

Podemos acercarnos a numerosos conceptos desde diferentes perspectivas.

Hilos conductores:

- Estadística
- Redes neuronales \Leftarrow Punto de continuación

Objetivo fundamental: aprendizaje estadístico

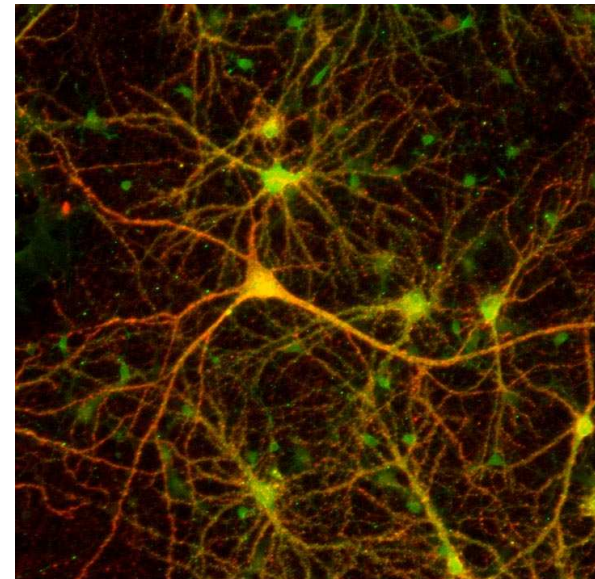
Diferentes tipos de motivaciones:

