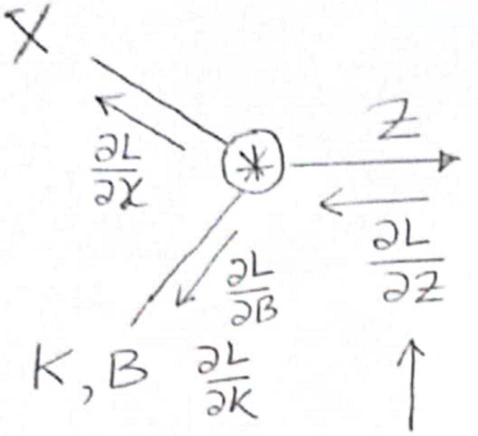


We have this from next layer

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{bmatrix} \circledast \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} + B = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$$

$$Z_{11} = X_{11}K_{11} + X_{12}K_{12} + X_{21}K_{21} + X_{22}K_{22} + B$$

$$Z_{12} = X_{12}K_{11} + X_{13}K_{12} + X_{22}K_{21} + X_{23}K_{22} + B$$

$$Z_{21} = X_{21}K_{11} + X_{22}K_{12} + X_{31}K_{21} + X_{32}K_{22} + B$$

$$Z_{22} = X_{22}K_{11} + X_{23}K_{12} + X_{32}K_{21} + X_{33}K_{22} + B$$

$\frac{\partial L}{\partial K} = \frac{\partial L}{\partial Z} \cdot \frac{\partial Z}{\partial K}$

$$\begin{bmatrix} \frac{\partial L}{\partial K_{11}} & \frac{\partial L}{\partial K_{12}} \\ \frac{\partial L}{\partial K_{21}} & \frac{\partial L}{\partial K_{22}} \end{bmatrix}$$

$$\frac{\partial L}{\partial K_{mn}} = \sum_{ij} \frac{\partial L}{\partial Z_{ij}} \cdot \frac{\partial Z_{ij}}{\partial K_{mn}}$$

$$\frac{\partial L}{\partial K_{11}} = \frac{\partial L}{\partial Z_{11}} \cdot \frac{\partial Z_{11}}{\partial K_{11}} + \frac{\partial L}{\partial Z_{12}} \cdot \frac{\partial Z_{12}}{\partial K_{11}} + \frac{\partial L}{\partial Z_{21}} \cdot \frac{\partial Z_{21}}{\partial K_{11}} + \frac{\partial L}{\partial Z_{22}} \cdot \frac{\partial Z_{22}}{\partial K_{11}}$$

$$= \frac{\partial L}{\partial Z_{11}} \cdot X_{11} + \frac{\partial L}{\partial Z_{12}} \cdot X_{12} + \frac{\partial L}{\partial Z_{21}} \cdot X_{21} + \frac{\partial L}{\partial Z_{22}} \cdot X_{22}$$

Similarly

$$\frac{\partial L}{\partial K_{12}} = \frac{\partial L}{\partial Z_{11}} \cdot X_{12} + \frac{\partial L}{\partial Z_{12}} \cdot X_{13} + \frac{\partial L}{\partial Z_{21}} \cdot X_{22} + \frac{\partial L}{\partial Z_{22}} \cdot X_{23}$$

COURTESY: Coding Lane Youtube Channel

$$\frac{\partial L}{\partial k_{21}} = \frac{\partial L}{\partial z_{11}} \cdot x_{21} + \frac{\partial L}{\partial z_{12}} \cdot x_{22} + \frac{\partial L}{\partial z_{21}} \cdot x_{31} + \frac{\partial L}{\partial z_{22}} \cdot x_{32}$$

$$\frac{\partial L}{\partial k_{22}} = \frac{\partial L}{\partial z_{11}} x_{21} + \frac{\partial L}{\partial z_{12}} x_{23} + \frac{\partial L}{\partial z_{21}} x_{32} + \frac{\partial L}{\partial z_{22}} x_{33}$$

\## $$\frac{\partial L}{\partial k} = \begin{bmatrix} \frac{\partial L}{\partial k_{11}} & \frac{\partial L}{\partial k_{12}} \\ \frac{\partial L}{\partial k_{21}} & \frac{\partial L}{\partial k_{22}} \end{bmatrix} \rightarrow \operatorname{conv}\left(X, \frac{\partial L}{\partial Z}\right)$$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \circledast \begin{bmatrix} \frac{\partial L}{\partial z_{11}} & \frac{\partial L}{\partial z_{12}} \\ \frac{\partial L}{\partial z_{21}} & \frac{\partial L}{\partial z_{22}} \end{bmatrix}$$

