Name: Juan David Martinez Vargas

1 Convolución 2D

Backpropagation a través de una capa convolucional es una de las operaciones más fundamentales en el aprendizaje profundo. En este documento la derivaremos, la implementaremos y proporcionaremos algo de intuición sobre lo que está sucediendo dentro de las capas convolucionales de las redes neuronales.

1.1 Forward Pass

Revisaremos la operación básica de una capa convolucional, que en realidad implementa la correlación cruzada en bibliotecas modernas como Pytorch. Para facilitar la comprensión, trabajaremos con un pequeño ejemplo numérico. Imagina un kernel $\mathbf{W}$ simple de 2×2 y una imagen $\mathbf{X}$ simple de 3×3 :

$$\mathbf{W} = \begin{pmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_{00} & x_{01} & x_{02} \\ x_{10} & x_{11} & x_{12} \\ x_{20} & x_{21} & x_{22} \end{pmatrix}$$

El mapa de características $\mathbf{Z}$, en este caso de 2×2 estará dado por la convolución (correlación cruzada) entre la imágen y el kernel:

$$\mathbf{Z} = \mathbf{X} * \mathbf{W} \tag{1}$$

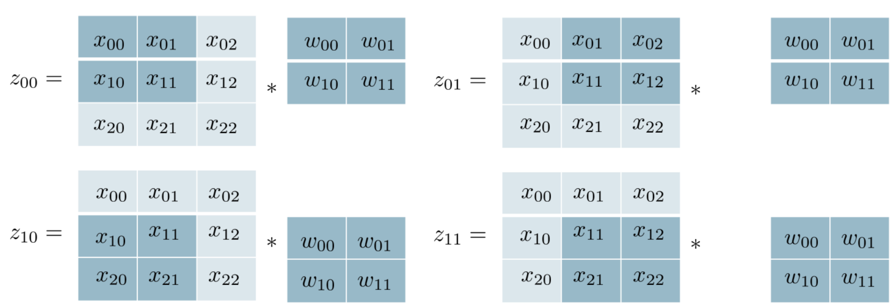

Figure 1: Convolución 2D

Cada uno de los valores del mapa de características está dado por las siguientes ecuaciones:

$$z_{00} = w_{00} \cdot x_{00} + w_{01} \cdot x_{01} + w_{10} \cdot x_{10} + w_{11} \cdot x_{11} \tag{2}$$
$$z_{01} = w_{00} \cdot x_{01} + w_{01} \cdot x_{02} + w_{10} \cdot x_{11} + w_{11} \cdot x_{12} \tag{3}$$
$$z_{10} = w_{00} \cdot x_{10} + w_{01} \cdot x_{11} + w_{10} \cdot x_{20} + w_{11} \cdot x_{21} \tag{4}$$
$$z_{11} = w_{00} \cdot x_{11} + w_{01} \cdot x_{12} + w_{10} \cdot x_{21} + w_{11} \cdot x_{22} \tag{5}$$

De forma general, podemos escribir estas ecuaciones de la siguiente forma:

$$z_{r,c} = \sum_{a=0}^{k_1-1} \sum_{b=0}^{k_2-1} x_{r+a,c+b} w_{a,b} \tag{6}$$

donde k_1 es el número de filas y k_2 es el número de columnas del kernel.

1.2 Backward pass

En el forward pass nos intereza calcular el mapa de características $\mathbf{Z}$ a partir de los datos de entrada $\mathbf{X}$ y $\mathbf{W}$. En el backward pass recibimos el gradiente de la función de costo w.r.t mapa de características $\frac{\partial \mathcal{L}}{\partial \mathbf{Z}}$ y queremos calcular el gradiente de la función de costo con respecto al kernel $\frac{\partial \mathcal{L}}{\partial \mathbf{W}}$ y a los datos $\frac{\partial \mathcal{L}}{\partial \mathbf{X}}$. Para esto, debemos calcular los gradientes internos $\frac{\partial \mathbf{Z}}{\partial \mathbf{X}}$ y $\frac{\partial \mathbf{Z}}{\partial \mathbf{W}}$

