

UACM

Universidad Autónoma
de la Ciudad de México

NADA HUMANO ME ES AJENO

Redes Neuronales

Sabino Miranda

Inteligencia Artificial (1)

- La Inteligencia Arificial (**IA**) tiene por objeto la comprensión y la construcción de entidades inteligentes.
- Abarca diversos campos
 - Generales: aprendizaje, razonamiento, percepción.
 - Específicos: jugar ajedrez, demostración de teoremas, coches autónomos, diagnóstico de enfermedades, ...
- El término IA se acuñó en 1956 por John McCarthy, proyecto de Dartmouth sobre IA (McCarthy, Minsky, Rochester, y Shannon, 2006).
- Motivación
 - Estudiar aspectos de aprendizaje: **es posible detallar ⇒ simular**
 - Solucionador General de Problemas (GPS)

Inteligencia Artificial (2)

- Sistemas Expertos
- Redes neuronales
- Cómputo evolutivo
- ...
- Presente en actividades consideradas para humanos.
 - Ajedrez: DeepBlue derrotó a Kasparov (1997)
 - Preguntas y Respuestas: Watson derrotó a jugadores humanos (2011)
 - Juego Go: AlphaGo derrotó a Lee Sedol (2016)
- Ramas cruciales: visión artificial, procesamiento del lenguaje natural (PLN).

Inteligencia Artificial (3)

- Visión artificial
 - Reconocimiento facial
 - Vehículos autónomos, drones
 - Diagnósticos médicos
 - Generación de imágenes basadas en texto: DALL-E, MidJourney, Stable Diffusion
 - ...
- PLN ⇒ máquinas entiendan, interpreten y generen lenguaje humano de manera coherente
 - Generación de texto
 - Análisis de sentimientos
 - Clasificación de emociones
 - Perfilado de usuarios
 - Diálogos

Inteligencia Artificial (4)

- Asistentes virtuales
- ...

Aprendizaje Automático (Machine Learning) (1)

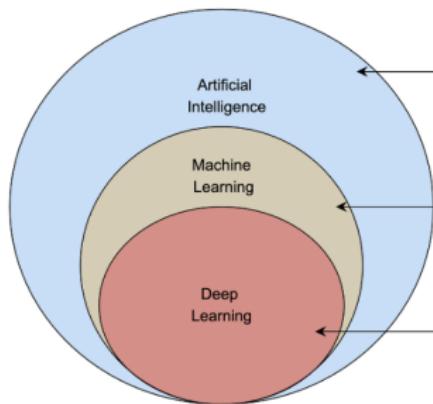


Figura 1: IA, ML y DL

Aprendizaje Automático (Machine Learning) (2)

- El Aprendizaje Automático, Aprendizaje de Máquina o *Machine Learning* (ML) es un subcampo de la IA que se enfoca en el desarrollo de algoritmos y modelos que permiten a las computadoras aprender de los datos. En lugar de ser explícitamente programadas para realizar una tarea, las computadoras utilizan ML para identificar patrones en los datos y tomar decisiones basadas en esos patrones.
- "Se dice que un programa de computadora aprende de la experiencia E con respecto a alguna clase de tareas T y la medida de desempeño P si su desempeño en las tareas en T, medido por P, mejora con la experiencia E"(T. Mitchell).

Aprendizaje Automático (Machine Learning) (3)

- Para definir un problema en aprendizaje automático, se requiere especificar:
 - La experiencia (conjunto de datos de entrenamiento)
 - La tarea (clasificación, regresión, ...)
 - La medida de desempeño (accuracy, precisión, error cuadrático, ...)

Tipos de aprendizaje automático

- **Aprendizaje supervisado**
 - Se entrena al modelo con un conjunto de datos etiquetados, es decir, donde las respuestas correctas son conocidas. Ejemplos incluyen la regresión lineal y la clasificación.
- **Aprendizaje no supervisado**
 - Se trabaja con datos sin etiquetas, y el objetivo es descubrir patrones o estructuras ocultas en los datos. Ejemplos incluyen el clustering y la reducción de dimensionalidad.
- **Aprendizaje por refuerzo**
 - En este caso, un agente aprende a tomar decisiones mediante prueba y error, recibiendo recompensas o castigos en función de las acciones que tome.