\# $$\frac{\partial L}{\partial B} = \sum_{ij} \frac{\partial L}{\partial z_{ij}} \cdot \frac{\partial z_{ij}}{\partial B} = \sum_{ij} \frac{\partial L}{\partial z_{ij}} = \operatorname{sum}\left(\frac{\partial L}{\partial Z}\right)$$

\# $$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Z} \cdot \frac{\partial Z}{\partial x}$$

$$\frac{\partial L}{\partial X_{mn}} = \sum_{ij} \frac{\partial L}{\partial z_{ij}} \cdot \frac{\partial z_{ij}}{\partial X_{mn}}$$

$$\frac{\partial L}{\partial x_{11}} = \frac{\partial L}{\partial z_{11}} \cdot k_{11}$$

$$\frac{\partial L}{\partial x_{12}} = \frac{\partial L}{\partial z_{11}} \cdot k_{12} + \frac{\partial L}{\partial z_{12}} \cdot k_{11}$$

$$\frac{\partial L}{\partial x_{13}} = \frac{\partial L}{\partial z_{12}} \cdot k_{12}$$

$$\frac{\partial L}{\partial x_{21}} = \frac{\partial L}{\partial z_{11}} \cdot k_{21} + \frac{\partial L}{\partial z_{21}} k_{11}$$

$$\frac{\partial L}{\partial x_{22}} = \frac{\partial L}{\partial z_{11}} \cdot k_{22} + \frac{\partial L}{\partial z_{12}} \cdot k_{21} + \frac{\partial L}{\partial z_{21}} k_{12} + \frac{\partial L}{\partial z_{22}} k_{11}$$

$$\frac{\partial L}{\partial x_{23}} = \frac{\partial L}{\partial z_{12}} \cdot k_{22} + \frac{\partial L}{\partial z_{22}} k_{12}$$

$$\frac{\partial L}{\partial x_{31}} = \frac{\partial L}{\partial z_{21}} \cdot \cancel{\partial} K_{21}$$

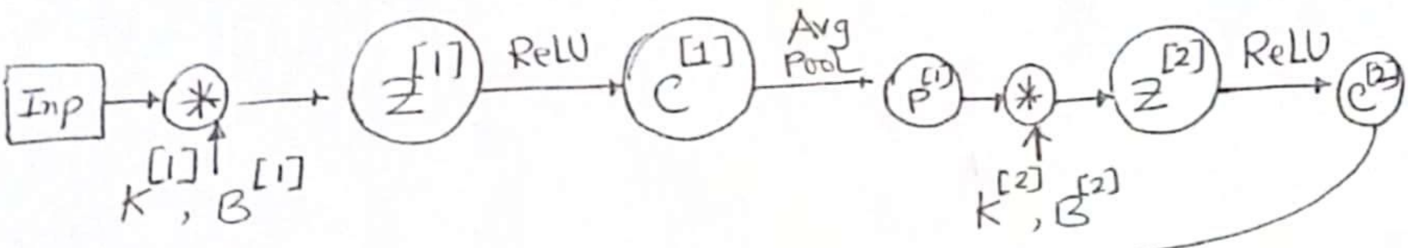
$$\frac{\partial L}{\partial x_{32}} = \frac{\partial L}{\partial z_{21}} \cdot K_{22} + \frac{\partial L}{\partial z_{22}} \cdot K_{21}$$

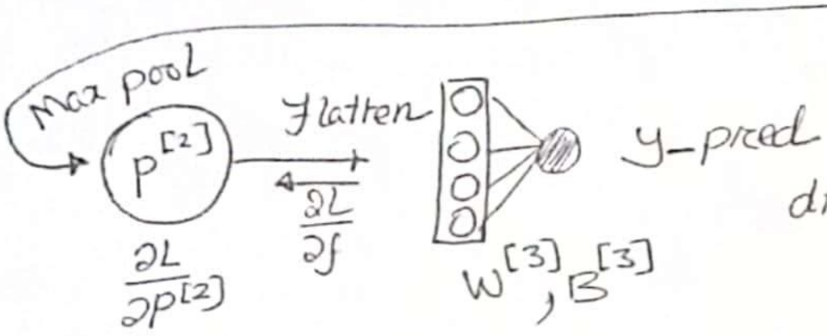
$$\frac{\partial L}{\partial x_{33}} = \frac{\partial L}{\partial z_{22}} \cdot K_{22}$$

