

Redes Neuronales Convolucionales

Juan Gamarra Moreno

Introducción

- Las redes neuronales convolucionales son una poderosa técnica de redes neuronales artificiales. Estas redes preservan la estructura espacial del problema y fueron desarrolladas para tareas de reconocimiento de objetos como el reconocimiento de dígitos escritos a mano.
- Son populares porque se están logrando resultados de vanguardia en tareas difíciles de procesamiento de lenguaje natural y visión por computadora.
- Las redes neuronales convolucionales se usan también en el aprendizaje profundo, también llamadas ConvNets o CNN.

Introducción

Acerca de las ConvNets o CNN es importante saber:

- Los componentes básicos que se utilizan en las CNN, como las capas convolucionales y las capas de agrupaciones.
- Cómo se combinan los bloques de construcción.
- Mejores prácticas para configurar CNN en las tareas de reconocimiento de objetos.

Ejemplo de caso para redes neuronales convolucionales

Dado un conjunto de datos de imágenes en escala de grises con el tamaño estandarizado de 32x32 píxeles cada una, una red neuronal feedforward tradicional requeriría 1.024 pesos de entrada (más un sesgo). Esto es bastante, el aplanamiento de la matriz de píxeles de la imagen a un vector largo de valores de píxeles pierde toda la estructura espacial de la imagen. A menos que se cambie el tamaño de todas las imágenes a la perfección, la red neuronal tendrá grandes dificultades con el problema.

Ejemplo de caso para redes neuronales convolucionales (cont.)

Las redes neuronales convolucionales esperan y preservan la relación espacial entre píxeles al aprender las representaciones de características internas utilizando pequeños cuadrados de datos de entrada. Las características se aprenden y utilizan en toda la imagen, lo que permite que los objetos de las imágenes se desplacen o trasladen en la escena y aún así sean detectables por la red. Esta es la razón por la que la red es tan útil para el reconocimiento de objetos en fotografías, seleccionando dígitos, rostros, objetos, etc. con orientación variable.

Ejemplo de caso para redes neuronales convolucionales (cont.)

En resumen, algunos de los beneficios del uso de redes neuronales convolucionales son:

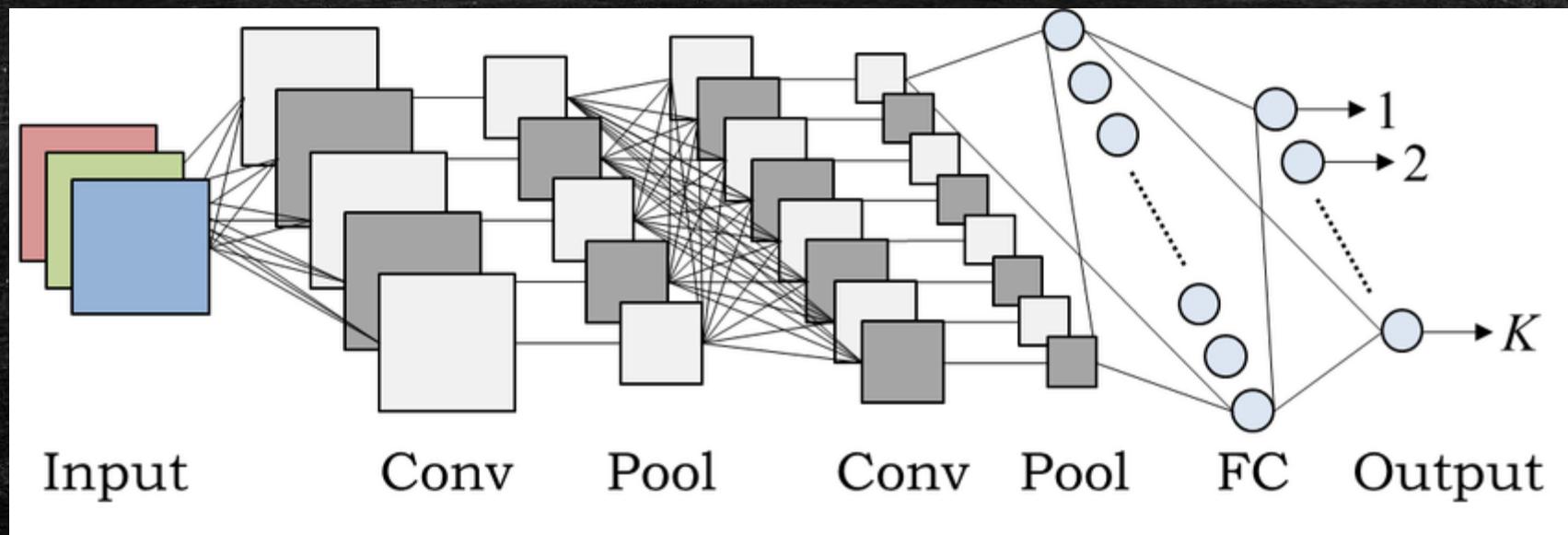
- Usan menos parámetros (pesos) para aprender que una red completamente conectada.
- Están diseñados para ser invariables con respecto a la posición del objeto y la distorsión en la escena.
- Aprenden y generalizan automáticamente características del dominio de entrada.