Redes Neuronales (*Deep Learning*) (1)

- Las Redes Neuronales Artificiales (*Deep learning*) (ANN o NN por sus siglas en inglés) son una técnica popular de Aprendizaje Automático (*Machine Learning*).
- Simulan el mecanismo de aprendizaje en organismos biológicos.
- Las RNA fueron desarrolladas para simular el sistema nervioso humano para tareas de aprendizaje automático, tratando unidades de cómputo en un sistema de aprendizaje de forma similar a las neuronas humanas.

Redes Neuronales (*Deep Learning*) (2)

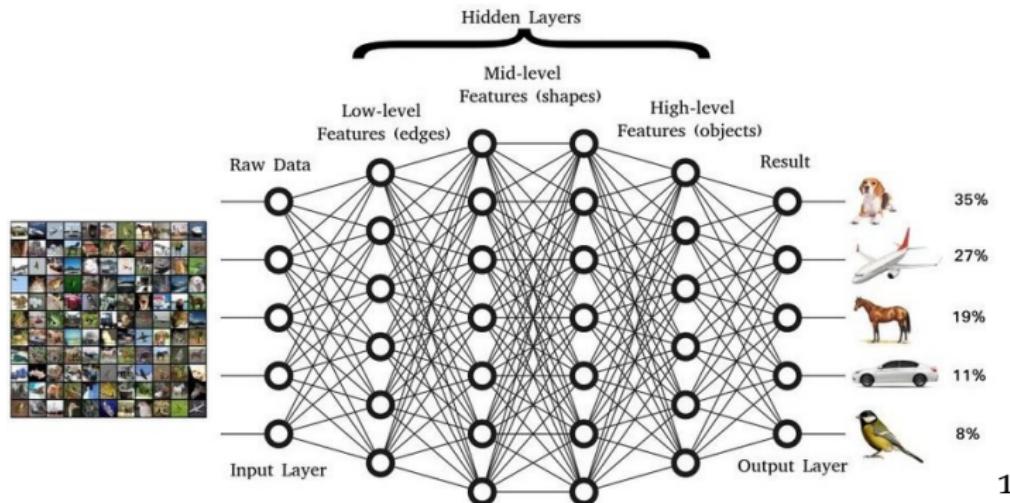


Figura 2: Red Neuronal

Panorama histórico de las RNA

Panorama histórico de las redes neuronales (1)

- **1911:** Se empieza a conocer la anatomía del tejido nervioso, a partir de los trabajos de Santiago Ramon y Cajal.
 - Los investigadores trataron de conocer la forma cómo este tejido y los órganos que constituye, especialmente el cerebro, procesan la información que reciben de los órganos receptores, para dar una respuesta adecuada a sus estímulos.
 - Se demostró que las neuronas son células individuales que se comunican a través de conexiones específicas.
- **1943:** Warren McCulloch y Walter Pitts introduciendo el concepto de una neurona artificial, un modelo matemático para redes neuronales.

Panorama histórico de las redes neuronales (2)

- **1949:** Donald Hebb propone *la regla de Hebb* que describe cómo conexiones sinápticas se fortalecen cuando dos o más neuronas se activan de forma contigua en el tiempo y en el espacio. Idea fundamental en el aprendizaje sináptico.
- **1958:** Frank Rosenblatt desarrolla el *perceptrón*, un modelo de red neuronal simple capaz de aprendizaje supervisado, considerado uno de los primeros algoritmos de aprendizaje automático.
- **1960:** Bernard Widrow y Marcian Hoff desarrollan el modelo *ADALINE* (Adaptive Linear Neuron) y el algoritmo de retropropagación de error, que influirían en desarrollos futuros.

Panorama histórico de las redes neuronales (3)

- **1969:** Marvin Minsky y Seymour Papert publican *Perceptrons*, muestran las limitaciones de los perceptrones de una capa para resolver problemas no lineales, como el XOR. Esto llevó a un estancamiento en la investigación de redes neuronales.
- **1970's (Invierno de la IA):** El interés en las redes neuronales disminuye debido a las limitaciones técnicas y teóricas, así como al auge de otros enfoques como la IA simbólica.
- **1982:** John Hopfield introduce las redes de Hopfield, un tipo de red neuronal recurrente que puede ser utilizada como memoria asociativa.

