

INFOGRAFÍA

Arquitecturas de Aprendizaje Profundo: Análisis, Estructura y Aplicaciones

Comparación estructural y funcional entre MLP, CNN, RNN y Transformers, ilustrando cómo cada arquitectura procesa la información y en qué contextos resulta más efectiva.

1. Redes Densas (Fully Connected / MLP)

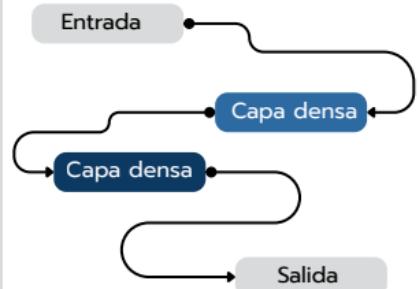
- Modelo matemático:

$$h = \sigma(Wx + b)$$

Ecuación para capas ocultas:

$$h^{(l)} = \sigma\left(W^{(l)}h^{(l-1)} + b^{(l)}\right), \quad l = 1, 2, \dots, L \quad \hat{y} = h^{(L)}$$

- Esquema:



- Aplicaciones:

- Clasificación tabular
- Regresión
- Arquitecturas base en muchas redes

2. Redes Convolucionales (CNN)

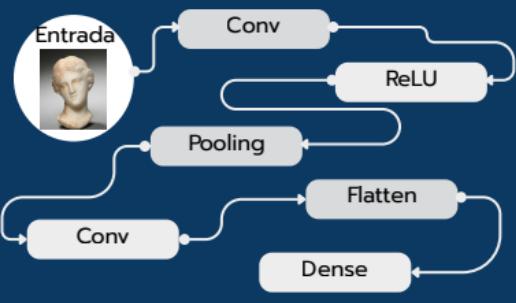
- Modelo matemático:

$$h_{i,j,k} = \sigma\left(\sum_{m,n,c} x_{i+m, j+n, c} K_{m, n, c, k} + b_k\right)$$

Pooling:

$$y_{i,j} = \max_{m,n} x_{i+m, j+n}$$

- Esquema:



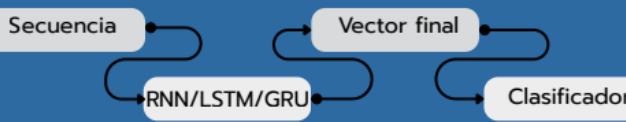
3. Redes Recurrentes (RNN / LSTM / GRU)

- Modelo matemático:

$$h_t = \sigma(Wx_t + Uh_{t-1} + b)$$

$$\begin{aligned} f_t &= \sigma(W_fx_t + U_fh_{t-1} + b_f) & \tilde{c}_t &= \tanh(W_cx_t + U_ch_{t-1} + b_c) \\ i_t &= \sigma(W_ix_t + U_ih_{t-1} + b_i) & c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ o_t &= \sigma(W_ox_t + U_oh_{t-1} + b_o) & h_t &= o_t \odot \tanh(c_t) \end{aligned}$$

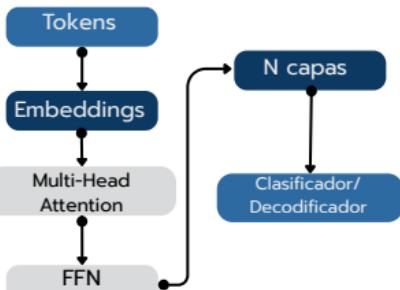
- Esquema:



- Aplicaciones:

- Series de tiempo
- Reconocimiento de voz
- Traducción automática clásica
- Modelado de lenguaje (pre-Transformers)

- Esquema:



4. Transformers (Atención / Self-Attention)

- Modelo matemático:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

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

- Aplicaciones:

- Modelos modernos de NLP (ChatGPT, BERT, LLaMA)
- Audio / Speech
- Visión computacional (Vision Transformers - ViT)