Bloques de construcción de CNNs

Hay tres tipos de capas en una red neuronal convolucional:

1. Capas convolucionales.
2. Capas de agrupación. (Pooling)
3. Capas totalmente conectadas. (Fully-Connected)

Bloques de construcción de CNNs



Capas Convolucionales

Las capas convolucionales se componen de filtros y mapas de características.

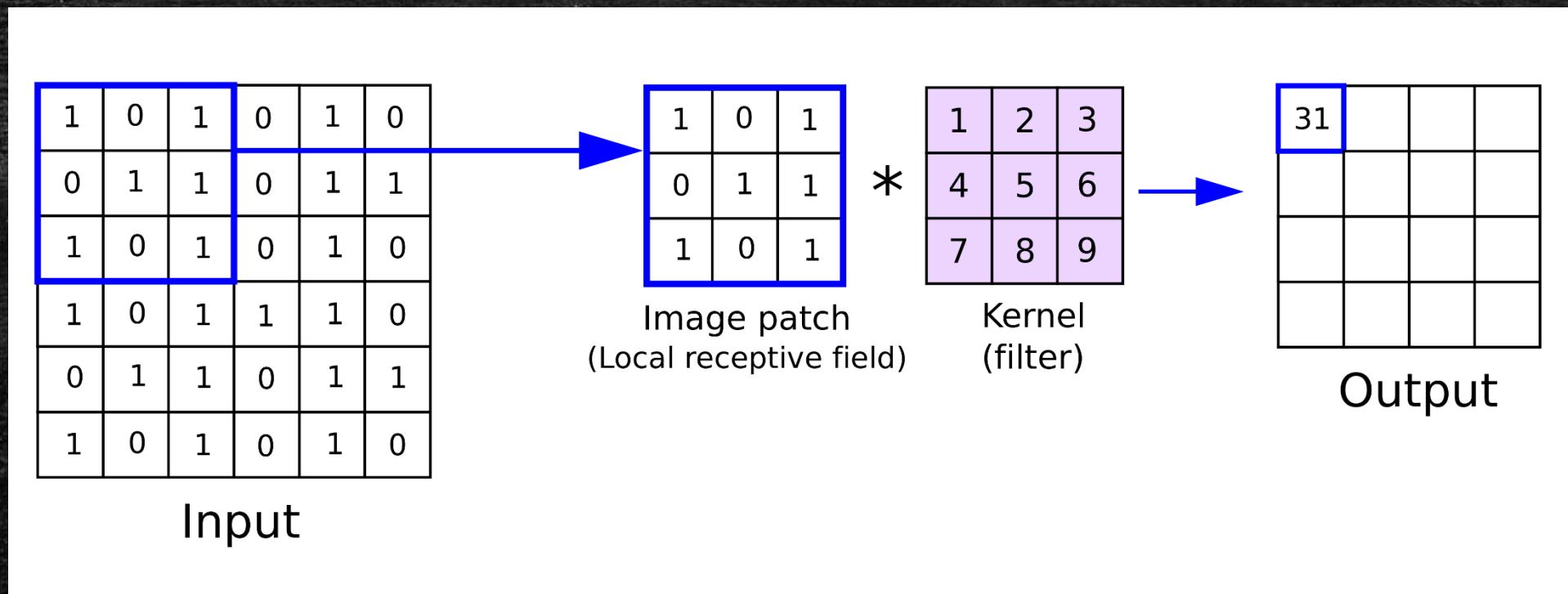
- **Filtros:** Los filtros son esencialmente las neuronas de la capa. Estas tienen a la vez entradas ponderadas y generan un valor de salida como una neurona. El tamaño de entrada es un cuadrado fijo llamado parche o campo receptivo. Si la capa convolucional es una capa de entrada, entonces el parche de entrada serán valores de píxeles. Si están más profundas en la arquitectura de la red, entonces la capa convolucional tomará datos de un mapa de características de la capa anterior.

