

Comparación de arquitecturas en Deep Learning

REDES DENSAS (MLP)

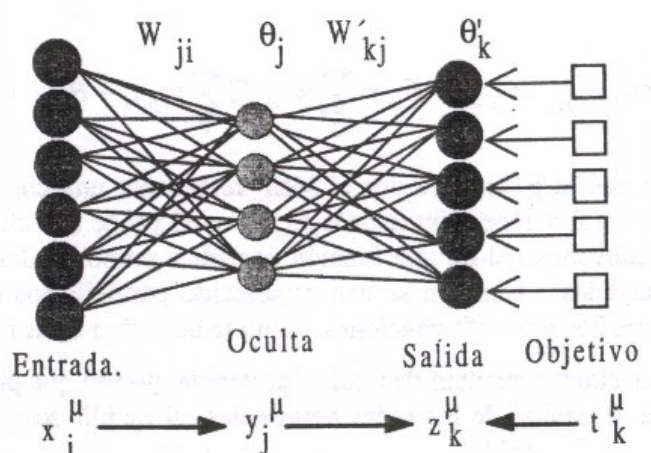
Son redes totalmente conectadas donde cada neurona recibe todas las entradas. No aprovechan estructura espacial ni temporal, pero son versátiles y simples.

Modelo matemático fundamental:

- Para una capa:
$$z = \sigma(Wx + b)$$
- Para varias capas:
$$a^l = \phi(W^l a^{l-1} + b^l)$$

Aplicaciones:

Datos tabulares, regresión, clasificación simple y capas finales de otras redes.



REDES CONVOLUCIONALES (CNN)

Usan filtros locales que detectan patrones (bordes, texturas). Aprovechan la estructura espacial de imágenes o señales 2D.

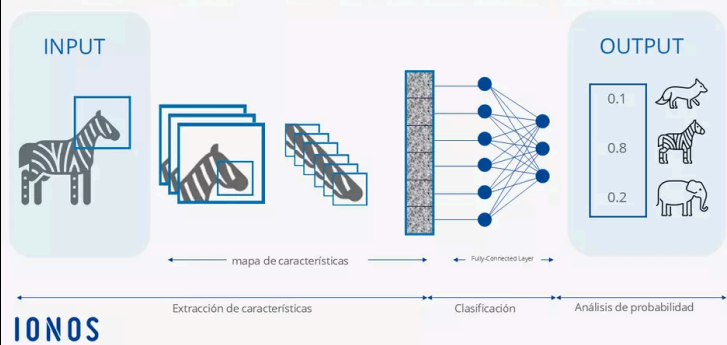
Modelo matemático fundamental:

$$y_k(i, j) = \sum_c \sum_{u, v} W_{k,c}(u, v) x_c(i - u, j - v) + b_k$$

Aplicaciones:

Visión por computador: clasificación, detección, segmentación.

CONVOLUTIONAL NEURAL NETWORK (CNN)



REDES RECURRENTE (RNN)

Procesan secuencias paso a paso, manteniendo un estado oculto que recuerda información previa.

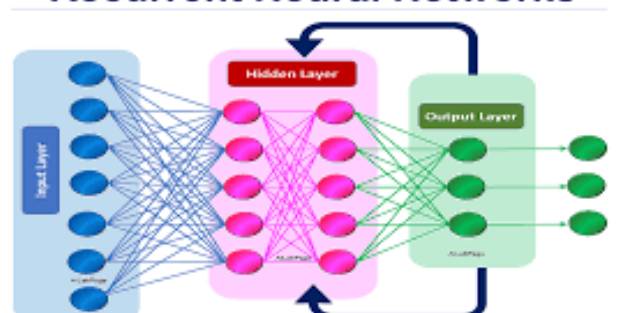
Modelo matemático fundamental:

$$h_t = \phi(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \rightarrow y_t = W_{hy}h_t + b_y$$

Aplicaciones:

Series temporales, voz, texto, secuencia-a-secuencia.

Recurrent Neural Networks



ARQUITECTURAS DE ATENCIÓN / TRANSFORMERS

No procesan secuencias paso a paso, sino que cada token "atiende" a los demás. Muy paralelizables y dominantes en NLP y multimodal.

Modelo matemático fundamental:

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

Aplicaciones:

Traducción, resumen, modelos de lenguaje, visión (ViT), audio multimodal.

