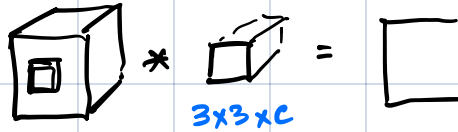


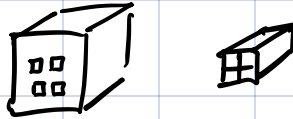
1. Convolution

1. Standard convolution



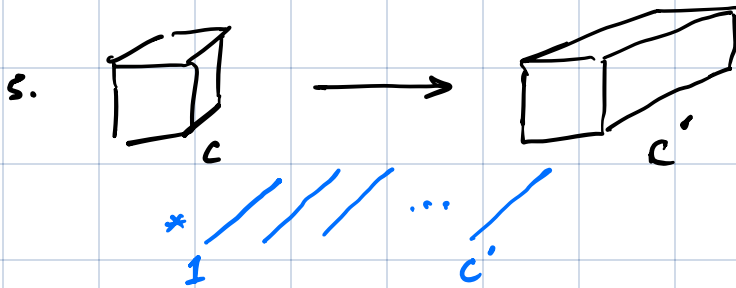
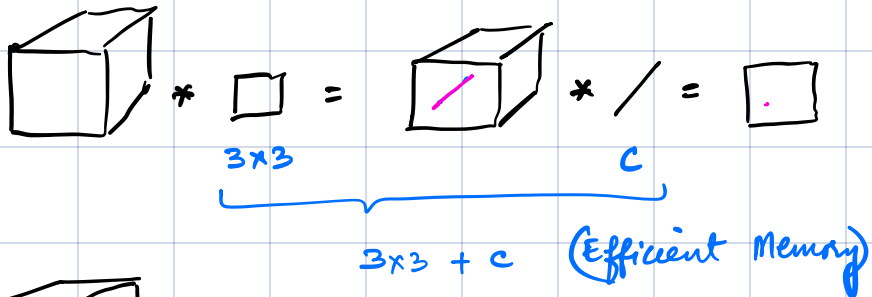
2. Padding, stride