Capas Convolucionales (cont.)

Las capas convolucionales se componen de filtros y mapas de características.

- **Filtros:** Los filtros son esencialmente las neuronas de la capa. Estas tienen a la vez entradas ponderadas y generan un valor de salida como una neurona. El tamaño de entrada es un cuadrado fijo llamado parche o campo receptivo. Si la capa convolucional es una capa de entrada, entonces el parche de entrada serán valores de píxeles. Si están más profundas en la arquitectura de la red, entonces la capa convolucional tomará datos de un mapa de características de la capa anterior.

Capas Convolucionales (cont.)



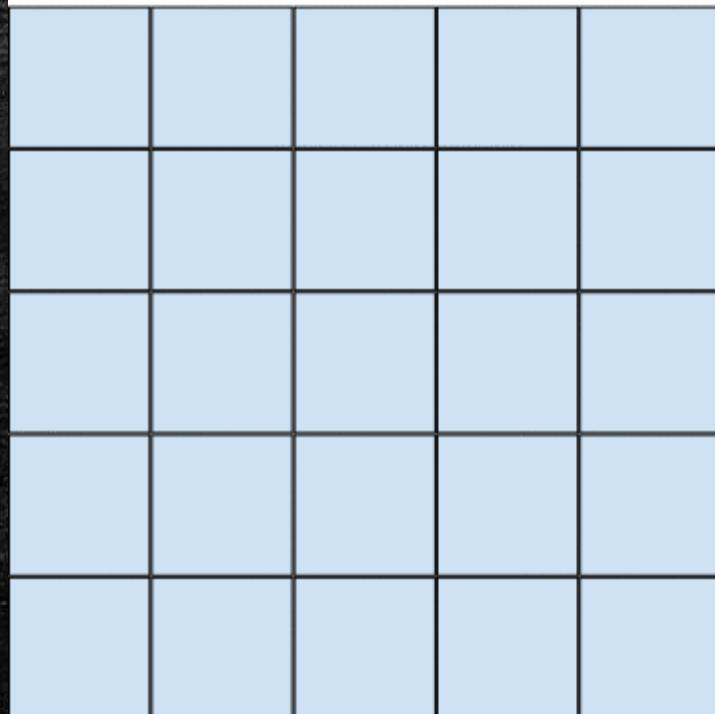
Capas Convolucionales (cont.)

- Mapas de características: El mapa de características es la salida de un filtro aplicado a la capa anterior. Un filtro determinado se superpone en toda la capa anterior y se mueve un pixel a la vez. Cada posición da como resultado una activación de la neurona y la salida se recopila en el mapa de características. Si el campo receptivo se mueve un pixel de activación a activación, entonces el campo se superpondrá con la activación anterior por los valores de entrada (ancho de campo, 1) .

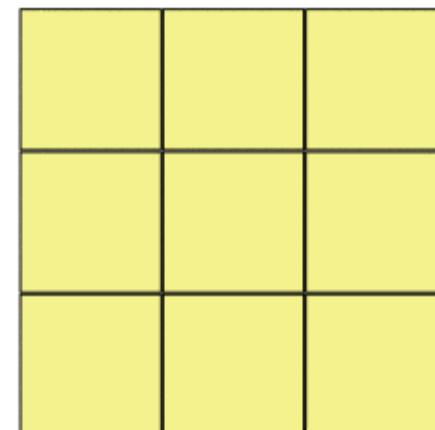
Capas Convolucionales (cont.)

