

Resumen semana 10

Jose Ignacio Calderón Díaz

I. RESPUESTAS DEL QUIZ

- 1) Función de activación para clasificación multiclase es la funcion softmax
- 2) Técnicas: time warping, mascara de tiempo, mascara de frecuencia
- 3) Derivadas que al ser multiplicadas dan el resultado:

$$\frac{\partial L}{\partial a} = 2(a^l - y)$$

$$\frac{\partial a}{\partial z} = \sigma'(z)$$

$$\frac{\partial z}{\partial w} = a_k^{l-1}$$

II. CONVOLUTIONAL NEURAL NETWORKS

Hemos visto hasta ahora:

- Transformamos con una serie de capas ocultas. Transforma la data de entrada a una salida.
- Una capa de salida.
- Aprendemos la secuencia de pixeles:
Errores si movemos los objetos de lugar.
Si aplicamos rotación al objeto.

Las imagenes actuales tienen al menos 3 canales de color. Por ejemplo el dataset CIFAR-10 tiene 10 clases para clasificar. Tamaño $32 \times 32 \times 3$. Si mantenemos el enfoque de fully conected, tendríamos un total de 3072 pesos. No es escalable, cuando las imágenes son mas grandes vamos a tener el mismo problema (costoso computacionalmente).

Si se tiene una imagen de 200×200 . Se tendrían $200 \times 200 \times 3 = 120.000$ pesos.

A. Desventajas de Fully connected

- Requiere gran cantidad de parámetros
- Su entrenamiento y rendimiento es lento

B. Ventajas de ConvNet

- Neuronas organizadas en 3 dimensiones: largo, ancho y profundidad (canales)
- Estas neuronas estarán conectadas a pequeñas regiones de la capa anterior, en lugar de fully-connected.
- Reducir el tamaño de la imagen a un vector
- Capa de salida: $1 \times 1 \times 10$
- La idea es que sirvan de extractor de características de las imágenes.
- Se especializan en ciertas partes/colores
- En el vector se realiza el clasificador

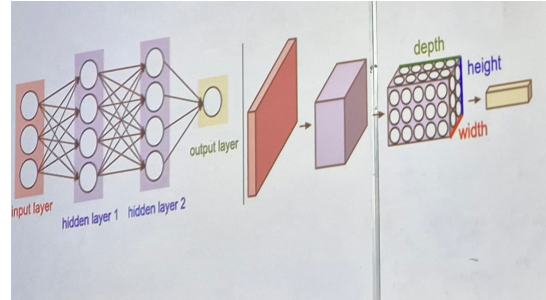


Fig. 1. ConvNet

III. ARQUITECTURA DE UNA RED CONVOLUCIONAL

Incluye tres tipos principales de capas:

1) Capas convolucionales:

- Entrada: imagen con dimensiones ancho, alto y canales.
- Se aplican n filtros para extraer características.
- Cada filtro aprende un conjunto de pesos (por ejemplo, $5 \times 5 \times 3 = 75$ pesos).
- Se aplica una función de activación y se generan mapas de características (feature maps).

2) Capas de pooling:

- Reducen las dimensiones espaciales (ancho y alto).
- No afectan la cantidad de canales.
- Ejemplo: de $(32, 32, 12)$ a $(16, 16, 12)$.
- No tienen parámetros.

3) Capas fully connected:

- Transforma el vector en probabilidades de pertenecer a cada clase.
- Realiza la clasificación final.

IV. ALEXNET

- Arquitectura que demuestra el potencial del deep learning.
- Disminuye la cantidad de parámetros y mejora el rendimiento.

A. Aplicaciones

- Clasificación de imágenes.
- Segmentación de objetos.
- Segmentación de instancias.
- Procesamiento de imágenes.

V. DETALLES DE LA CAPA DE CONVOLUCIÓN

Filtro o kernel:

- Matriz bidimensional (ej. 3×3 , 5×5).

- Se desliza sobre la imagen y realiza convolución.
- Ejemplo: filtro gaussiano produce desenfoque y reduce ruido.

VI. CAMPOS RECEPTIVOS LOCALES

- Cada neurona se enfoca en una región local de la imagen.
- Ejemplo: imagen de entrada $32 \times 32 \times 3$ con campo receptivo 5×5 .
- Cada neurona tiene $5 \times 5 \times 3 = 75$ pesos.

VII. STRIDE Y PADDING

Stride:

- Número de pasos con los que se mueve el filtro sobre la imagen.
- Afecta la dimensión de la salida.

Padding:

- Añade píxeles alrededor de la imagen para mantener su tamaño original tras la convolución.
- Útil para preservar la dimensionalidad.
-

$$\text{padding} = \frac{k - 1}{2}$$

donde k es el tamaño del filtro.

VIII. CÁLCULO DE DIMENSIONES FINALES

Para determinar el tamaño de salida tras la convolución se utiliza la fórmula correspondiente, que depende de:

- m , Cantidad de píxeles en una fila/columna.
- k , tamaño del kernel.
- p , tamaño del padding.
- s , cantidad de pasos

$$\text{Dimensión resultado} = \frac{(m - k + 2p)}{s} + 1$$