26 Iterations



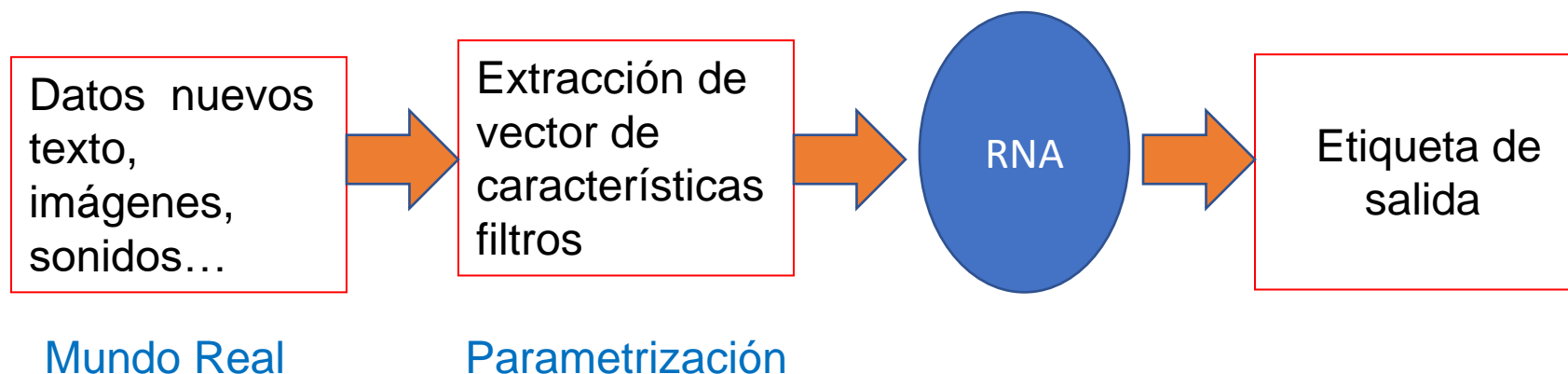
Fair Condition: RProp, 29 Iterations



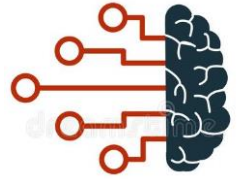
# Aprendizaje



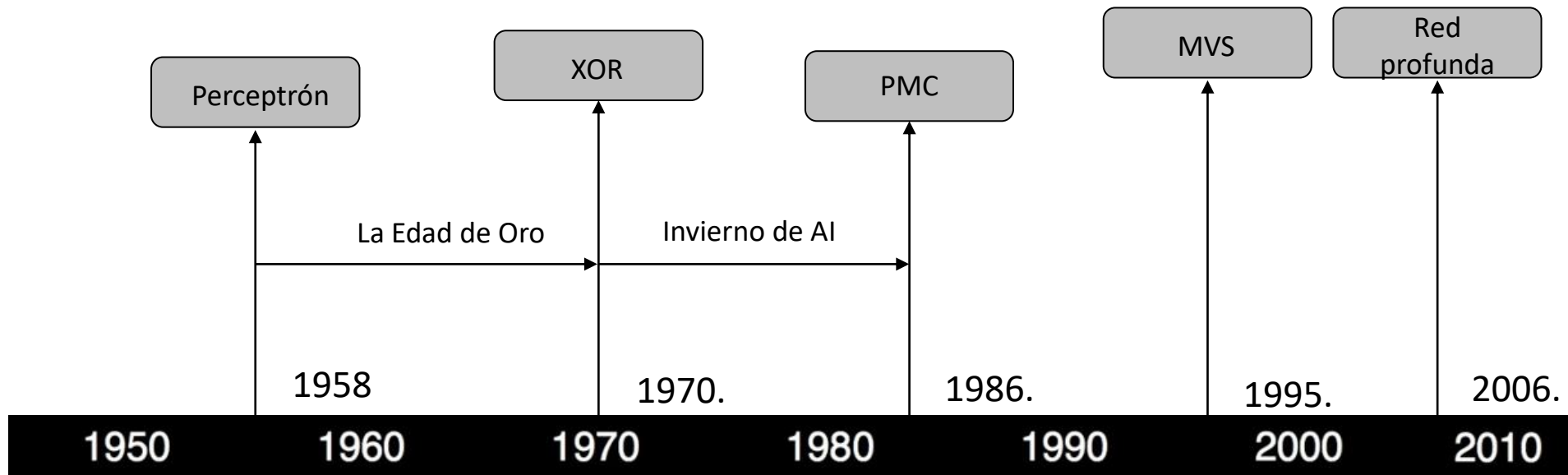
# Reconocimiento

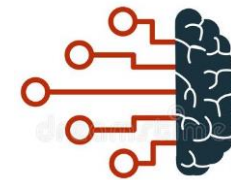






# Historia del desarrollo de las RNA





Gracias.....