- Mapas de características:

Input Feature Map



Output Feature Map

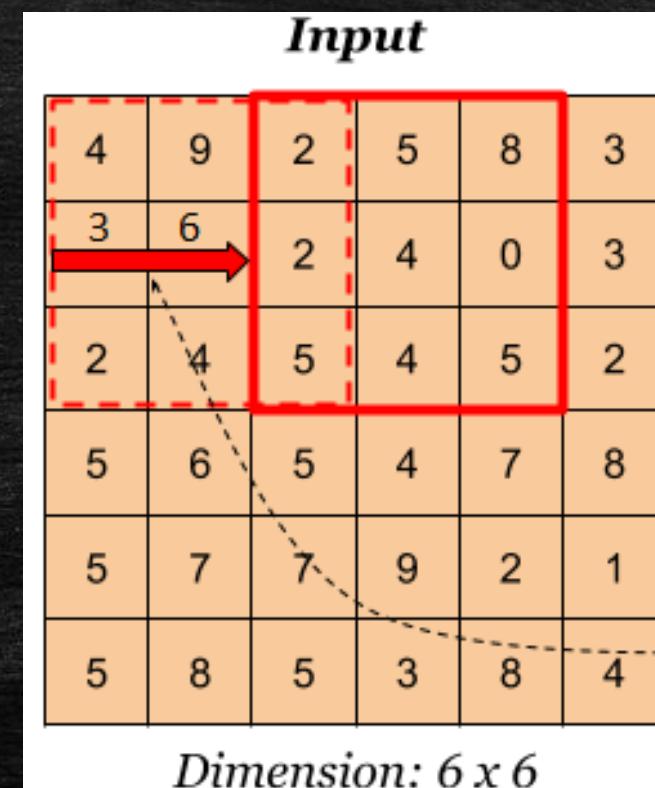
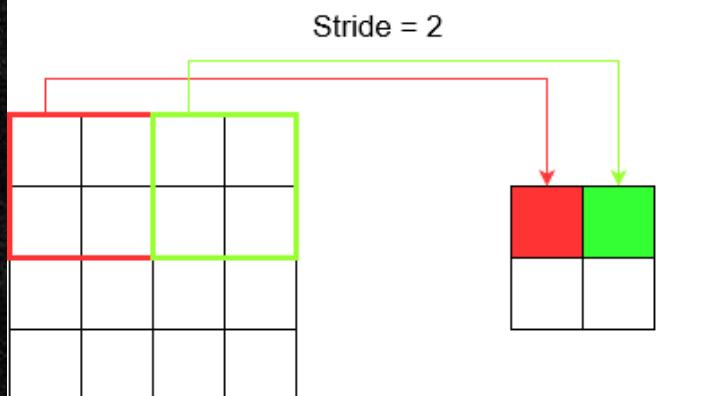
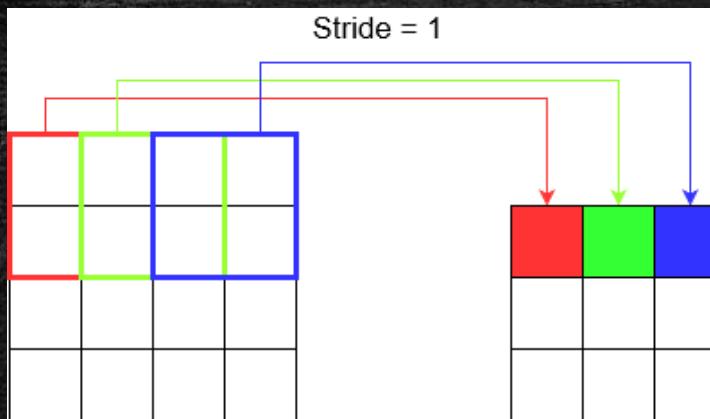


Capas Convolucionales (cont.)

- Mapas de características (cont): La distancia que el filtro se mueve a través de la entrada desde la capa anterior a cada activación se denomina zancada (stride). Si el tamaño de la capa anterior no es claramente divisible por el tamaño del campo receptivo del filtro y el tamaño de la zancada, entonces es posible que el campo receptivo intente leer el borde del mapa de características de entrada. En este caso, se pueden utilizar técnicas como el relleno de ceros para inventar entradas simuladas con valores cero para que las lea el campo receptivo.

Capas Convolucionales (cont.)

