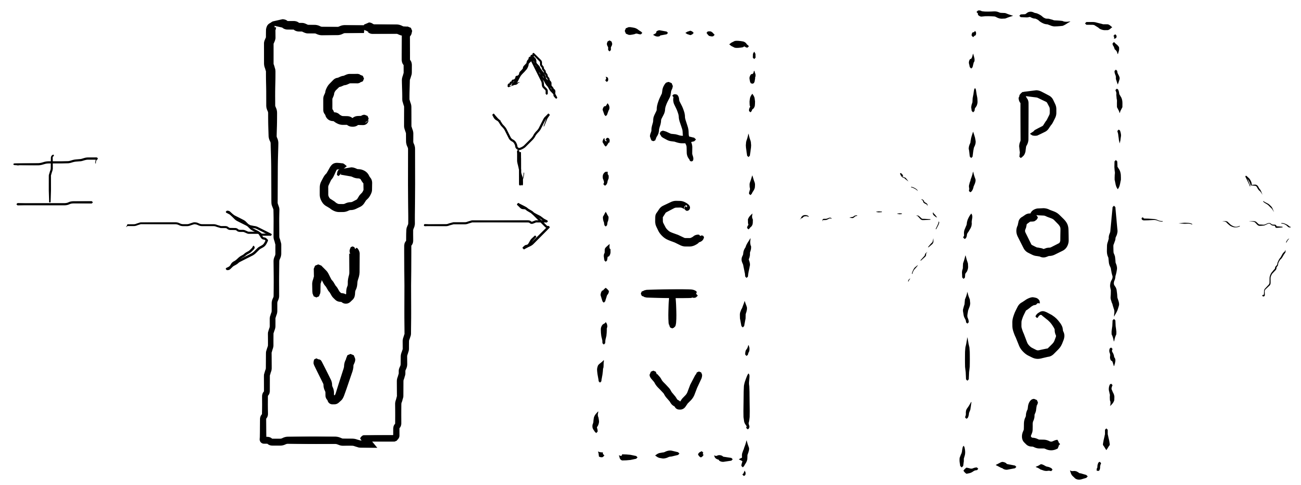
Back-propagation en una capa de convolución



Consideraciones:

- I 1 solo canal (escala de grises) $\Rightarrow I\,[H, W]$
- w 1 solo filtro (kernel) $\Rightarrow w\,[k_1, k_2]$
- padding $\rightarrow$ valid $\Rightarrow \hat{Y}\,[H-k_1+1,\ W-k_2+1]$
- Stride $\rightarrow 1$

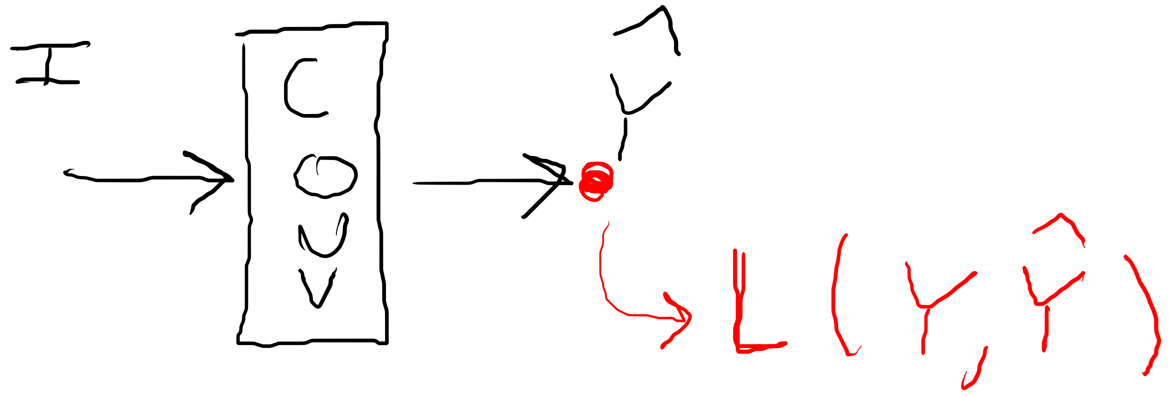
Operaciones

convolución $$(w * I)_{i,j} = \sum_{m=0}^{k_1-1}\sum_{n=0}^{k_2-1} I(i-m, j-n)\, w(m,n)$$

correlación cruzada $$(w \circledast I)_{i,j} = \sum_{m=0}^{k_1-1}\sum_{n=0}^{k_2-1} I(i+m, j+n)\, w(m,n)$$