3. Dilated

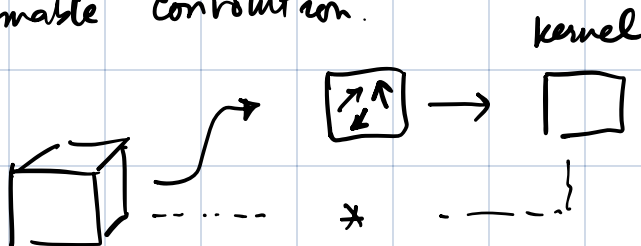


a. Receptive field

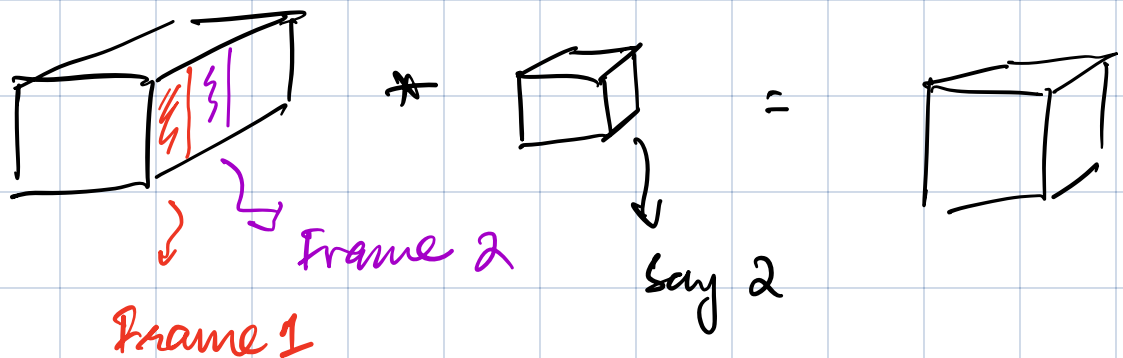
4. Point-wise / separable convolution



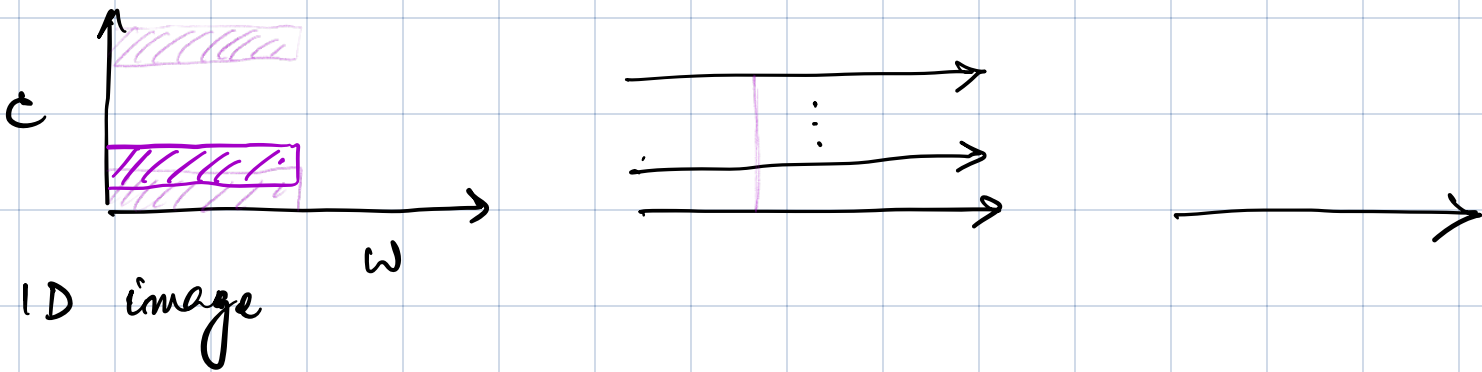
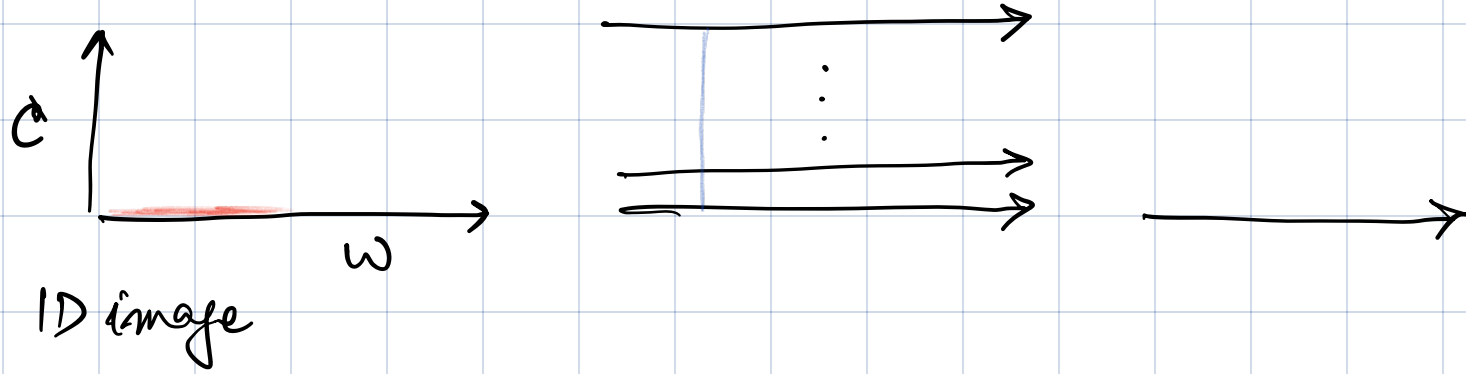
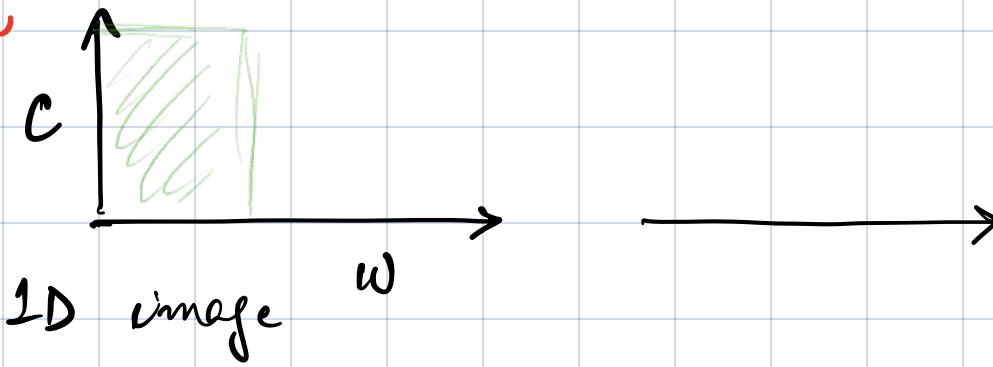
6. Deformable convolution.