- Mapas de características (cont):



Filter

$$\begin{matrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{matrix}$$

*

=

Parameters:

Size: $f = 3$

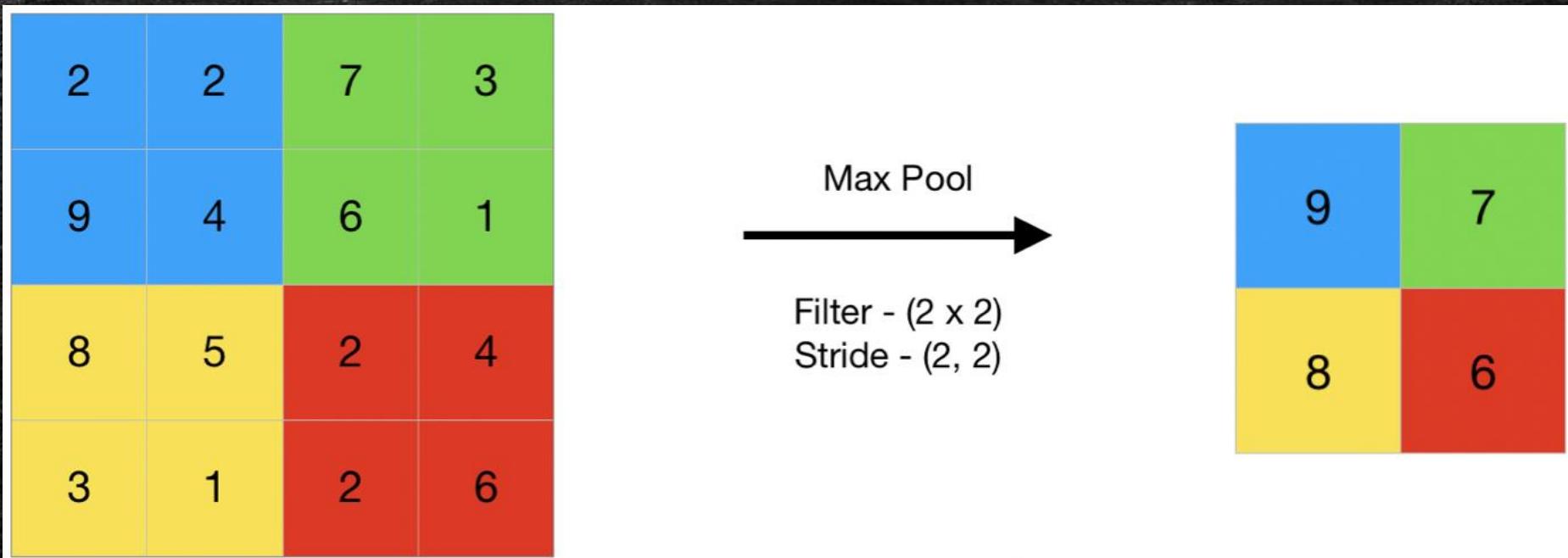
Stride: $s = 2$

Padding: $p = 0$

Capas de agrupación (Pooling)

Las capas de agrupación reduce la muestra del mapa de características de las capas anteriores. Las capas de agrupación siguen una secuencia de una o más capas convolucionales y están destinadas a consolidar las características aprendidas y expresadas en el mapa de características de la capa anterior. Como tal, la agrupación puede considerarse una técnica para comprimir o generalizar las representaciones de características y, en general, reducir el exceso de ajuste de los datos de entrenamiento por parte del modelo.

Capas de agrupación (Pooling)

