

# Arquitecturas de Deep Learning



UNIVERSIDAD  
NACIONAL  
DE COLOMBIA

Las arquitecturas modernas de deep learning se basan en operaciones como transformaciones densas, convoluciones, recurrencia y atención, cada una optimizada para extraer patrones específicos en distintos dominios de datos.

## REDES DENSAS (MLP)

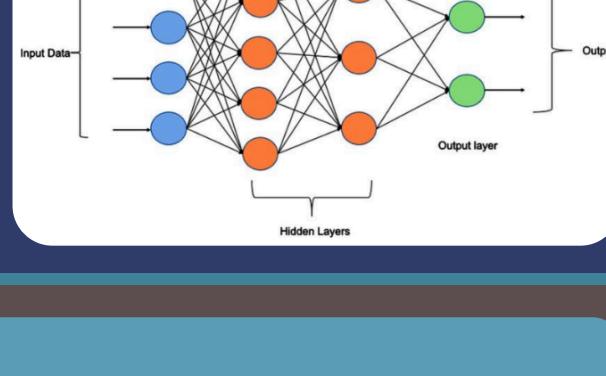
$$\hat{y} = \text{sign}(\mathbf{w}^T \mathbf{x})$$

Modelo

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} - \sum_{\mathbf{x}_n \in \mathcal{X}} \langle \mathbf{w}^T \mathbf{x}_n, y_n \rangle$$

Entrenamiento

Las MLP se utilizan ampliamente en tareas de clasificación, predicción en modelos estadísticos, reconocimiento de patrones simples, y como capas finales de decisión en arquitecturas más avanzadas donde se requiere combinar características extraídas por otros modelos.



## REDES CONVOLUCIONALES (CNN)



$$[\mathbf{H}]_{i,j,d} = \sum_{a=-\Delta}^{\Delta} \sum_{b=-\Delta}^{\Delta} \sum_c = [\mathbf{V}]_{a,b,c,d} [\mathbf{X}]_{i+a,j+b,c}$$

Modelo matemático

Las CNN se utilizan en tareas de visión por computadora como clasificación de imágenes, detección de objetos, reconocimiento facial, y en general en cualquier problema que requiera extraer patrones visuales complejos de datos en forma de imágenes o video.

## REDES RECURRENTES (RNN/LSTM/GRU)

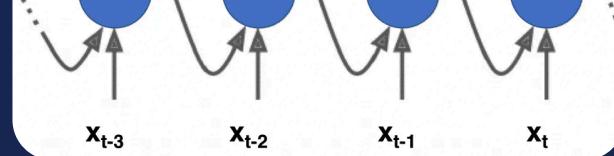
$$h_t = \phi(W_x x_t + W_h h_{t-1} + b) \quad y_t = W_y h_t + c$$

Modelo matemático

RNN (Recurrent Neural Network): Procesa secuencias mediante un estado oculto que se actualiza paso a paso.

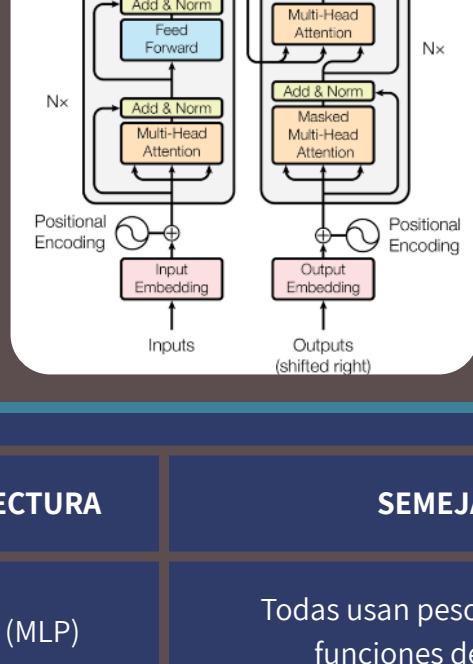
LSTM (Long Short-Term Memory): Variante de RNN con puertas que controlan memoria para capturar dependencias largas.

GRU (Gated Recurrent Unit): Versión simplificada de LSTM con menos puertas y memoria más eficiente.



Las redes recurrentes, se usan en tareas secuenciales como modelado de lenguaje, análisis de series temporales, reconocimiento de voz, predicción de secuencias, clasificación de señales, y en general en cualquier problema que requiera procesar información ordenada en el tiempo.

## TRANSFORMERS (BASADAS EN ATENCIÓN)



$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$

Ecuación clásica de atención escalada

Las arquitecturas basadas en atención se utilizan ampliamente en modelado de lenguaje, traducción automática, resumen de texto, generación de contenido, análisis de secuencias, sistemas de diálogo, clasificación de texto, recuperación de información, visión por computadora con Vision Transformers (ViT), y en general en tareas que requieren capturar dependencias globales en secuencias o elementos.

ARQUITECTURA	SEMEJANZAS	DIFERENCIAS (LIMITACIONES)
Densas (MLP)	Todas usan pesos entrenables y funciones de activación	No capturan estructura espacial ni temporal
Convolucionales (CNN)	Comparten el uso de funciones de activación y entrenamiento por gradiente	Solo modelan patrones locales, limitadas para dependencias largas
Recurrentes (RNN, LSTM, GRU)	Comparten principio de transformación no lineal y modelan secuencias.	Dificultad para paralelizar, lentas en secuencias largas y complejidad al entrenar
Transformers (Atención)	Utilizan activaciones y capas densas internamente	Consumen mucha memoria, requieren grandes conjuntos de datos.