Panorama histórico de las redes neuronales (4)

- **1986:** Geoffrey Hinton, David Rumelhart, y Ronald Williams popularizan el algoritmo de *backpropagation* (retropropagación) para redes neuronales multicapa, lo que permite entrenar (ajustar los parámetros) a los modelos.
- **1995:** Vladimir Vapnik y Corinna Cortes desarrollan las *Máquinas de Soporte de Vectores* (SVM), un enfoque alternativo que rivaliza con las redes neuronales en muchas aplicaciones de clasificación.
- **1997:** Se desarrollan las primeras *Long Short-Term Memory* (LSTM) por Hochreiter y Schmidhuber, un tipo de red neuronal recurrente diseñada para superar el problema del desvanecimiento del gradiente, permitiendo el aprendizaje de dependencias a largo plazo.

Panorama histórico de las redes neuronales (5)

- **2006:** Geoffrey Hinton y su equipo popularizan el concepto de *deep learning* (aprendizaje profundo) a través de la técnica de preentrenamiento no supervisado utilizando máquinas de Boltzmann restringidas y autoencoders.
- **2009:** ImageNet, una base de datos masiva de imágenes, se convierte en un catalizador para la investigación en visión por computadora y redes neuronales convolucionales (CNNs).
- **2012:** *AlexNet*, una CNN profunda desarrollada por A. Krizhevsky, I. Sutskever y G. Hinton, gana la competencia ImageNet, marcando un hito en la visión por computadora y popularizando las CNNs en la comunidad de IA.

Panorama histórico de las redes neuronales (6)

- **2014:** Ian Goodfellow introduce las Redes Neuronales Antagónicas *Generative Adversarial Networks* (GANs), que permiten la generación de datos sintéticos, abriendo nuevas posibilidades en el campo del aprendizaje.
- **2017:** Se presenta la arquitectura Transformer en el artículo "Attention Is All You Need", revolucionando el procesamiento de lenguaje natural al introducir el mecanismo de atención, que permite paralelización y mejora el rendimiento en diversas tareas como la traducción automática.

Panorama histórico de las redes neuronales (7)

- **2020:** OpenAI lanza *GPT-3*, un modelo de lenguaje, basado en la arquitectura de Transformers, con 175 mil millones de parámetros, que muestra capacidades avanzadas en procesamiento de lenguaje natural (NLP).
- **2021:** DeepMind presenta *AlphaFold 2*, un modelo basado en redes neuronales profundas que resuelve el problema del plegamiento de proteínas con una precisión sin precedentes, revolucionando la biología computacional y la medicina.
- **2022:**
 - OpenAI introduce *DALL·E 2*, un modelo generativo que combina Transformers y GANs para crear imágenes realistas a partir de descripciones textuales, marcando un hito en la generación multimodal.

Panorama histórico de las redes neuronales (8)

- **Noviembre**: OpenAI lanza *ChatGPT*, basado en GPT-3.5, revolucionando la interacción humano-máquina con capacidades conversacionales avanzadas y acceso masivo al público general.
- **2023**:
 - **Marzo**: OpenAI presenta *GPT-4*, un modelo multimodal con mayor precisión y capacidades de razonamiento, adoptado en productos como Microsoft Bing Chat.
 - Surgen modelos de código abierto competitivos: *Mistral AI* (Francia) lanza su primera versión de modelos eficientes (e.g., Mistral 7B), destacándose por su optimización en hardware limitado.
 - **DeepSeek** (China) emerge con modelos especializados en tareas de búsqueda y razonamiento, alcanzando desempeños comparables a los modelos GPTs de OpenAI.

Panorama histórico de las redes neuronales (9)

- **2024:**
 - OpenAI lanza *ChatGPT-5*, con capacidades de razonamiento multi-paso e integración nativa con herramientas externas (API's, robótica), acercándose a una AGI (IA General) limitada.
 - *Mistral* lanza modelos con arquitecturas híbridas (Transformers + SSM, State Space Models, en análisis de series temporales), reduciendo costos computacionales y permitiendo aplicaciones en Edge Computing.
- **2025:**
 - Consolidación de ecosistemas de IA: *ChatGPT-5* y *Gemini* (Google) dominan el mercado comercial, mientras *DeepSeek*, *Mistral*, *Llama* lideran nichos de código abierto.

