

Introducción a las Redes de Neuronas

Departamento de Informática
Universidad Carlos III de Madrid
Avda. de la Universidad, 30. 28911 Leganés (Madrid)

Introducción de Redes de Neuronas

- Introducción
- Fundamentos biológicos
- Modelo computacional
- Características
- Historia de las redes neuronales

Introducción I

- Objeto de la IA: construcción de sistemas inteligentes
- Sistema Inteligente: dispositivo físico o lógico capaz de realizar tareas como: toma de decisiones, resolución de problemas, aprendizaje...
- Áreas
 - Simbólica – Top-down
 - Subsimbólica – Botton-up

Introducción II

- IA simbólica (Top-down)
 - consiste en programar máquinas capaces de contener representaciones simbólicas de su entorno
 - es necesario disponer de una aproximación a la solución del problema y diseñarla completamente.
 - Ej: sistemas expertos que emulan el razonamiento de expertos humanos en campos muy concretos, como el diagnóstico médico
 - gran dificultad para representar la enorme cantidad de conocimientos, de distinta naturaleza (incluido el sentido común) que posee un ser humano

Introducción III

- IA subsimbólica (Bottom – up)
 - No se realizan diseños de alto nivel, sino que se parte de sistemas genéricos que se van adaptando.
 - Imita los mecanismos que nos capacitan como seres inteligentes.
 - Diseña sistemas inspirados en las características biológicas (sistemas nerviosos, evolución,..) que se adaptan y generan sistemas capaces de resolver problemas
 - Las Redes de neuronas se encuadran en la IA subsimbólica. Intentan emular los sistemas neuronales biológicos

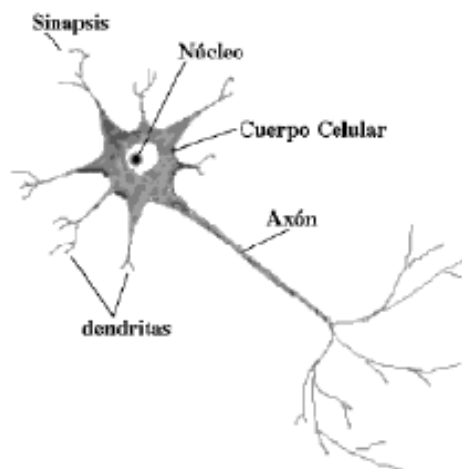
Fundamentos biológicos I

- Sistema de comunicación neuronal:
 - Receptores
 - Situados en las células sensoriales. Recogen información interior/exterior (estímulos)
 - Sistema nervioso
 - Recibe información, la elabora, almacena parcialmente y reenvía
 - Órganos diana o efectores
 - Ej: Músculos y glándulas. Reciben la información y la interpretan

Fundamentos biológicos II

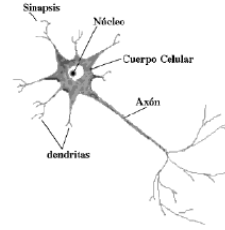
- Elemento fundamental: la neurona
- Funciones de las neuronas
 - Recogida información
 - Integración en un código de activación
 - Transmisión codificada en forma de frecuencia de impulsos
 - Distribución espacial a través del axón
 - Transmisión
 - Neuronas subsiguientes
 - Células efectoras

Fundamentos biológicos III



Fundamentos biológicos IV

- Reciben los estímulos por las dendritas
- La señal de activación se emiten a través del axón
- Sinapsis:
 - elementos de unión entre Axón y dendritas
 - Espacio líquido con determinadas concentraciones iónicas (K, Na)
 - Se convierten en **Potenciadores** o **Inhibidores** de la señal procedente de los axones



Fundamentos biológicos V

- Las neuronas no son elementos lineales
- Si se supera un umbral, se produce una señal de activación
- El sistema nervioso funciona como una enorme malla que propaga señales electroquímicas de unas células a otras, modificando sucesivamente la concentración de iones de las sinapsis.