Nuestra capa de CONV realiza la sig. operación:

$$\hat{Y}(i,j) = (W \circledast I)_{(i,j)} + b = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} I(i+m, j+n)\, W(m,n) + b$$



Nuestra función de costo L en la salida de dicha capa va a ser función de el valor verdadero Y y la salida de nuestra capa $\hat{Y}$. Queremos hallar:

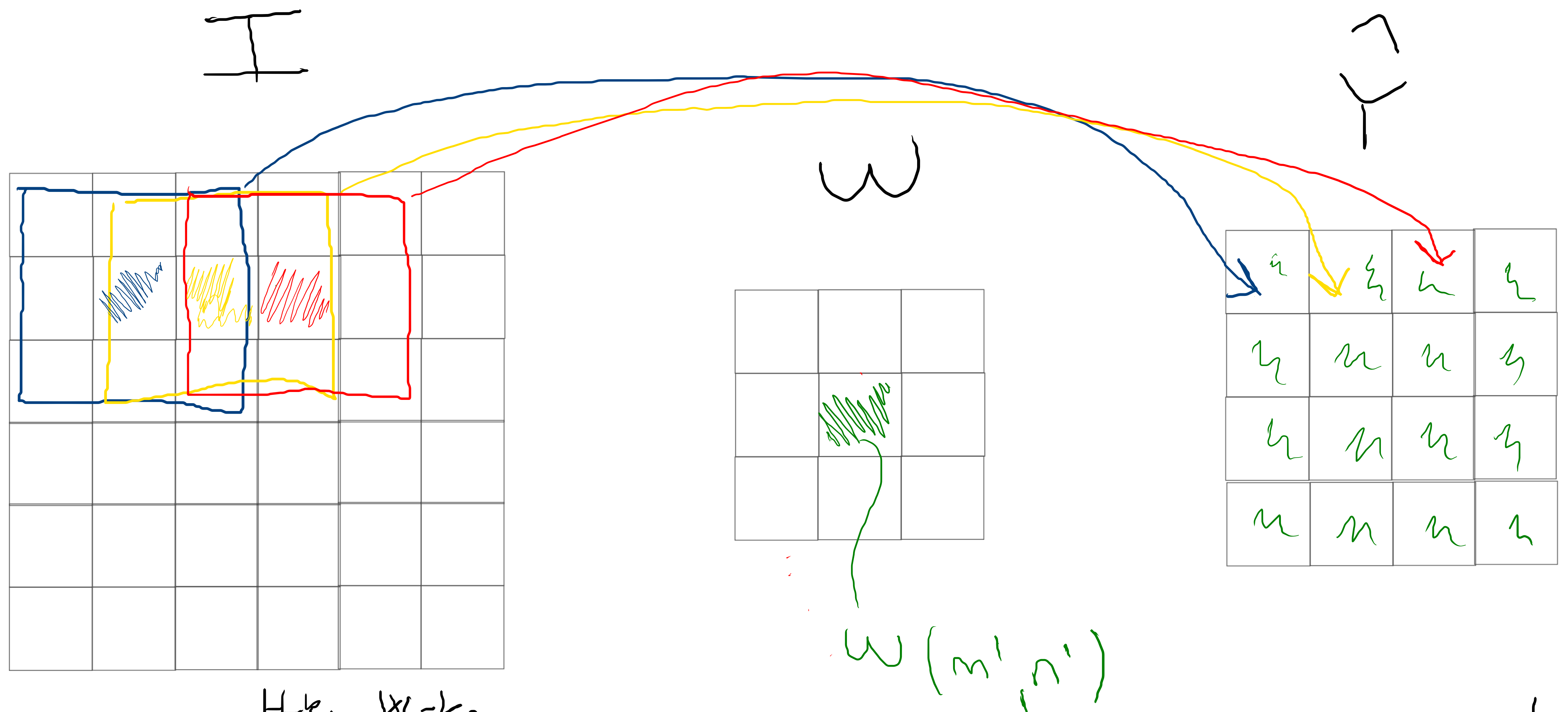
1° $\dfrac{\partial L}{\partial W}$ → para poder otimizar los parámetros del kernel