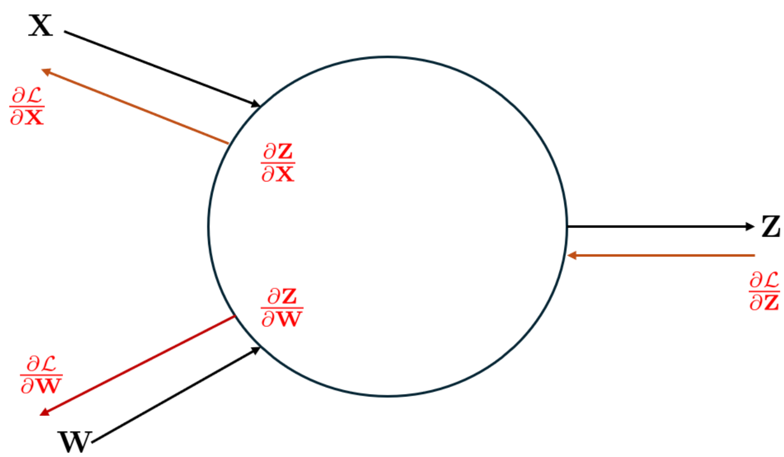

Figure 2: Backweard pass en la Convolución 2D

Si calculamos por ejemplo la derivada:

$$\frac{\partial z_{00}}{\partial w_{00}} = x_{00}$$

$$\frac{\partial z_{10}}{\partial w_{01}} = x_{11}$$

En todos los posibles casos podemos notar que tendremos un solo término diferente de 0:

$$\frac{\partial z_{r,c}}{\partial w_{a',b'}} = x_{r+a',c+b'}$$

Ahora, utilizando la regla de la cadena para calcular $\frac{\partial \mathcal{L}}{\partial w_{a',b'}}$ obtenemos:

$$\frac{\partial \mathcal{L}}{\partial w_{a',b'}} = \sum_{r=0}^{k_1-1} \sum_{c=0}^{k_2-1} \frac{\partial \mathcal{L}}{\partial z_{r,c}} \frac{\partial z_{r,c}}{\partial w_{a',b'}} \tag{7}$$

$$= \sum_{r=0}^{k_1-1} \sum_{c=0}^{k_2-1} \frac{\partial \mathcal{L}}{\partial z_{r,c}} x_{r+a',c+b'} \tag{8}$$

Un caso particular sería calcular $\frac{\partial \mathcal{L}}{\partial w_{0,0}}$. De las ecuaciones 2 a 5 y utilizando la regla de la cadena, podemos intuir que:

$$\frac{\partial \mathcal{L}}{\partial w_{0,0}} = \frac{\partial \mathcal{L}}{\partial z_{0,0}} \frac{\partial z_{0,0}}{\partial w_{0,0}} + \frac{\partial \mathcal{L}}{\partial z_{0,1}} \frac{\partial z_{0,1}}{\partial w_{0,0}} + \frac{\partial \mathcal{L}}{\partial z_{1,0}} \frac{\partial z_{1,0}}{\partial w_{0,0}} + \frac{\partial \mathcal{L}}{\partial z_{1,1}} \frac{\partial z_{1,1}}{\partial w_{0,0}} \tag{9}$$

Reemplazando los gradientes internos obtenemos:

$$\frac{\partial \mathcal{L}}{\partial w_{0,0}} = \frac{\partial \mathcal{L}}{\partial z_{0,0}} x_{0,0} + \frac{\partial \mathcal{L}}{\partial z_{0,1}} x_{0,1} + \frac{\partial \mathcal{L}}{\partial z_{1,0}} x_{1,0} + \frac{\partial \mathcal{L}}{\partial z_{1,1}} x_{1,1} \tag{10}$$

Esta doble suma sobre escalares multiplicados debería recordarnos a la correlación cruzada, porque esta es su definición. Por lo tanto, podemos reinterpretar la ecuación anterior como una correlación cruzada entre el gradiente de la capa superior y la imagen de entrada:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}} = \mathbf{X} * \frac{\partial \mathcal{L}}{\partial \mathbf{Z}} \tag{11}$$