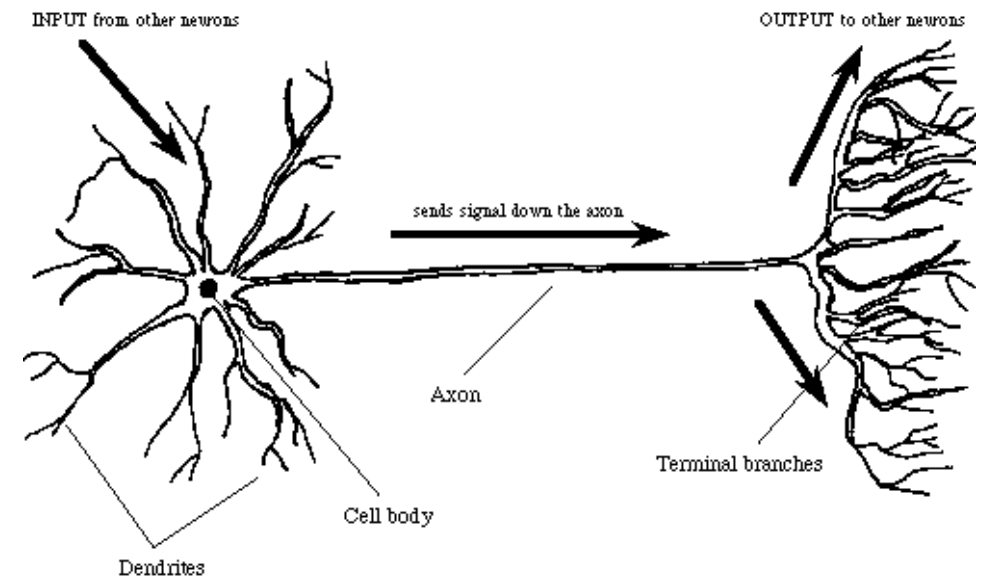
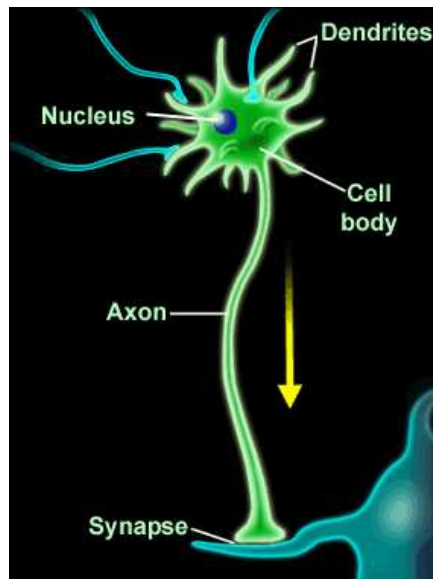
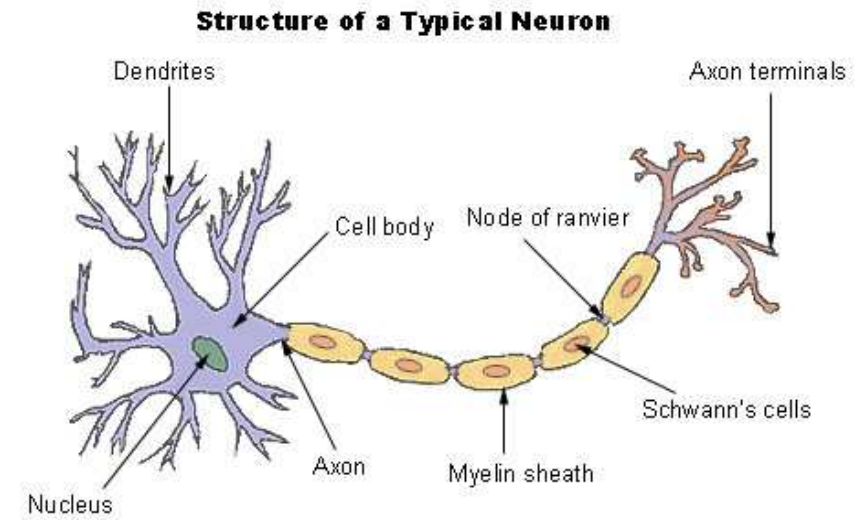
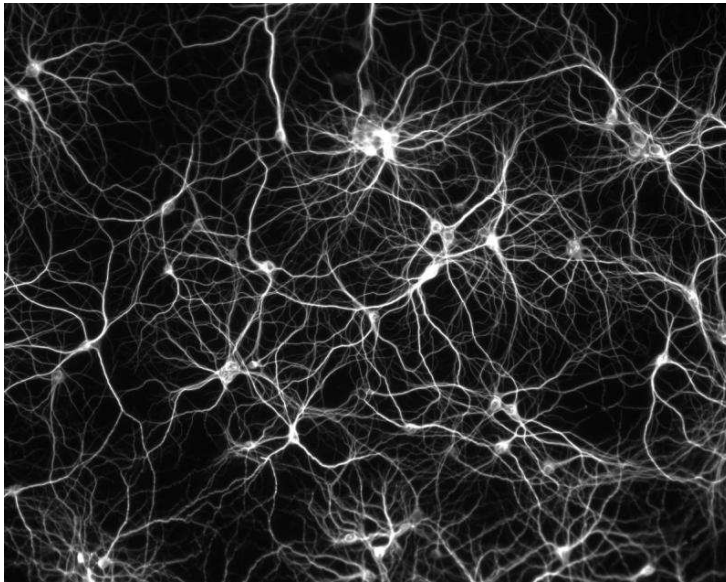
- Biológica
- Física
- Matemática
- Sistémica / Computacional
- Ingenieril

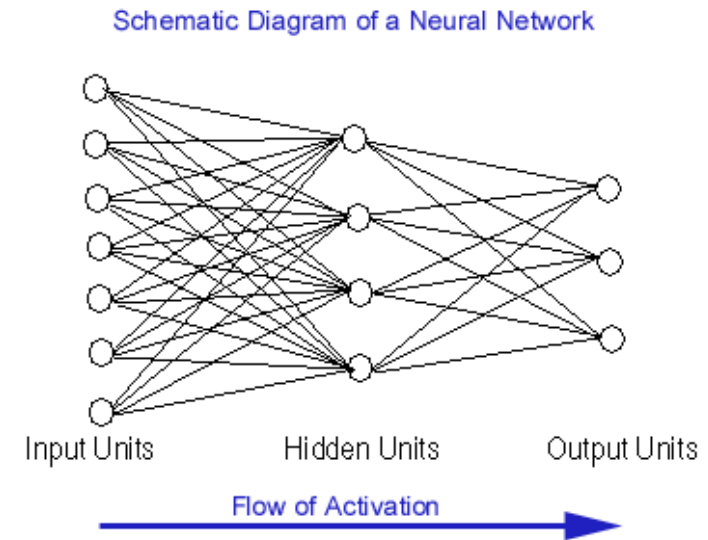
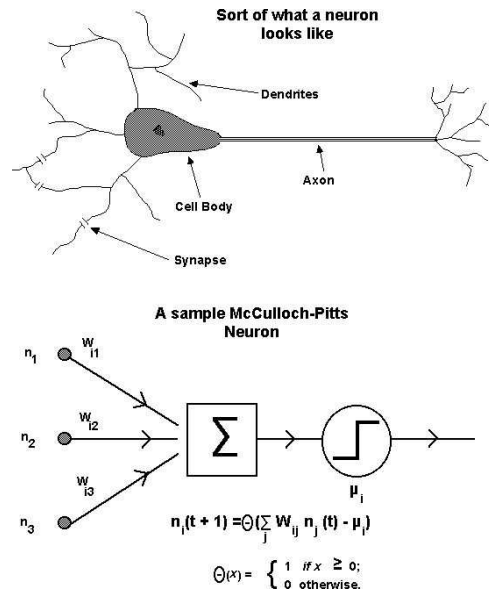
Familia muy variada de modelos o arquitecturas, con diferentes finalidades