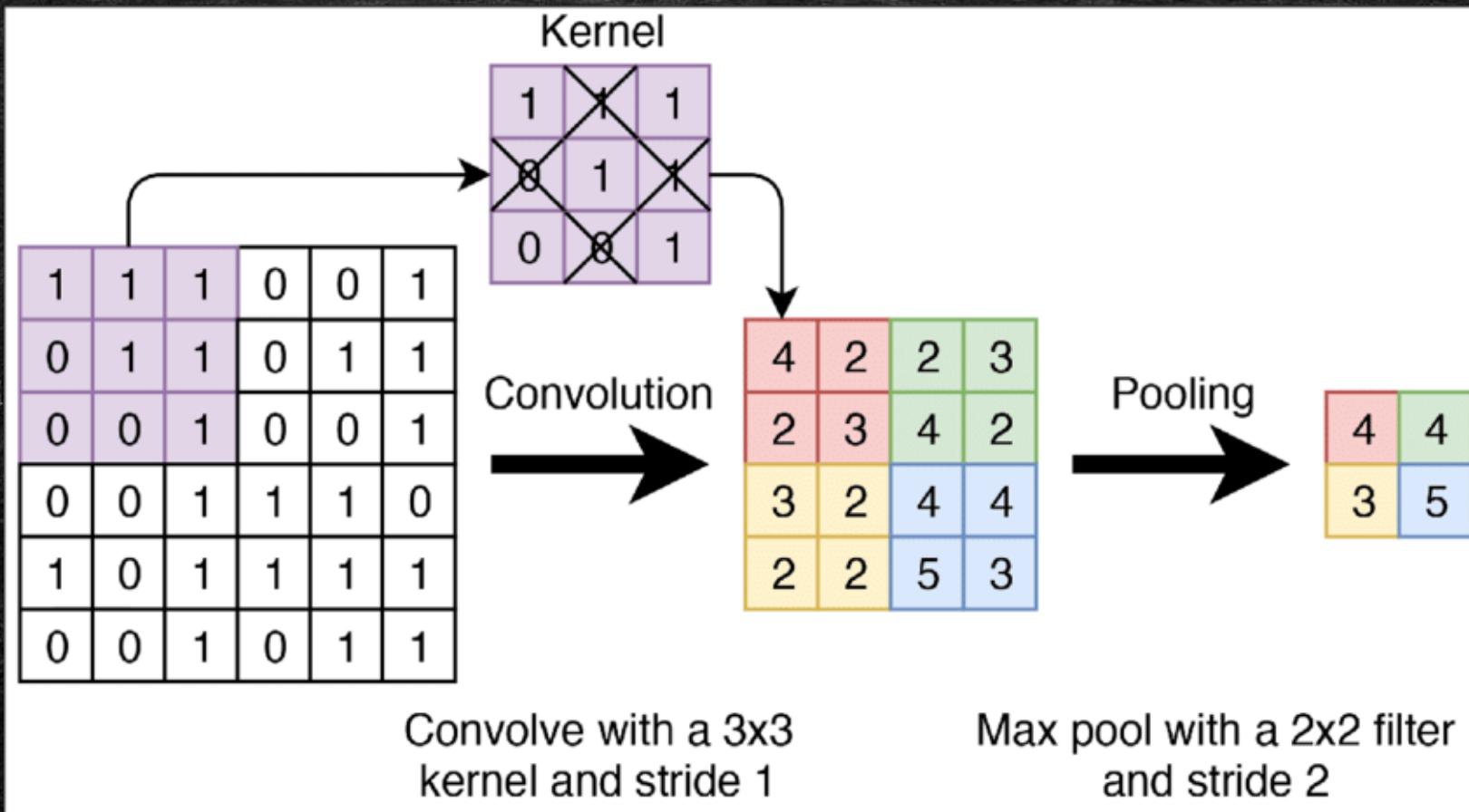


Capas de agrupación (Pooling)(cont.)

Estos también tienen un campo receptivo, a menudo mucho más pequeño que la capa convolucional.

Además, el paso o el número de entradas que se mueve el campo receptivo para cada activación es a menudo igual al tamaño del campo receptivo para evitar cualquier superposición. La agrupación de capas suele ser muy simple, tomando el promedio o el máximo del valor de entrada para crear su propio mapa de características.

Capas de agrupación (Pooling)(cont.)