Fundamentos biológicos VI

- Robustez y tolerancia a fallos
- Adaptación y plasticidad
- Manejar información: inconsistente, difusa y con ruido
- Alta dimensión: aprox. 10^{11} neuronas, cada una conectada con entre 1.000 y 10.000 células
- Lentitud de proceso individual (milisegundos), rapidez de proceso de estímulos:
- Paralelismo
- Aprendizaje continuo
- Determinación genética

Modelo computacional

Una red viene determinada por

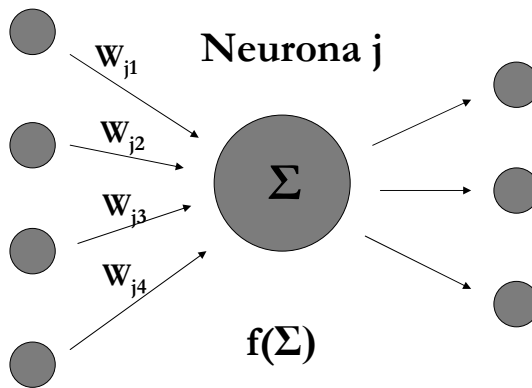
- Conjunto de unidades de proceso.
- Conexiones entre las unidades
- Regla de propagación, para emitir señales a través de las conexiones
- Regla de aprendizaje para modificar los pesos de las conexiones

Modelo computacional

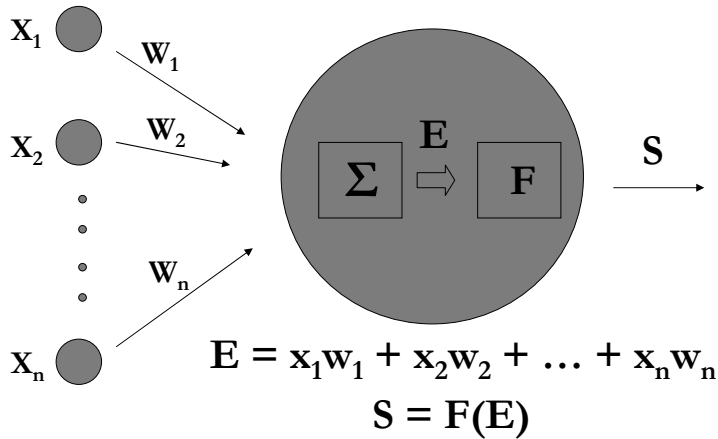
La neurona artificial I

- Cada neurona esta caracterizada por un estado interno denominado **nivel de activación**
 - Conjunto de estados posibles de la neurona. Ejemplos:
 - $S=\{0,1\}$, 0: desactivada, 1: activada
 - $S=\{0,1,2,3,\dots,255\}$, distintos niveles de gris de un pixel
 - $S=[0,1]$, intervalo continuo de valores
- **Función de activación:** permite cambiar el nivel de activación a partir de las señales de entrada
- Estas ultimas se combinan entre sí, generando la entrada total

La neurona artificial II



La neurona artificial III



La neurona artificial IV

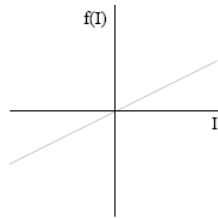
$$E = x_1 w_1 + x_2 w_2 + \dots + x_n w_n$$

- Esto puede ser definido en forma vectorial como sigue:

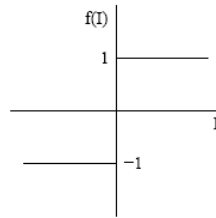
$$E = X^T W$$

- Función de salida:
 - Lineal.- $S = K * E$ con K constante.
 - Umbral.- $S = 1$ si $E \geq \theta$, $S = 0$ si $E < \theta$. Siendo θ el umbral constante.
 - Cualquier función.- $S = F(E)$. Siendo F una función cualquiera.