This is 180° rotation of ~~filter~~ filter K

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial L}{\partial z_{11}} & \frac{\partial L}{\partial z_{12}} & 0 \\ 0 & \frac{\partial L}{\partial z_{21}} & \frac{\partial L}{\partial z_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \circledast \begin{bmatrix} K_{22} & K_{21} \\ K_{12} & K_{11} \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial x_{11}} & \frac{\partial L}{\partial x_{12}} & \frac{\partial L}{\partial x_{13}} \\ \frac{\partial L}{\partial x_{21}} & \frac{\partial L}{\partial x_{22}} & \frac{\partial L}{\partial x_{23}} \\ \frac{\partial L}{\partial x_{31}} & \frac{\partial L}{\partial x_{32}} & \frac{\partial L}{\partial x_{33}} \end{bmatrix}$$

$\frac{\partial L}{\partial x} = \text{conv}\left(\text{padded}\left(\frac{\partial L}{\partial z}\right), 180^\circ \text{ rotated fiter } K \right)$





$f = \text{flatten}(P^{[2]})$

$dP^{[2]} =: \frac{\partial L}{\partial P^{(2)}} = df.\text{reshape}(P^{[2]}.\text{shape})$

Gradient of Max Pooling

$$C^{[2]} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \rightarrow P^{[2]} = [4].$$

Let's assume $\frac{\partial L}{\partial P^{[2]}} = [2]$

$$\therefore \frac{\partial L}{\partial C^{[2]}} = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\frac{\partial L}{\partial C^{[2]}_{mn}} = \begin{cases} \frac{\partial L}{\partial P^{[2]}_{xy}}, & \text{if } C_{mn} \text{ is the max element} \\ 0, & \text{otherwise} \end{cases}$$

$$C^{[2]} = \text{ReLU}(Z^{[2]})$$

$$\frac{\partial L}{\partial Z^{[2]}} = \frac{\partial L}{\partial C^{[2]}} \frac{\partial C^{[2]}}{\partial Z^{[2]}}$$

$$\frac{\partial C^{[2]}}{\partial Z^{[2]}_{mn}} = \begin{cases} 1, & \text{if } Z_{mn} > 0 \\ 0, & \text{otherwise} \end{cases}$$

$\frac{\partial L}{\partial P^{[1]}} \rightarrow \frac{\partial L}{\partial X}$ of earlier calculation

$$C^{[1]} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \therefore P^{(1)} = [2.5]$$

Let's assume

$$\frac{\partial L}{\partial P^{[1]}} = [2]$$

$$\frac{\partial L}{\partial C^{[1]}} = ? \rightarrow \begin{bmatrix} \frac{1}{4}\cdot 2 & \frac{1}{4}\cdot 2 \\ \frac{1}{4}\cdot 2 & \frac{1}{4}\cdot 2 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

Courtesy: Code Lane youtube channel.

$$\frac{\partial L}{\partial C^{[1]}_{mn}} = \frac{1}{4} * \frac{\partial L}{\partial P^{[2]}_{xy}} \quad \text{For} \quad \begin{array}{l} x = \text{floor}(m/2) \\ y = \text{floor}(n/2) \end{array}$$

Gradient of cross entropy with softmax.

For each class k, softmax prob. $P_K = \frac{e^{f_K}}{\sum_j e^{f_j}}$

f is the mapping from input x to output layer

cross entropy loss L_i (for i^{th} datapoint) $= -\log(P_{y_i})$

$$\frac{\partial L_i}{\partial f_K} = -\frac{\partial}{\partial f_K} \log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) = -\frac{\sum_j e^{f_j}}{e^{f_{y_i}}} \frac{\partial}{\partial f_K} \frac{e^{f_{y_i}}}{\sum_j e^{f_j}}$$

$\boxed{\text{for } K = y_i} \rightarrow$

$$\frac{\partial L_i}{\partial f_K} = -\frac{\sum_j e^{f_j}}{e^{f_{y_i}}} \cdot \frac{1}{(\sum_j e^{f_j})^2} \cdot \left(e^{f_K} \cdot \sum_j e^{f_j} - e^{f_K} \cdot e^{f_K}\right)$$

$$= -\frac{1}{e^{f_K} \cdot \sum_j e^{f_j}} \left[e^{f_K} \cdot \sum_j e^{f_j} - e^{f_K} \cdot e^{f_K}\right]$$

$$= \boxed{P_k - 1}$$

For $\boxed{K \neq y_i}$

$$\frac{\partial L_i}{\partial f_K} = -\frac{\sum_j e^{f_j}}{e^{f_{y_i}}} \cdot \frac{1}{(\sum_j e^{f_j})^2} \left[\sum_j e^{f_j} \cdot 0 - e^{f_{y_i}} e^{f_K}\right]$$

$$= -\frac{1}{e^{f_{y_i}} \sum_j e^{f_j}} \left(-e^{f_K} \cdot e^{f_{y_i}}\right)$$

$$= \frac{e^{f_K}}{\sum_j e^{f_j}} = \boxed{P_K}$$

End