Nos interesará perspectiva:

Matemática \Rightarrow Sistémica / Computacional \Rightarrow Ingeniería







Combinación de dos perspectivas:

- *Cómo están constituidas* \Rightarrow Teoría de Marr (sistemas de procesamiento de información)
- *Para qué sirven* \Rightarrow Teoría de resolución de problemas. (Aplicación fundamental: aprendizaje \Rightarrow Teoría del aprendizaje en máquinas)

Herramienta organizacional para estudiar sistemas complejos de procesamiento de información

3 niveles de descripción

- Computacional: objetivo de la computación
- Algorítmico: algoritmo para computar el objetivo
- Implementacional: implementación detallada del algoritmo

Determina herramientas matemáticas a emplear en cada contexto

Paradigma de computación:

- Teoría de Marr (nivel implementacional y algorítmico)

⇒ Sistema dinámico con modularidad específica

Especificidad de redes neuronales: todas ellas comparten propiedades comunes a nivel algorítmico e implementacional ⇒ caracterizables mediante modelado matemático sencillo (sistema dinámico con modularidad específica) que facilita implementación

(También habitualmente...)

Paradigma de aprendizaje:

- Teoría de aprendizaje en máquinas (estadístico, computacional)

⇒ Máquina de aprendizaje

Las **redes neuronales** artificiales están formadas por un número (normalmente elevado) de unidades de procesamiento sencillas (**nodos**), que funcionan en paralelo, con un grado de **interconexión** muy alto, y que se comunican de forma sencilla. Las redes neuronales artificiales pueden adaptar o ajustar su funcionalidad (aprender) **modificando los “pesos”** de las conexiones entre las unidades.

- Caracterización de la **neurona**, unidad básica de computación: entrada, pesos, salida
- Definición de una **topología** de interconexión: conectividad, flujo de información, etc.
- Definición de unas **reglas de ajuste de pesos** (habitualmente, aprendizaje)

Propiedades derivadas de sus ingredientes: no lineal, distribuida y adaptable

Tipo particular de autómata:

- Entrada $I = (x_1, \dots, x_m)$
- Pesos: $W = (w_1, \dots, w_m)$
- Estado interno S : estado de activación de la neurona
- Salida: $O = \mathcal{O}(S)$, habitualmente, no lineal (escalón o una función sigmoideal)

$$\left. \begin{array}{l} \text{Entrada } I \\ \text{Pesos: } W \end{array} \right\} Net = W^T \cdot I$$

Caracterizable Entrada/Salida:

$$O_{t+1} = \mathcal{O}(F(S_t, W_t^T \cdot I_t)), \quad I, W \in R^m$$

Ajuste de pesos de interconexión entre las neuronas

Existen diferentes criterios de ajuste según el problema a resolver y la arquitectura seleccionada.