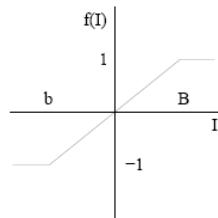
La neurona artificial V



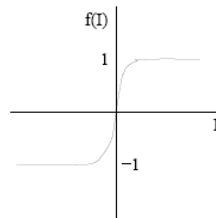
Funcion de Salida Lineal



Funcion de Salida Umbral



Funcion de Salida Saturacion

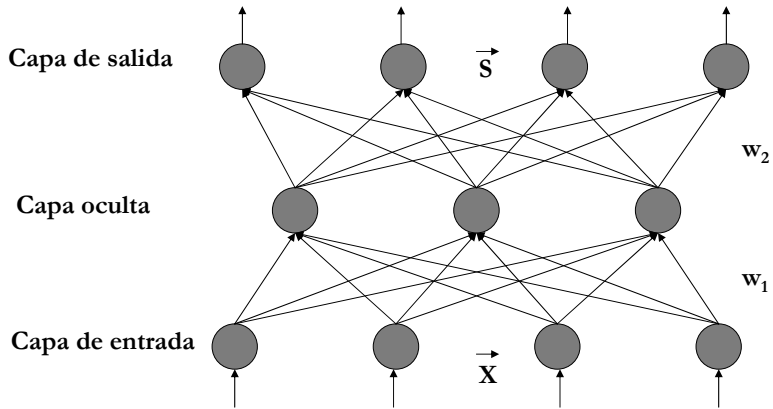


Funcion de Salida Continua

Estructura básica de la red

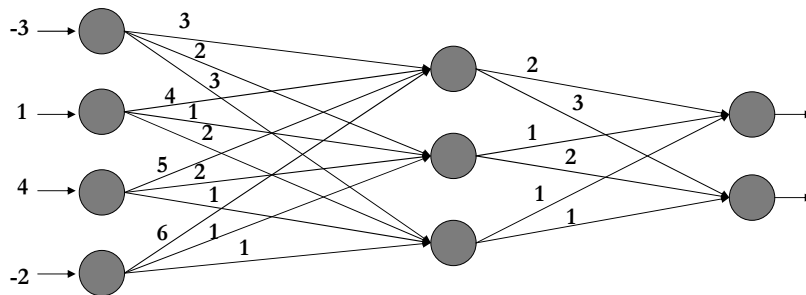
- Arquitectura: forma de conexión entre neuronas
- La estructura básica
 - Red multicapa
 - Capa de entrada
 - Capas ocultas
 - Capa de salida
 - Propagación de la señal hacia delante
 - El valor de la señal se modifica según el peso de la conexión
 - Los pesos se ajustan en la fase de aprendizaje

Estructura básica de la red

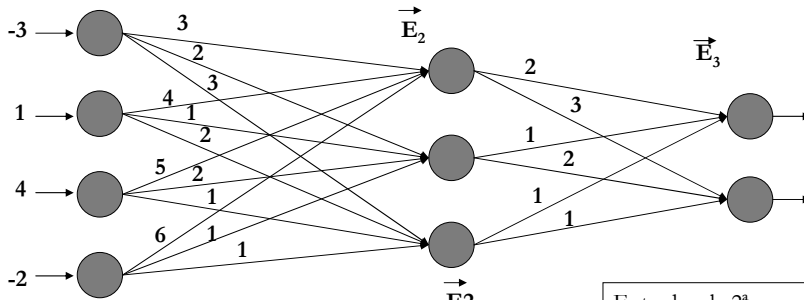


$$\vec{S} = F(F(\vec{x} \cdot w_1) \cdot w_2)$$

Estructura básica de la red



Estructura básica de la red



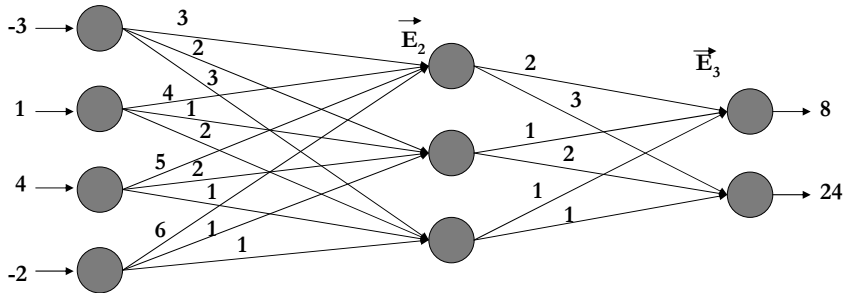
$$\vec{S} = F(F(\vec{x} \cdot w_1) \cdot w_2)$$

$$\vec{E}_2 = \vec{x} \cdot w_1 = \begin{bmatrix} -3 & 1 & 4 & -2 \end{bmatrix} \begin{pmatrix} 3 & 2 & 3 \\ 4 & 1 & 2 \\ 5 & 2 & 1 \\ 6 & 1 & 1 \end{pmatrix} = \begin{bmatrix} 3 & 1 & -5 \end{bmatrix}$$

Entrada a la 2ª capa

\vec{E}_2

Estructura básica de la red



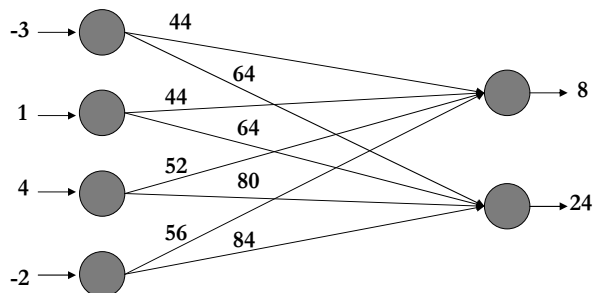
$$F(x) = k \cdot x$$

$$\vec{S} = F(F(\vec{E}_2) \cdot w_2) = F\left(\begin{bmatrix} 6 & 2 & -10 \end{bmatrix} \begin{pmatrix} 2 & 3 \\ 1 & 2 \\ 1 & 1 \end{pmatrix} \right) = F\left(\begin{bmatrix} 4 & 12 \end{bmatrix} \right) = \begin{bmatrix} 8 & 24 \end{bmatrix}$$

