

14. Redes Convolucionales

Ejercicio 14.1. Invarianza

Dada una imagen representada por una función 2D $f(x, y)$, donde x e y son las coordenadas de los píxeles, considera las siguientes transformaciones t de la imagen y las funciones que la procesan. Determina cuáles de estas funciones son **invariantes** bajo las transformaciones (es decir: $f(t(I)) = f(I)$).

Transformaciones:

- **Traslación con padding 0:** $t_1(x, y) = (x + 1, y)$ (desplaza la imagen un píxel a la derecha con padding 0)
- **Traslación con padding circular:** $t_2(x, y) = (x + 1, y)$ (desplaza la imagen un píxel a la derecha con padding circular)
- **Rotación:** $r(x, y)$ rota la imagen 90° en el sentido de las agujas del reloj.

Funciones:

1. $f_1(x, y) = \sum$ de las intensidades de los píxeles
2. $f_2(x, y) =$ intensidad máxima de píxel en la imagen
3. $f_3(x, y) =$ ubicación del píxel más brillante
4. $f_4(x, y) =$ promedio de intensidades de los píxeles
5. $f_5(x, y) =$ booleano que detecta si hay al menos una persona en la imagen
6. $f_6(x, y) =$ máscara de segmentación de la imagen

Ejercicio 14.2. Dada la señal unidimensional de entrada $x = [0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0]$ y el kernel 1D $k = (-1, 0, 1)$, resolver:

1. Graficar la señal de entrada. Calcular la señal de salida resultado de aplicar una convolución 1D: $y = x * k$ (mapa de características) sin padding y con stride 1 para los primeros 10 valores de t . ¿Qué tamaño tiene la salida?
2. Interpretar brevemente los valores de la señal de salida y en relación con la detección de características en x , ¿qué observan?

Ejercicio 14.3. Considerar la siguiente matriz de entrada X y el kernel K de una operación de Conv2D simple:

- **Entrada X (5×5):**

$$X = \begin{pmatrix} 1 & 2 & 0 & 0 & 1 \\ 3 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 2 & 0 & 1 & 0 & 1 \\ 3 & 1 & 0 & 0 & 1 \end{pmatrix}$$

- **Kernel K (3×3):**

$$K = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{pmatrix}$$

Responder:

1. ¿Qué tamaño tiene la salida de $Y = X * K$ (stride 1, sin padding)?
2. Calcula la matriz Y resultante.
3. Calcular ahora con las mismas características $Y' = Y * K$, ¿qué sucede con las dimensiones y por qué?

Ejercicio 14.4. Considerar la siguiente matriz de entrada X y el kernel K de una operación de Conv2D simple:

- **Entrada X (5×5):**

$$X = \begin{pmatrix} 1 & 2 & 3 & 0 & 1 \\ 0 & 1 & 2 & 3 & 2 \\ 1 & 0 & 1 & 1 & 0 \\ 2 & 1 & 0 & 2 & 1 \\ 0 & 2 & 3 & 1 & 0 \end{pmatrix}$$

- Kernel K (3×3):

$$K = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

Responder:

1. Calcula $Y = X * K$ (stride 1, sin padding). ¿Qué tamaño tiene Y y cuál es su matriz?
2. Sobre esa Y , aplica max-pool 2×2 con stride 2. ¿Cuál es la forma final y cuánto le dió?
3. Comenta brevemente cómo cambian las dimensiones al combinar Conv2D y Max-Pooling en esta secuencia de operaciones.

Ejercicio 14.5. Considere la siguiente matriz:

$$X = \begin{pmatrix} 1 & 3 & 2 & 4 \\ 5 & 6 & 1 & 0 \\ 2 & 8 & 3 & 1 \\ 0 & 1 & 4 & 7 \end{pmatrix}$$

Le queremos aplicar dos tipos de *pooling* diferentes: *Max-Pooling* y *Average-Pooling*. Aplique ambas operaciones a la matriz usando un kernel de tamaño 2×2 y un stride de 2, ¿que observa en cada caso? ¿cuando serviría usar una sobre la otra?