Clasificación habitual (de aprendizaje en máquinas):

- Supervisado
 - Heurísticos: perceptrón
 - Minimizan función de coste (algoritmos tradicionales de adaptación): PMC, RBF
- Mediante refuerzo (con recompensa/castigo)
 - Actor/crítico
 - Q-learning
- No supervisado
 - Hebbiano (Hopfield, Oja)
 - Competitivo (SOM)
 - Máquinas de Boltzmann

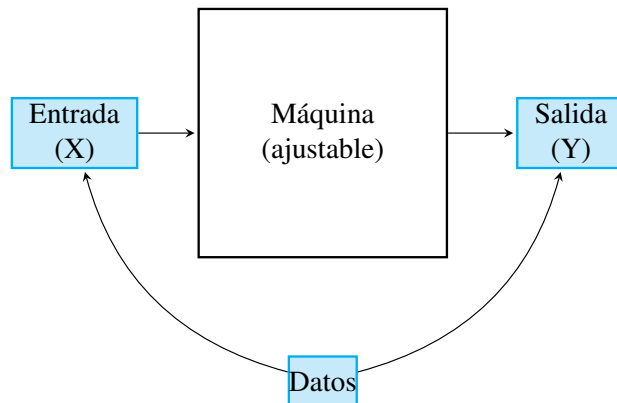
Comprende varios aspectos:

- Definición de conectividad
- Temporalización o sincronización del flujo de información
 - Tiempo continuo-discreto
 - Secuencia de computaciones: Flujo directo o Bucles de realimentación

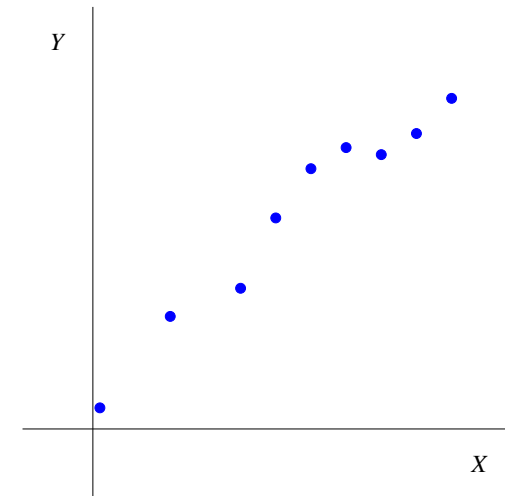
Aprendizaje supervisado versus aprendizaje no supervisado

Aprendizaje supervisado versus aprendizaje no supervisado

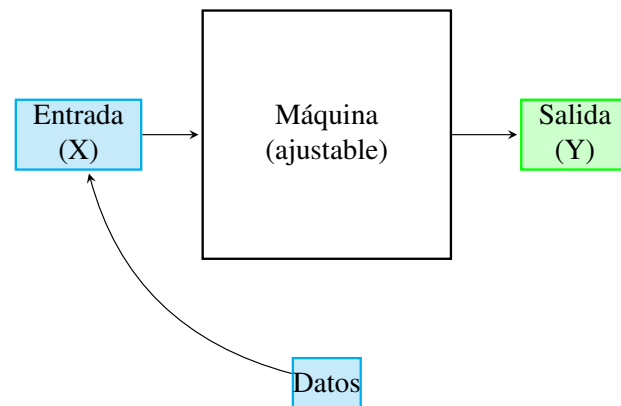
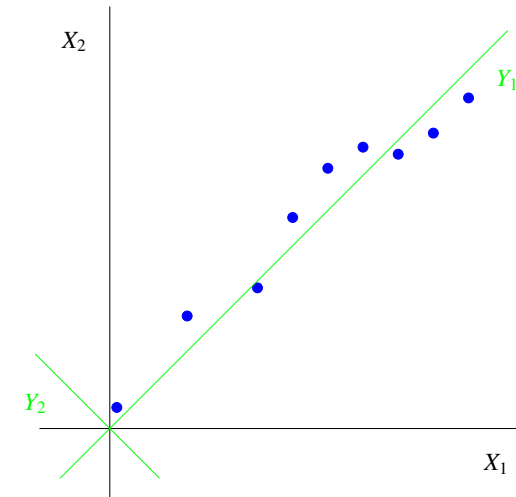
Aprendizaje supervisado



