

# Apuntes sobre IA

Sergio de Mingo

2 de febrero de 2026

## Contents

### 1 Redes Neuronales Convolucionales

1

## 1 Redes Neuronales Convolucionales

Anteriormente aplanamos una imagen de  $28 \times 28$  a 784. Al hacer eso, la red ya no sabe si el píxel 1 está al lado del píxel 2 o del píxel 28. Simplemente ve una lista de números. Piensa en esto: ¿Cómo podrías tú reconocer una cara si te dieran todos los píxeles de una foto mezclados en una bolsa? Sería imposible. Las Redes Neuronales Convolucionales o CNN vienen a resolver precisamente eso. Hasta ahora, las redes veían los píxeles como una lista plana. Si movíamos el dibujo de un “8” un poco a la derecha, la red se confundía porque los píxeles ya no caían en las mismas neuronas de entrada. Las CNN solucionan esto imitando el córtex visual humano.

En lugar de conectar cada píxel a una neurona, usamos un filtro. Como si fuera una pequeña rejilla de  $3 \times 3$  que se desliza sobre la imagen. Este filtro no mira toda la imagen a la vez; mira solo un pequeño trozo, extrae una característica (como una línea vertical o un borde) y se mueve al siguiente píxel. La red ahora no aprende a reconocer «píxeles en tal posición», sino que aprende los valores de ese filtro que detectan rasgos importantes (bordes, curvas, texturas).

La estructura maestra de una CNN podría resumirse en los siguientes puntos:

- **Capa de convolución:** Es el motor. Aquí es donde los filtros escanean la imagen. Si aplicas 32 filtros diferentes, obtendrás 32 mapas de características (versiones de la imagen donde resaltan cosas distintas).
- **Capa de pooling:** Es el sintetizador. Su trabajo es reducir el tamaño de la imagen. Toma, por ejemplo, un cuadrado de  $2 \times 2$  y se queda solo con el valor más alto (el más brillante). Hacemos esto porque si detectamos un borde, no nos importa el píxel exacto, nos importa saber que ahí hay un borde. Esto hace que la red sea invariante a la traslación.
- **Capa totalmente conectada:** Al final de la red, después de que los filtros hayan extraído toda la información, volvemos a usar las capas que ya conoces para tomar la decisión final (¿Es un 8 o es un 3?).