Ejercicio 14.6. Se tiene una red con dos capas secuenciales:

1. Conv2D de kernel 3×3 , padding = 1, stride = 1.
2. Conv2D de kernel 5×5 , padding = 0, stride = 2.

Responder:

1. Para un píxel en la salida final (capa 2), dibuja esquemáticamente su campo receptivo en la entrada original.
2. Calcula numéricamente el tamaño (ancho \times alto) de ese campo receptivo.
3. Explica en qué influye cada capa (kernel, stride, padding) en la expansión del campo receptivo.

Ejercicio 14.7. Verdadero o Falso:

- (a) En una red convolucional, el *stride* controla cuántas posiciones se desplaza el filtro sobre la imagen de entrada en cada paso.
- (b) Una capa de Max Pooling reduce la resolución de las características, lo que ayuda a disminuir el tamaño de los datos y a prevenir el sobreajuste.
- (c) Una capa convolucional con un kernel de dimensiones 3×3 tiene un total de 9 pesos más 1 sesgo, independientemente de las dimensiones de la imagen a la cuál se le aplica.
- (d) Las capas densas (*fully connected*) en una red convolucional tienen más parámetros que las capas convolucionales, en general, porque cada nodo está conectado a todos los nodos de la capa anterior.
- (e) Las capas convolucionales pueden ser utilizadas para extraer features espaciales y temporales en datos secuenciales, como video o audio.
- (f) Las capas de Max Pooling aumentan la resolución espacial de los *features maps*, lo que aumenta la cantidad de parámetros y hace que la red sea más compleja.
- (g) El Global Average Pooling toma el valor medio de cada *feature map* completo y lo usa como una entrada a la capa densa, reduciendo significativamente el número de parámetros.
- (h) Global Pooling permite a una misma arquitectura procesar imágenes de distintas dimensiones/resoluciones.

Ejercicio 14.8. Responda según corresponda, considerando que todas las triples representan (*alto* \times *ancho* \times *canales/kernels*):

1. Dada una entrada de tamaño $(32 \times 32 \times 16)$, ¿cuál será el tamaño de la salida luego de aplicar una operación de MaxPooling de tamaño 2 y stride 2?

2. Queremos aplicar un filtro con un kernel size de 7×7 y utilizar un stride de 1 para una entrada de $(63 \times 63 \times 16)$, ¿qué tamaño de padding deberemos usar para poder mantener el tamaño de la entrada?
3. ¿Qué stride estamos utilizando si al aplicar un kernel de tamaño 5×5 con padding de 1, obtenemos una salida de tamaño $(41 \times 41 \times 20)$ al aplicársela a una imagen de tamaño $(123 \times 123 \times 3)$?y cuantos filtros se utilizaron?
4. Dada una imagen de tamaño $(64 \times 64 \times 3)$ a la cual se le aplican varios sucesivos pasos/capas convolucionales con kernel de tamaño 3×3 con stride 1 y sin padding, ¿que sucederá con la imagen?
5. En un feature map de dimensión $(28 \times 28 \times 512)$ se aplica una operación de global average pooling, ¿qué tamaño tendrá la salida de esta operación?

Ejercicio 14.9. Considera una red neuronal convolucional que se utiliza para clasificar imágenes en diez clases. La tabla 3 describe la estructura de la red convolucional a utilizar. Complete la tabla con el número de parámetros entrenables asociados a cada capa. Todas las convoluciones se rellenan para conservar las dimensiones de las imágenes.

Número de Capa	Propiedades de la Capa	Cantidad de Parámetros
Capa 0	Imágenes en color de 32x32 y 3 canales	0
Capa 1	Capa convolucional con 40 filtros de 3x3	
Capa 2	Capa convolucional con 40 filtros de 3x3	
Capa 3	Capa de Max Pooling con stride de 2 y tamaño de pooling de 2x2	
Capa 4	Capa convolucional con 64 filtros de 3x3	
Capa 5	Capa convolucional con 64 filtros de 3x3	
Capa 6	Capa densa con 128 unidades	
Capa 7	Capa de salida densa con 10 unidades	
TOTAL:		

Tabla 3: Descripción de la estructura de una red neuronal convolucional y la cantidad de parámetros entrenables asociados a cada capa. El stride de cada capa es 1 salvo que se aclare.