2° $\dfrac{\partial L}{\partial I}$ → para poder pasar el gradiente a la capa previa

tambien será necesario hallar $\dfrac{\partial L}{\partial b}$ para optimizar b...

Comencemos con $\frac{\partial L}{\partial w}$

¿Como afecta a la salida $\hat{Y}$ variar 1 peso $w(m',n')$?



$\Delta w(m',n')$ afecta a todos los pixeles de $\hat{Y}$

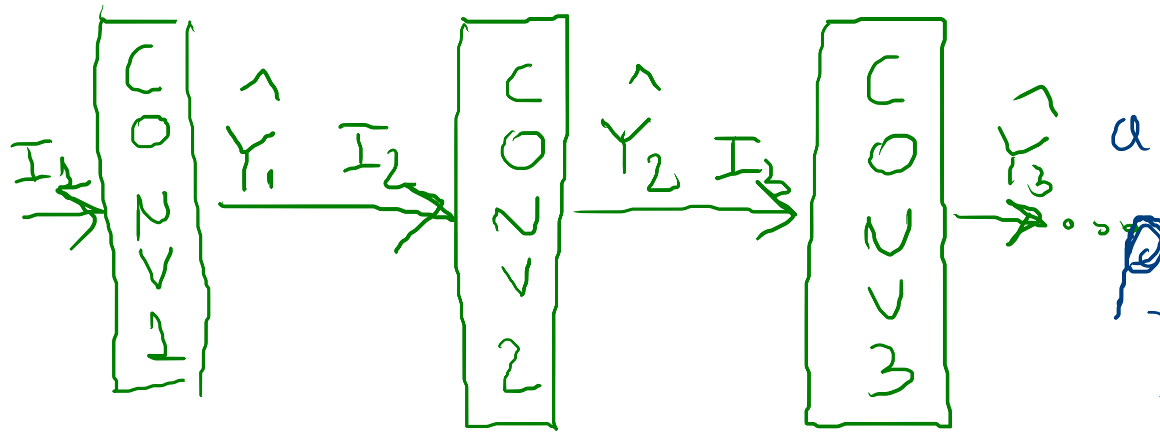
$\Downarrow$

L se ve afectado por cada pixel de $\hat{Y}$

$$\frac{\partial L}{\partial w(m',n')} = \underbrace{\sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \frac{\partial L}{\partial \hat{Y}(i,j)}}_{\text{sumatoria en todo } \hat{Y}} \cdot \frac{\partial \hat{Y}(i,j)}{\partial w(m',n')}$$

entonces :

$$\frac{\partial L}{\partial w(m',n')} = \sum_{i=0}^{H-k_1}\sum_{j=0}^{W-k_2} \underbrace{\frac{\partial L}{\partial \hat{Y}(i,j)}}_{\substack{\text{Esto me lo aporta la capa} \\ \text{posterior; es el } \frac{\partial L}{\partial I} \text{ pero de la} \\ \text{capa que sigue}}} \cdot \underbrace{\frac{\partial \hat{Y}(i,j)}{\partial w(m',n')}}_{\substack{\text{acá tenemos} \\ \text{que trabajar}}} \qquad (1)$$



La entrada de una capa es la salida de su previa

$\hat{Y}(i,j)$ por definición...

