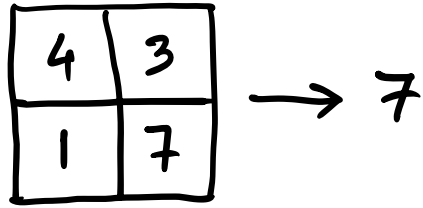


Maxpool Layer



- Provides some level of **shift and rotation** - invariance.
- Often used to reduce the spatial dimension of the data.

Forward pass: $y = \max(4, 3, 1, 7)$

But how would you backpropagate the gradient?

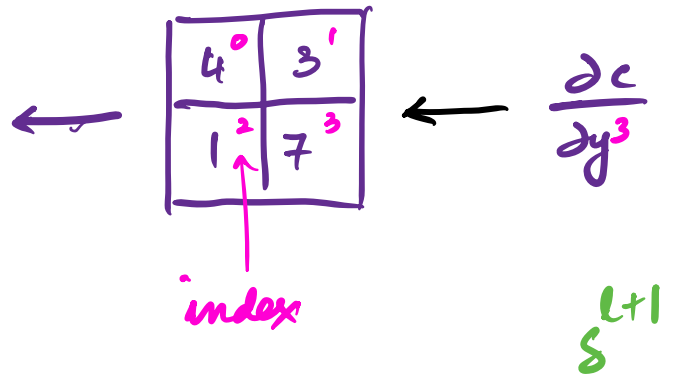
- ① There is no gradient w.r.t. non-maxima values.
- ② Max operation is locally linear with slope 1 w.r.t. the input that receives max.

Gradient is passed back only to the neuron that achieved the max. All other neurons will receive a zero.

$$\frac{\partial c}{\partial x} = \frac{\partial c}{\partial y} \cdot \frac{\partial x}{\partial y}$$

In the case of a maxpool $\frac{\partial x}{\partial y}$ is either a 1 or a 0.

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{\partial c}{\partial y^3} \end{bmatrix} \delta^L$$



This means that the maxpool layer should remember

the index where the max value was found.

- In practice maxpool performs better than minpool or averagepool operations.