Ejercicio 1: CNN Fundamentals (LeNet)

Jordi Blasco Lozano

Planteamiento

Se considera la arquitectura *LeNet* mostrada en el enunciado, con entrada una imagen en escala de grises de tamaño 28×28. Se pide:

a) Calcular el tamaño de las activaciones (dimensiones y número de activaciones) y el número de parámetros entrenables de la red.

b) Calcular el *forward pass* y el *backward pass* de la primera capa convolucional (expresiones y dimensiones).

Arquitectura LeNet considerada

La red (capas y *hyperparámetros*) es:

- **Entrada:** $28 \times 28 \times 1$.
- **C1:** Convolución 5×5, $c_1 = 6$ filtros, **stride** $s = 1$, **padding** $p = 2 \Rightarrow 28 \times 28 \times 6$.
- **S2:** *AvgPool* 2×2, $s = 2 \Rightarrow 14 \times 14 \times 6$.
- **C3:** Convolución 5×5, $c_3 = 16$ filtros, $s = 1$, $p = 0 \Rightarrow 10 \times 10 \times 16$.
- **S4:** *AvgPool* 2×2, $s = 2 \Rightarrow 5 \times 5 \times 16$.
- **F5:** Capa densa (equivalente a conv 5×5 sobre $5 \times 5 \times 16$): 120 neuronas.
- **F6:** Capa densa: 84 neuronas.
- **Salida:** Capa densa: 10 neuronas.

Notación y fórmulas

Tamaño de mapa de características

Para una convolución 2D con entrada $h \times w$, filtro $f \times f$, padding p y stride s:

$$h_{\text{out}} = \left\lfloor \frac{h + 2p - f}{s} \right\rfloor + 1, \qquad w_{\text{out}} = \left\lfloor \frac{w + 2p - f}{s} \right\rfloor + 1.$$

Parámetros entrenables

En una capa convolucional con c_{in} canales de entrada, c_{out} filtros y kernel $f \times f$:

$$\#\text{params} = \big(f \cdot f \cdot c_{\text{in}} + 1\big)\, c_{\text{out}},$$

donde el $+1$ corresponde al sesgo (*bias*) por filtro. En una capa densa con n_{in} entradas y n_{out} salidas:

$$\#\text{params} = n_{\text{in}}\, n_{\text{out}} + n_{\text{out}}.$$

1. Apartado a: Tamaño de activación y número de parámetros

Cálculo de dimensiones por capa

- **C1:** $h = w = 28,\ f = 5,\ p = 2,\ s = 1$:
$$h_{\text{out}} = \frac{28 + 2 \cdot 2 - 5}{1} + 1 = 28, \qquad w_{\text{out}} = 28 \Rightarrow 28 \times 28 \times 6.$$
- **S2 (pool):** reduce a la mitad con 2×2 y $s = 2$:
$$28 \times 28 \times 6 \ \longrightarrow\ 14 \times 14 \times 6.$$
- **C3:** $h = w = 14,\ f = 5,\ p = 0,\ s = 1$:
$$h_{\text{out}} = \frac{14 - 5}{1} + 1 = 10, \qquad w_{\text{out}} = 10 \Rightarrow 10 \times 10 \times 16.$$
- **S4 (pool):**
$$10 \times 10 \times 16 \ \longrightarrow\ 5 \times 5 \times 16.$$
- **F5:** aplanado $5 \cdot 5 \cdot 16 = 400$ entradas $\to 120$.
- **F6:** $120 \to 84$.
- **Salida:** $84 \to 10$.

Tabla resumen (activaciones y parámetros)

Sea "#Act." el número total de activaciones (#elementos del tensor de salida de la capa).