De la neurona biológica a la neurona artificial

Neurona Biológica (1)

- El sistema nervioso controla todo lo que hacemos desde respirar y caminar hasta pensar y sentir.
- El sistema nervioso de los vertebrados y en particular el del ser humano, integran un eje cerebro espinal (cerebro y médula espinal) y los nervios
- El cerebro y la médula espinal forman el sistema nervioso central.
- Las neuronas forman parte del cerebro y se conectan entre ellas.

Neurona Biológica (2)

- Una neurona (Fig. 3) consta de un cuerpo celular (o soma), del que sale una rama principal (axón) y otras ramas más cortas llamadas dendritas. El axón se ramifica en su extremo (sinapsis) para conectar otras neuronas.

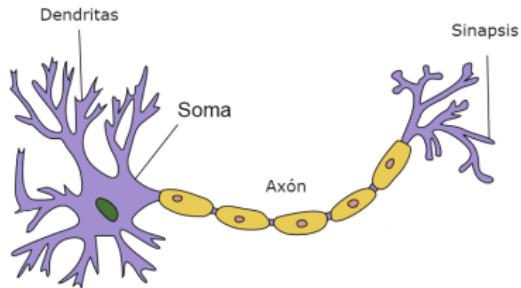


Figura 3: Neurona biológica

Neurona Biológica (3)

- Una de las características que diferencian a las neuronas del resto de células vivas, es su capacidad de comunicación.
- En términos generales, las dendritas y el cuerpo celular reciben señales de entrada, el cuerpo celular las combina e integra y emite señales de salida.
- El axón transporta esas señales a sus terminales, los cuales se encargan de distribuir la información a un nuevo conjunto de neuronas.
- Las señales que se encuentran en una neurona biológica son de naturaleza eléctrica (axón) y química (sinapsis-dendritas) mediante neurotransmisores.

Neurona Biológica (4)

- Se estima que el cerebro humano puede tener alrededor de 67-86 mil millones de neuronas. También se estima que hay alrededor de 85 mil millones de otras células que no son neuronas¹.

¹von Bartheld, C.S., Bahney, J. and Herculano-Houzel, S., 2016. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *Journal of Comparative Neurology*, 524(18), pp.3865-3895

Modelo de Neurona Artificial (1)

- En la Fig. 4, se muestra un modelo de neurona artificial.
- La información que recibe la neurona es el vector $X = (x_1, x_2, \dots, x_n)$.

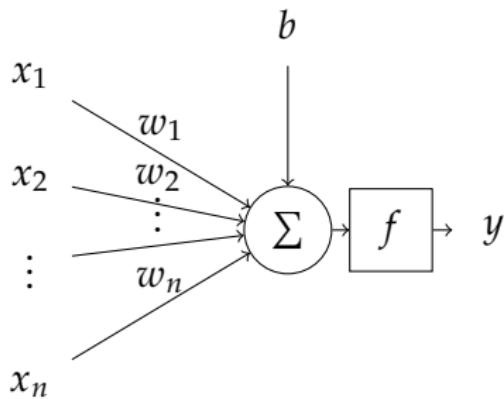


Figura 4: Modelo de neurona artificial

Modelo de Neurona Artificial (2)

- La información recibida por la neurona es modificada por un vector W de pesos sinápticos (emular la sinapsis) entre las neuronas biológicas.

$$\sum x_i * w_i$$

- El parámetro b es el *bias* o umbral de la neurona.
- f es la función de activación, hace que la neurona se active o no se active.
$$f(\sum x_i * w_i + b)$$
- y es la salida de la neurona.

Modelo de Red Neuronal (1)

- Una de las arquitecturas modernas más sencilla es el Perceptrón Multicapa o *Multilayer Perceptron* (MLP) o *Feedforward Neural Networks*.
- Consta de:
 - Capa de entrada: datos de entrada $[x_1, x_2, \dots, x_n]$
 - Capas ocultas
 - Capa de salida
 - Pesos w_{ij} (Parámetros de la red)
 - *Bias* o sesgo b_i (Parámetros de la red)
 - Función de activación para la neurona
 - Número de épocas para el entrenamiento
 - Tasa de aprendizaje (*learning rate*)
 - Función de error o Función de costo o Función de Pérdida

Modelo de Red Neuronal (2)