Estructura básica de la red

- Si la función de activación es lineal existe una red de una capa equivalente
 - Independientemente del número de capas
 - Su cálculo es trivial
 - Su potencial para solucionar problemas genéricos es pobre

Estructura básica de la red



$$\vec{S} = F(F(\vec{x} \cdot \mathbf{w}_1) \cdot \mathbf{w}_2) = k \cdot (k \cdot \vec{x} \cdot \mathbf{w}_1) \cdot \mathbf{w}_2 = \vec{x} \cdot \overbrace{k^2 \cdot \mathbf{w}_1 \cdot \mathbf{w}_2}^{\mathbf{w}_t}$$

$$\mathbf{w}_t = k^2 \cdot \mathbf{w}_1 \cdot \mathbf{w}_2 = 4 \cdot \begin{pmatrix} 3 & 2 & 3 \\ 4 & 1 & 2 \\ 5 & 2 & 1 \\ 6 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 44 & 64 \\ 44 & 64 \\ 52 & 80 \\ 56 & 84 \end{pmatrix}$$

$$\mathbf{S} = \mathbf{x} \cdot \mathbf{w}_t = (8 \ 24)$$

Aprendizaje

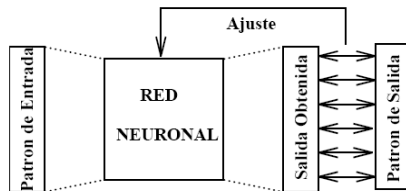
- Aprendizaje: determinación de los valores precisos de los pesos para la resolución eficiente de un problema
- Componente fundamental
- La capacidad de aprendizaje está ligada a los ejemplos del conjunto de entrenamiento
 - Significativo (cantidad)
 - Representativo (variedad)

Aprendizaje

- El proceso de aprendizaje consiste en
 - Introducir paulatinamente todos los ejemplos de aprendizaje
 - Modificar los pesos siguiendo un esquema de aprendizaje
 - Cuando se han introducido todos, se comprueba si se cumple cierto criterio de convergencia. Si no se cumple, se repite el proceso.
- La modificación de los pesos se puede hacer
 - después de introducir cada patrón
 - después de introducir todos los patrones
- La finalización del aprendizaje se hace en función de un criterio de convergencia determinado por
 - Número fijo de ciclos
 - Error menor que un valor dado
 - Modificación de pesos irrelevante

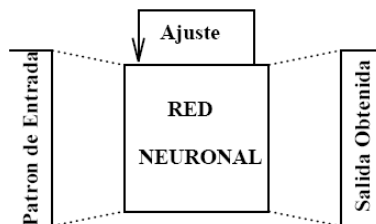
Aprendizaje

- Aprendizaje supervisado
 - Los ejemplos de aprendizaje contienen información relativa a la solución del problema
 - Utiliza profesor externo
 - Maneja información de error o de control
 - La utiliza de forma realimentada para guiarlo.



Aprendizaje

- Aprendizaje no supervisado
 - No utiliza información externa
 - Reajuste automático de los parámetros
 - Autoorganiza la información
 - La red trata de determinar características de los datos: redundancias, regularidades, rasgos significativos



Aprendizaje

- Aprendizaje por refuerzo
 - Utiliza información externa de control
 - No existe medida de error, solo si lo ha habido

Generalización

- Capacidad de la red de responder adecuadamente ante entradas no conocidas
- Problema de sobreajuste
 - La red no generaliza ante casos nuevos
 - Solución: dividir el conjunto de entrenamiento en subconjuntos de
 - Entrenamiento
 - Usado para ajustar el valor de los pesos
 - Validación
 - Usado para medir la eficacia de la red
 - Características
 - Independiente del de aprendizaje
 - Debe ser significativo y representativo

Sobreajuste

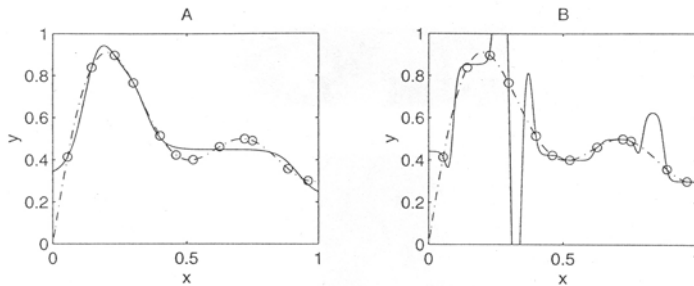


Figure 4.9: Effect of the number of hidden units on the network performance. The dashed line gives the desired function, the circles denote the learning samples and the drawn line gives the approximation by the network. 12 learning samples are used. a) 5 hidden units. b) 20 hidden units.