Ejemplo de aprendizaje supervisado: regresión



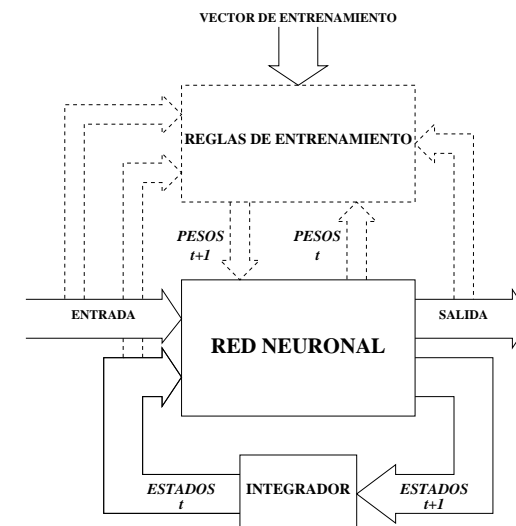
Se considera distribución de Y (una variable) condicionada a X

Aprendizaje no supervisado**Ejemplo de aprendizaje no supervisado: ACP**

Se considera distribución conjunta de (X_1, X_2) (multivariable)

Criterios de clasificación

- Entrada
 - Binaria (funciones discriminantes, etc.)
 - Continua (posible definir diferentes rangos de variación)
- Caracterización temporal
 - En tiempo continuo
 - En tiempo discreto
- Realimentación
 - Feedforward: definen una función u operador
 - Feedback: definen un sistema dinámico
- Entrenamiento
 - Supervisado
 - Con Recompensa/Castigo
 - No Supervisado

Diagrama de bloques general

Como paradigma de computación

- Implementabilidad hardware
- Alto grado de paralelismo
- Robustez frente a posibles fallos (información distribuida)

Habitualmente se simulan mediante software en arquitecturas tradicionales

Como paradigma de aprendizaje (estadístico)

- Entrenamiento supervisado: define espacio parametrizado
- Entrenamiento no supervisado: implementa heurísticos

Máquina de aprendizaje: mecanismos sencillos y generales de adaptación

- McCulloch & Pitts (1943). Estudio biológico del cerebro → Modelo de neurona formal (concepto de umbral).
- Hebb (1949). Hebbian Learning: aprendizaje mediante adaptación de sinapsis (reforzar conexión).
- Minsky (1951). Stochastic Neural Analog Reinforcement Calculator (SNARC).
- Rosenblatt (1959).
{Perceptrón= Estructura + regla de aprendizaje.}
Cambio temporal de pesos sinápticos. Regla del perceptrón.
- Widrow & Hoff (1960's). Filtrado adaptativo (\approx aprendizaje).
Redes neuronales: ADALINE y MADALINE I.
Aprendizaje basado en LMS: Delta Rule.
Problema entrenamiento de varias capas (con función signo).
- Minsky & Papert (1969). Perceptrón única capa: clasificación de 1^{er} orden (no XOR). Problema entrenamiento varias capas.

- Werbos (1974). Algoritmo de retropropagación.
Uso de función sigmoidea
- Grossberg (1970's). Adaptive Resonance Theory
- Kohonen (1970's). Mapas topológicos y memorias asociativas
- Hopfield (1982). Redes de Hopfield
- Rumelhart & McClelland (1986). Perceptrón Multicapa. Popularización de Retropropagación
- Cybenko, Hornik et al., Funahashi (1989).
PMC como aproximador universal
- Retropropagación + VLSI + Computación paralela + Aplicaciones =
Popularización RN's
- Resaca
- Asentamiento y contextualización de aportaciones fundamentales
- Redescubrimiento (siglo XXI): deep learning

- Adaline, perceptrón, madaline
- Perceptrón multicapa (PMC o MLP)
- Funciones de base radial (RBFs)
- Redes de Hopfield, memorias autoasociativas
- Mapas asociativos, memorias asociativas bidireccionales (BAM)
- Mapas topológicos autoorganizativos de Kohonen
- Redes Hebbianas
- Redes recurrentes, TDNNs, CNNs, etc.
- Teoría de resonancia adaptativa (ART) de Grossberg
- Contrapropagación, correlación en cascada, etc.
- Redes jerárquicas multicapa (más allá de PMC, Neocognitrón)
⇒ Deep Learning
 - Supervisadas: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (Long Short-Term Memories, LSTM, etc), Recursive Neural Networks
 - No supervisadas (Unsupervised Pretrained Networks, UPNs):
Autoencoders, Restricted Boltzmann Machines, Deep Belief Networks (DBNs), Generative Adversarial Networks (GANs)