- Método para el ajuste de los parámetros de la red (pesos y bias)

Modelo de Red Neuronal (3)

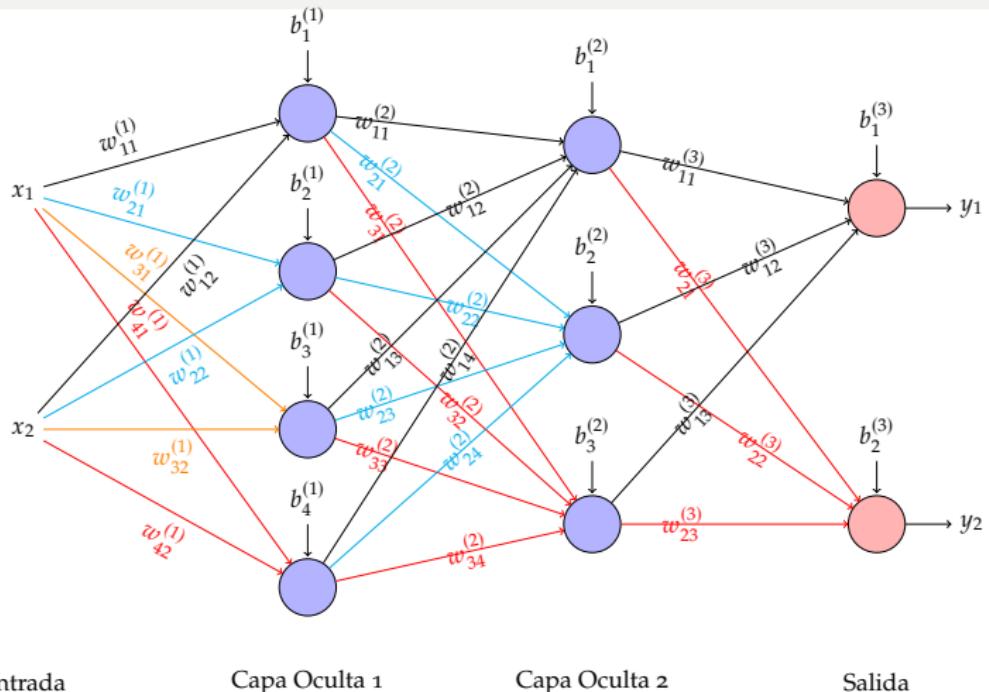


Figura 5: Modelo de Red Neuronal MLP

Modelo de Red Neuronal (1)

- Representación Típica. Para calcular la suma ponderada ($\sum w_i x_i$) (primer capa) se realiza la operación:

$$\mathbf{W} \cdot \mathbf{X} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

- La representación del *bias* es un vector columna:

$$\mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Modelo de Red Neuronal (2)

- La operación final sería:

$$\mathbf{Z} = \mathbf{W} \cdot \mathbf{X} + \mathbf{B} =$$

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Modelo de Red Neuronal

- Calcular la siguiente operación: $Z = W \cdot X + B$

$$\text{Sea } W = \begin{bmatrix} 1.2 & 0.5 & -0.7 & 2.1 \\ -0.3 & 1.8 & 0.4 & -1.5 \end{bmatrix}$$

$$\text{Sea } X = \begin{bmatrix} 0.5 \\ 1.0 \\ -2.0 \\ 3.5 \end{bmatrix}$$

$$\text{Sea } B = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix}$$

Referencias (1)

- ① Deep Learning. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. MIT Press, 2016.
<http://www.deeplearningbook.org>
- ② Dive into Deep Learning. Aston Zhang, Zachary C. Lipton, Mu li, and Alexander J. Smola. Cambridge University Press, 2023. <https://d2l.ai>
- ③ Neural Networks and Deep Learning A Textbook (2nd Edition). Charu C. Aggarwal. Springer, 2023.
<https://doi.org/10.1007/978-3-031-29642-0>
- ④ Deep Learning: Foundations and Concepts. Christopher M. Bishop and Hugh Bishop. Springer, 2024.
<https://doi.org/10.1007/978-3-031-45468-4>

Referencias (2)

- ⑤ PyTorch documentation.

<https://pytorch.org>

- ⑥ Numpy documentation.

<https://numpy.org>

- ⑦ Python documentation.

<https://www.python.org>