Comparación RN – RNA Diferencias

	RN	RNA
Conexión	Complejidad de las sinapsis (cada neurona -> miles de conexiones electroquímicas)	Conexiones simples
Sincronismo	Impulsos llegan en cualquier momento	Actualización periódica de parámetros
Patrones	Pueden aprender con pocos ejemplos	Necesitan muchos ejemplos
Arquitect.	Muy complejas	Simple

Comparación RN – RNA

Similitudes

- Forma de representación de la información
- Procesamiento distribuido de la información
- Redundancia
 - Tolerancia a fallos

RNA – Computación convencional

	Comput. convencional	RNA
Unidades de proceso	Una o muy pocas. Realiza muchas operaciones básicas	Muchas. Sólo realiza una operación básica
Poder de proceso	Núm de instrucc. por segundo	Núm de actualizaciones de pesos por segundo
Patrones	Necesita un programa detallado	No son programadas. Aprenden a partir de ejemplos
Almacenam. de información	La información se almacena en lugares específicos de memoria	Codifica la información de forma distribuida. Cada pieza de información puede estar distribuida en una parte extensa de la red Es redundante por naturaleza

RNA – Computación convencional

- Problemas específicos de las RNA
 - Tiempo de aprendizaje indeterminado
 - Diseño poco sistemático: prueba y error

Historia de las redes neuronales I

- Primer modelo matemático de RN. Warren McCulloch y Walter Pitts (1943)
 - Idea de neuronas como procesadores binarios
 - Introducción de función de paso por umbral
 - No tiene capacidad de aprendizaje
- Donald Hebb.- Realiza desarrollos matemáticos de aprendizaje. Aprendizaje hebbiano (1949)
- Marvin Minsky
 - Desarrollo una maquina (con reles, tubos y motores) con 40 neuronas, a partir de células de McCulloch-Pitts (1951)
 - Las conexiones se ajustaban mediante procesos hebbianos

Historia de las redes neuronales II

- Albert Uttley
 - Máquina teórica compuesta de "informes" (1956)
 - Cada informe es un separador lineal, ajustando sus parámetros mediante entropía (Shannon)
- Frank Rosenblatt: **PERCEPTRON SIMPLE**
 - Generalización de McC-P introduciendo aprendizaje (1957)
 - Dos niveles que ajustan las conexiones mediante ejemplos
 - Intento ampliarlo a tres niveles sin éxito, no sabía como ajustar los pesos de las capas ocultas
- Bernard Widrow: **ADALINE**
 - Creador del Adaline (1959), parecida al perceptron
 - Prueba matemática de que el modelo es capaz de minimizar el error tanto como se quiera
 - Problema de separabilidad lineal (como el Perceptron)

Historia de las redes neuronales III (los 60)

- Steinbuch.- Codificación de información en las RN.
Reconocimiento de escritura manuscrita, diagnósticos de fallos
- **Stephen Grossberg**
 - Relaciones entre fenómenos psicológicos y biológicos
 - Análisis matemático y elaboración de nuevos paradigmas de acceso directo, tiempo real, aprendizaje continuo
- Ahun-Ichi Amari.- RN biológicas y modelos matemáticos rigurosos. RN dinámicas y aleatoriamente conectadas
- James Anderson.- Memoria asociativa lineal. Nuevo método de corrección del error. Función umbral en rampa (Brain-state-in-a-bix)
- Fukushima.- (Neo)Cognitron. RN multicapa para visión

Historia de las redes neuronales IV

(los 70 y 80)

- Teuvo Kohonen (70's)
 - RN con conexiones aleatorias, memorias asociativas y matrices de correlación
 - **Mapas autoasociativos (SOM). Learning Vector Quantization (LVQ)**
- Rumelhart (80's)
 - **Perceptron Multicapa (Algoritmo de BackPropagation) (1986)**
- Sejnowski y Hinton (80's).- Maquina de bolzman. Extension a tres capas
- John Hopfield (80's).- Red autoasociativa. Estudios sobre complejidad y convergencia