$$\frac{\partial \hat{Y}(i,j)}{\partial w(m',n')} = \frac{\partial}{\partial w(m',n')}\left[\underbrace{\sum_{m=0}^{k_1-1}\sum_{n=0}^{k_2-1}}_{\substack{\text{sumo en} \\ \text{todo el kernel } w}} I(i+m, j+n)\cdot w(m,n) + b\right]$$

$$\frac{\partial \hat{Y}(i,j)}{\partial w(m',n')} = \frac{\partial}{\partial w(m',n')} \left[\underbrace{\sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} I(i+m,\, j+n) \cdot w(m,n) + b}_{\text{Hay 1 solo termino donde}} \right]$$

$m = m'$
$n = n'$

$$= \frac{\partial}{\partial w(m',n')} \left[\cdots + I(i+m',\, j+n') \cdot w(m',n') + \cdots + b \right]$$

Al derivar se eliminan todos los terminos que no tienen $w(m',n')$

$$\frac{\partial \hat{Y}(i,j)}{\partial w(m',n')} = I(i+m',\, j+n')$$

sustituimos en 1

$$\frac{\partial L}{\partial w(m',n')} = \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} \frac{\partial L}{\partial \hat{Y}(i,j)} \cdot \overbrace{I(i+m', j+n')}^{\text{lo hallado}} \qquad (1)$$

reordenamos

$$\frac{\partial L}{\partial w(m',n')} = \underbrace{\sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2}}_{\text{suma en } \hat{Y}} I(m'+i, n'+j) \frac{\partial L}{\partial \hat{Y}(i,j)} = \left(\frac{\partial L}{\partial Y} \circledast I \right)_{(m',n')}$$

$$\boxed{\frac{\partial L}{\partial w(m',n')} = \left(\frac{\partial L}{\partial Y} \circledast I \right)_{(m',n')}}$$

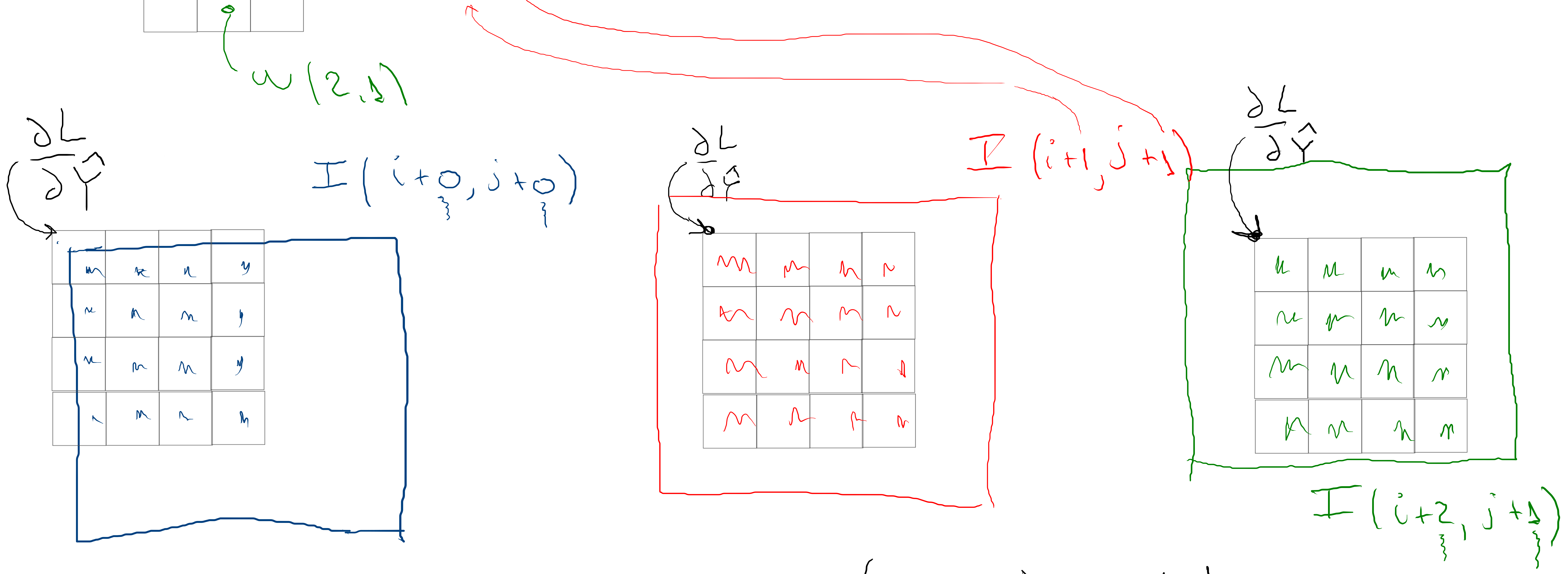
- Para cada elemento del kernel tendré 1 producto a resolver