Capa	**Salida**	**#Act.**	**#params**	**Cálculo params**
Entrada	$28 \times 28 \times 1$	$28 \cdot 28 \cdot 1 = 784$	0	–
C1 (conv)	$28 \times 28 \times 6$	$28 \cdot 28 \cdot 6 = 4704$	156	$(5 \cdot 5 \cdot 1 + 1) \cdot 6$
S2 (pool)	$14 \times 14 \times 6$	$14 \cdot 14 \cdot 6 = 1176$	0	–
C3 (conv)	$10 \times 10 \times 16$	$10 \cdot 10 \cdot 16 = 1600$	2416	$(5 \cdot 5 \cdot 6 + 1) \cdot 16$
S4 (pool)	$5 \times 5 \times 16$	$5 \cdot 5 \cdot 16 = 400$	0	–
F5 (densa)	120	120	48120	$400 \cdot 120 + 120$
F6 (densa)	84	84	10164	$120 \cdot 84 + 84$
Salida (densa)	10	10	850	$84 \cdot 10 + 10$
Total	–	8878	**61706**	–

2. Apartado b: Forward y Backward de la primera capa convolucional (C1)

Forward pass (C1)

Sea la entrada $X \in \mathbb{R}^{28 \times 28 \times 1}$. En C1 se usa:
$$f = 5, \quad s = 1, \quad p = 2, \quad c_{\text{out}} = 6.$$
Tras aplicar padding, definimos $X_{\text{pad}} \in \mathbb{R}^{32 \times 32 \times 1}$ añadiendo $p = 2$ ceros en cada borde.

Los pesos son $W^{[1]} \in \mathbb{R}^{5 \times 5 \times 1 \times 6}$ y los sesgos $b^{[1]} \in \mathbb{R}^{6}$. La salida preactivación es $Z^{[1]} \in \mathbb{R}^{28 \times 28 \times 6}$ y se calcula elemento a elemento como:
$$Z^{[1]}_{i,j,k} = \sum_{m=0}^{4} \sum_{n=0}^{4} \sum_{c=1}^{1} W^{[1]}_{m,n,c,k}\, X_{\text{pad}\,i+m,\,j+n,\,c} \ +\ b^{[1]}_{k}, \qquad i, j \in \{0, \ldots, 27\},\ k \in \{1, \ldots, 6\}.$$
Finalmente, la activación de salida de la capa es
$$A^{[1]} = \sigma\Big(Z^{[1]}\Big) \in \mathbb{R}^{28 \times 28 \times 6},$$
donde $\sigma(\cdot)$ es la función de activación elegida (en las transparencias se denota genéricamente por σ).

Backward pass (C1)

Sea $dA^{[1]} = \frac{\partial \mathcal{L}}{\partial A^{[1]}}$ el gradiente que llega desde la capa siguiente (pooling S2). Primero se deriva a través de la activación:

$$dZ^{[1]} = dA^{[1]} \odot \sigma'\Big(Z^{[1]}\Big), \qquad dZ^{[1]} \in \mathbb{R}^{28\times 28\times 6}.$$

Gradiente respecto al sesgo ($db^{[1]}$). Como hay un sesgo por filtro:

$$db^{[1]}_k = \sum_{i=0}^{27}\sum_{j=0}^{27} dZ^{[1]}_{i,j,k}, \qquad db^{[1]} \in \mathbb{R}^6.$$

Gradiente respecto a los pesos ($dW^{[1]}$). Cada peso acumula contribuciones de todas las posiciones donde se usó el kernel:

$$dW^{[1]}_{m,n,c,k} = \sum_{i=0}^{27}\sum_{j=0}^{27} X_{\text{pad}\,i+m,\,j+n,\,c}\; dZ^{[1]}_{i,j,k}, \qquad dW^{[1]} \in \mathbb{R}^{5\times 5\times 1\times 6}.$$

Gradiente respecto a la entrada (dX). Primero se obtiene el gradiente sobre la entrada con padding:

$$dX_{\text{pad}} \in \mathbb{R}^{32\times 32\times 1}.$$

Como en el forward se usa correlación (no se rota el kernel), el gradiente sobre la entrada se expresa como una convolución "completa" de dZ con el kernel rotado 180°:

$$dX_{\text{pad}}(:,:,c) = \sum_{k=1}^{6} dZ^{[1]}(:,:,k) \;*\; \text{rot180}\Big(W^{[1]}(:,:,c,k)\Big),$$

donde $*$ denota convolución 2D completa (con el padding necesario para recuperar tamaño 32×32) y $\text{rot180}(\cdot)$ rota el kernel 180°.

Finalmente, se elimina el padding para recuperar el gradiente de la entrada original:

$$dX = dX_{\text{pad}}\big[\,p{:}p+28-1,\; p{:}p+28-1,\; :\,\big] \in \mathbb{R}^{28\times 28\times 1}.$$