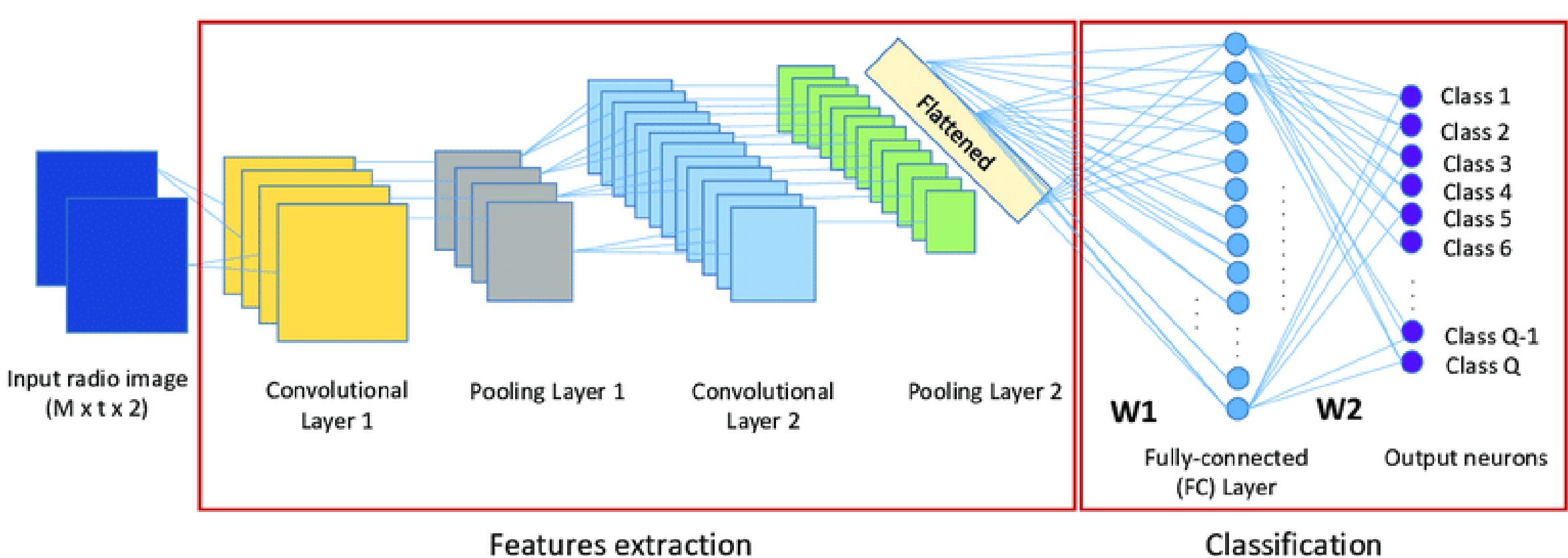


Capa Completamente Conectada

Las capas completamente conectadas son las capas de una red neuronal feedforward plana normal.

Estas capas pueden tener una función de activación no lineal o una activación softmax para generar probabilidades de predicciones de clase. Las capas completamente conectadas se utilizan al final de la red después de que las capas convolucionales y de agrupamiento hayan realizado la extracción y consolidación de características. Se utilizan para crear combinaciones finales no lineales de características y para hacer predicciones por la red.

Capa Completamente Conectada (cont.)



Recomendaciones con las redes neuronales convolucionales

- Dimensiones del campo receptivo de entrada: el valor predeterminado es 2D para imágenes, pero podría ser 1D, para palabras en una oración, o 3D para video que agrega una dimensión de tiempo.
- Tamaño del campo receptivo: el parche debe ser lo más pequeño posible, pero lo suficientemente grande para ver características en los datos de entrada. Es común usar 3×3 en imágenes pequeñas y 5×5 o 7×7 y más en tamaños de imagen más grandes.

Recomendaciones con las redes neuronales convolucionales (cont)

- Ancho de zancada (stride): utilice la zancada predeterminada de 1. Es fácil de entender y no necesita relleno para manejar el campo receptivo que sale del borde de sus imágenes. Esto podría aumentarse a 2 o más para imágenes más grandes.
- Número de filtros: los filtros son los detectores de características. Generalmente se utilizan menos filtros en la capa de entrada y cada vez se utilizan más filtros en las capas más profundas.

Recomendaciones con las redes neuronales convolucionales (cont.)

- Relleno: se establece en cero y se denomina relleno de cero al leer datos que no son de entrada. Esto es útil cuando no puede o no desea estandarizar los tamaños de la imagen de entrada o cuando desea utilizar tamaños de campo y zancada receptivos que no dividen claramente el tamaño de la imagen de entrada.

Recomendaciones con las redes neuronales convolucionales (cont.)

- Pooling: Pooling es un proceso destructivo o de generalización para reducir el sobreajuste. El tamaño del campo receptivo casi siempre se establece en 2×2 con un paso de 2 para descartar el 75% de las activaciones de la salida de la capa anterior.
- Preparación de datos: considere la posibilidad de estandarizar los datos de entrada, tanto las dimensiones de las imágenes como los valores de píxeles.

Recomendaciones con las redes neuronales convolucionales (cont.)

- Arquitectura de patrones: es común crear patrones de capas en la arquitectura de su red. Puede ser una, dos o varias capas convolucionales seguidas de una capa de agrupación. A continuación, esta estructura se puede repetir una o más veces. Finalmente, las capas completamente conectadas son a menudo solamente usadas en el extremo de salida y se puede apilar uno, dos o más en profundidad.

Recomendaciones con las redes neuronales convolucionales (cont.)

- Dropout: Las CNN tienen la costumbre de sobreajustarse, incluso con capas de agrupamiento. Dropout debe usarse, por ejemplo, entre capas completamente conectadas y quizás después de capas de agrupamiento.