viene dado por la capa siguiente

es la entrada

$$\frac{\partial L}{\partial w(m',n')} = \sum_{i=0}^{H-k_1} \sum_{j=0}^{W-k_2} I(m'+i, n'+j) \cdot \frac{\partial L}{\partial \hat{Y}(i,j)}$$

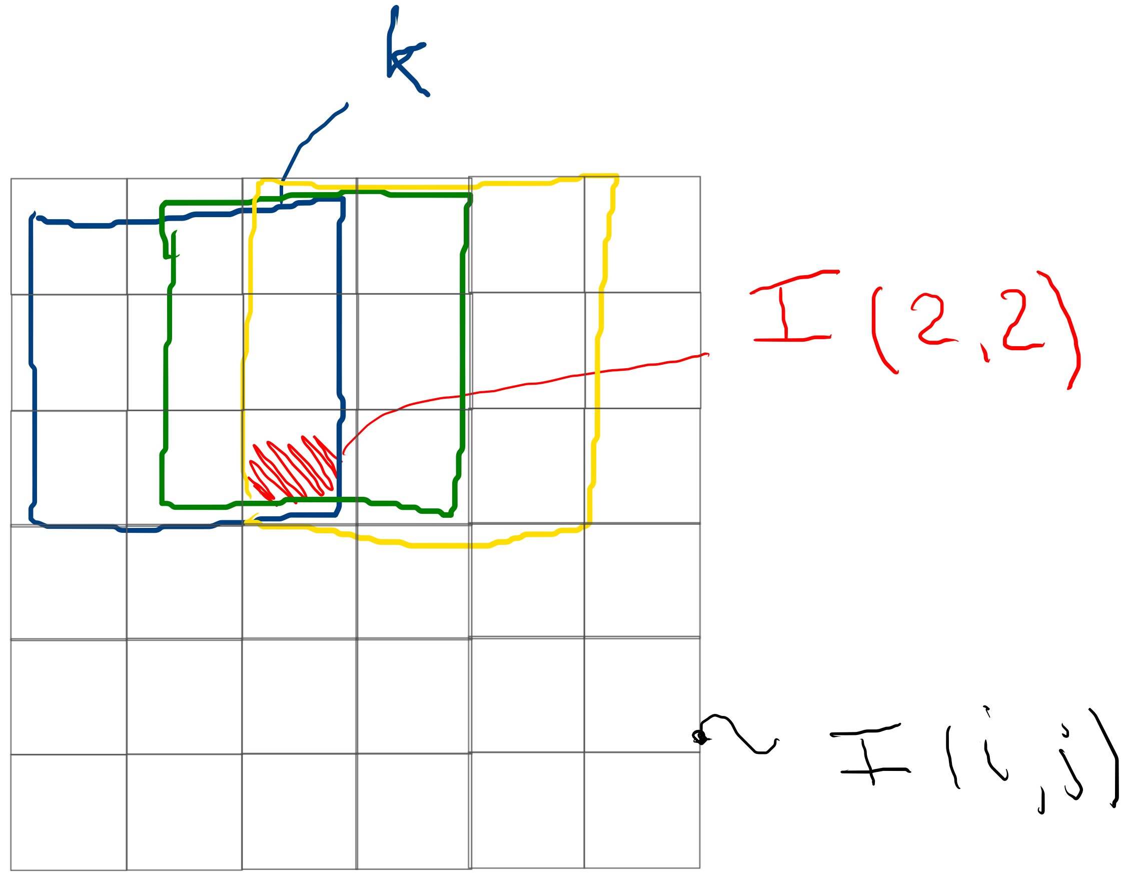
¿como se interpreta?

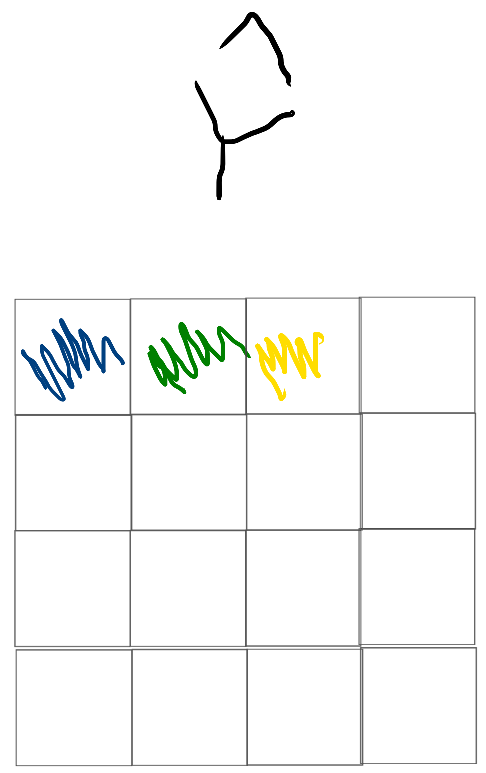


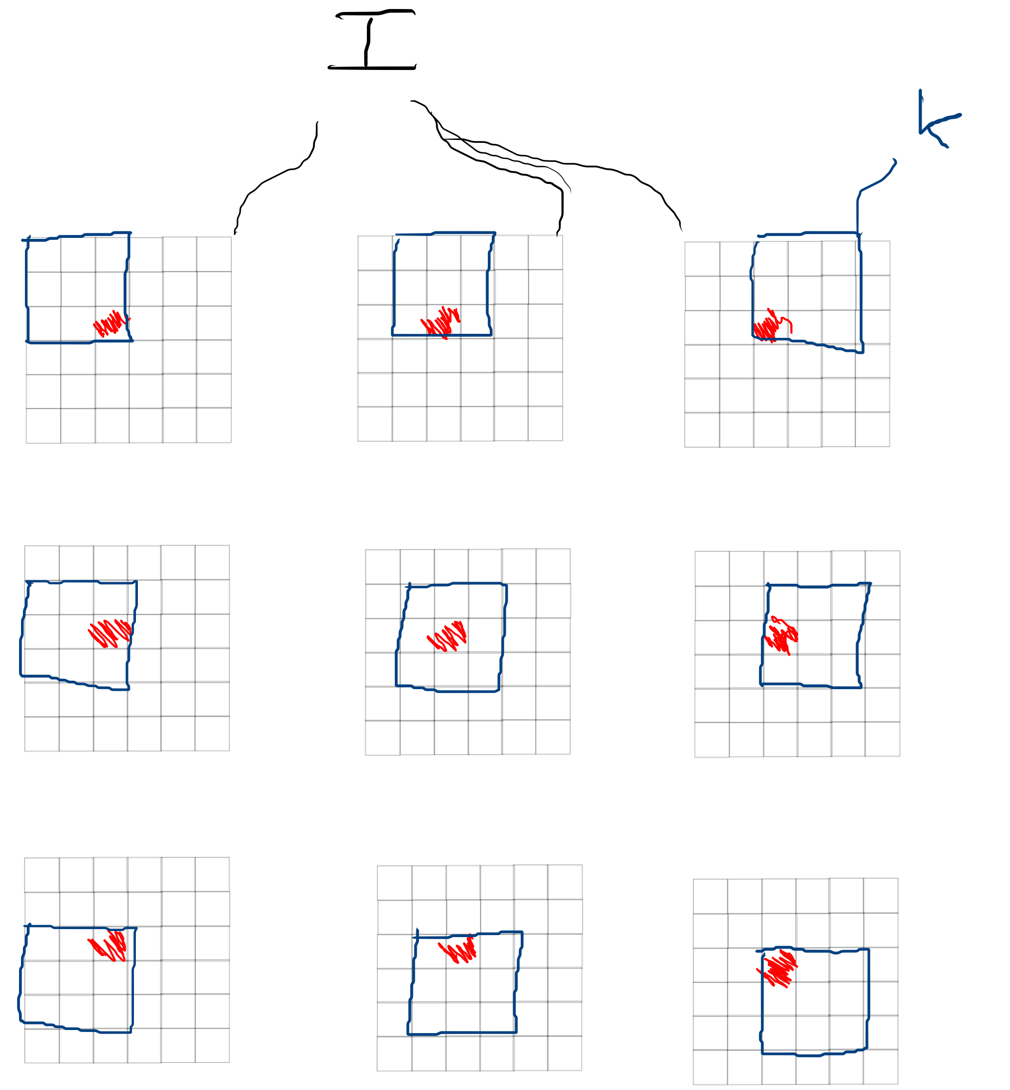
... son 9 $(k_1 \cdot k_2)$ en total...

Ahora seguimos con $\frac{\partial L}{\partial I}$ (para poder pasar el grad hacia la capa previa)

¿Como varia $\hat{Y}$ cuando se varia 1 pixel de I? $\Rightarrow$ dependerá del tamaño del kernel

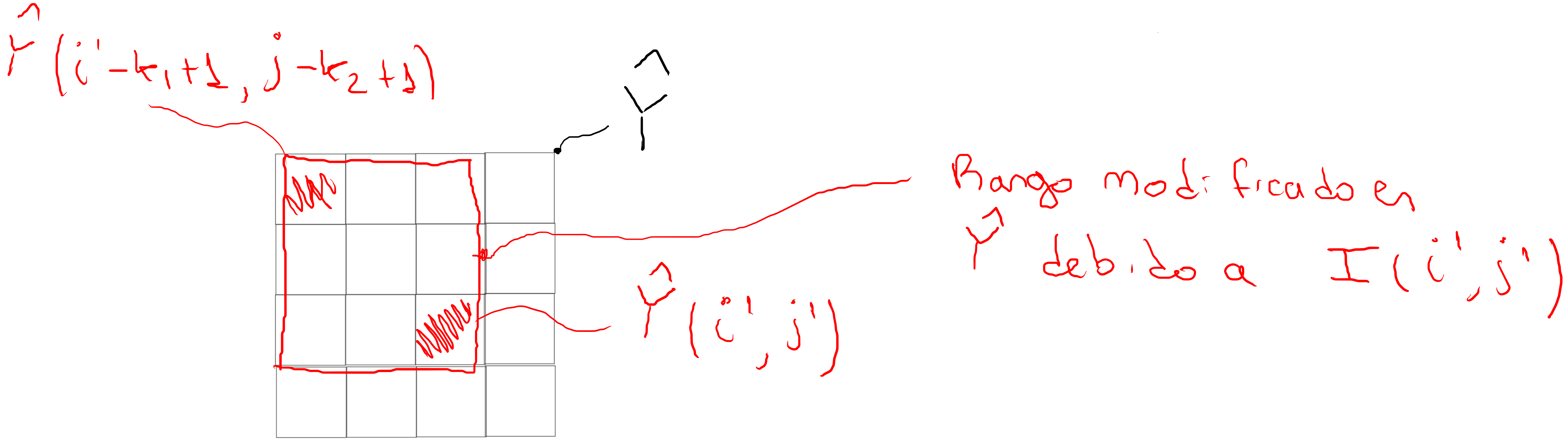






En este ejemplo hay 9 $(k_1 \cdot k_2)$ posiciones afectadas en $\hat{Y}$ por el pixel $I(2,2)$

Un pixel $I(i',j')$ va a modificar un area de $\hat{Y}$ en el rango $[\,i'-k_1+1 : i' \,;\, j'-k_2+1 : j\,]$



Entonces:

$$\frac{\partial L}{\partial I(i',j')} = \underbrace{\sum_{m=0}^{k_1-1}\sum_{n=0}^{k_2-1}}_{\text{en el rango del kernel}} \underbrace{\frac{\partial L}{\partial \hat{Y}(i'-m,j'-n)}}_{\text{de la capa posterior}} \cdot \overbrace{\frac{\partial \hat{Y}(i'-m,j'-n)}{\partial I(i',j')}}^{\text{a trabajar}} \tag{2}$$

el pixel (i',j') modifica a $\hat{Y}$ en el rango del kernel

$$\frac{\partial \hat{Y}(i'-m, j'-n)}{\partial I(i', j')} = \frac{\partial}{\partial I(i', j')} \Big[\underbrace{\sum_{p=0}^{k_1-1} \sum_{q=0}^{k_2-1}}_{\text{rango del kernel}} \overbrace{\underbrace{I\big((i'-m)+p, (j'-n)+q\big)}_{\text{hay 1 solo término donde}}}^{} \cdot w(p,q) + b \Big]$$

$\hat{Y}(i'-m, j'-n)$ por definición

hay 1 solo término donde

$p = m$

$q = n$

$I(i'-m+m, j'-n+n) = I(i', j')$!

entonces

$$\frac{\partial}{\partial I(i', j')} \cdot \Big[\ldots + I(i', j') \cdot w(m,n) + \ldots + b \Big]$$

$$\left.\frac{\partial \hat{Y}(i'-m, j'-n)}{\partial I(i', j')} = w(m,n) \right]$$

sustituimos en 2

$$\frac{\partial L}{\partial I(i',j')} = \underbrace{\sum_{m=0}^{k_1-1}\sum_{n=0}^{k_2-1}}_{\text{rango del kernel}} \frac{\partial L}{\partial \hat{Y}(i'-m, j'-n)} \cdot \overbrace{\omega(m,n)}^{\text{lo hallado}} \qquad (2')$$

$$\boxed{\frac{\partial L}{\partial I(i',j')} = \left(\omega * \frac{\partial L}{\partial \hat{Y}}\right)(i',j')}$$

para cada elemento de la entrada tengo que resolver el producto

viene dado por la capa siguiente

Para un (i', j') dado ...

$$\boxed{\frac{\partial L}{\partial I(i',j')} = \sum_{m=0}^{k_1-1}\sum_{n=0}^{k_2-1} \frac{\partial L}{\partial \hat{Y}}_{(i'-m,\, j'-n)} \cdot w(m,n)}$$

¿como se interpreta?

$\underbrace{i'-m \quad j'-n}$

con $m = k_1-1$, $n = k_2-1$

$\left(\frac{\partial L}{\partial \hat{Y}}\right)^*$

$\left.\begin{array}{l} i'=2 \\ j'=2 \end{array}\right\}$ con $m=0$, $n=0$

es la porcion de $\hat{Y}$ afectada por $I(i',j')$

... hay que resolver ese producto elemento a elemento.

$\left(\frac{\partial L}{\partial \hat{Y}}\right)^*$ es $\frac{\partial L}{\partial \hat{Y}}$ espejado en sus diagonales

$Q = \begin{array}{c|c} a & b \\ \hline c & d \end{array} \quad \xrightarrow{*} \quad Q^* = \begin{array}{c|c} d & c \\ \hline b & a \end{array}$