7. 3D convolution



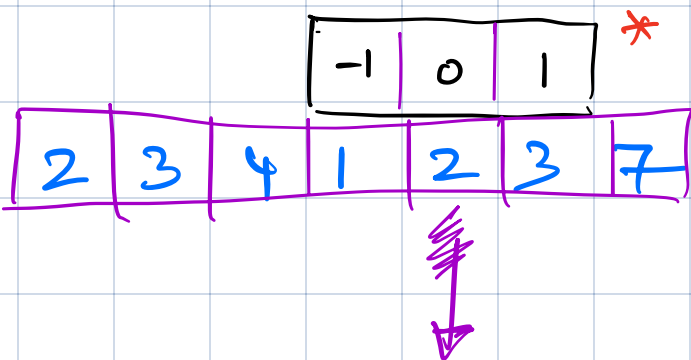
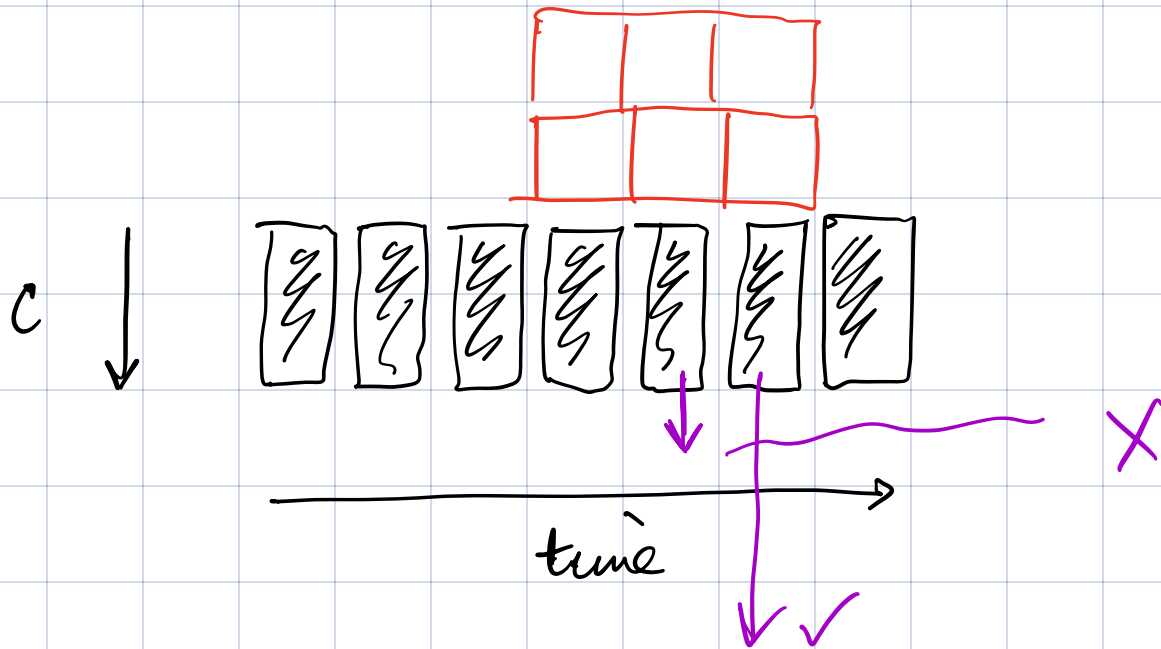
Axis



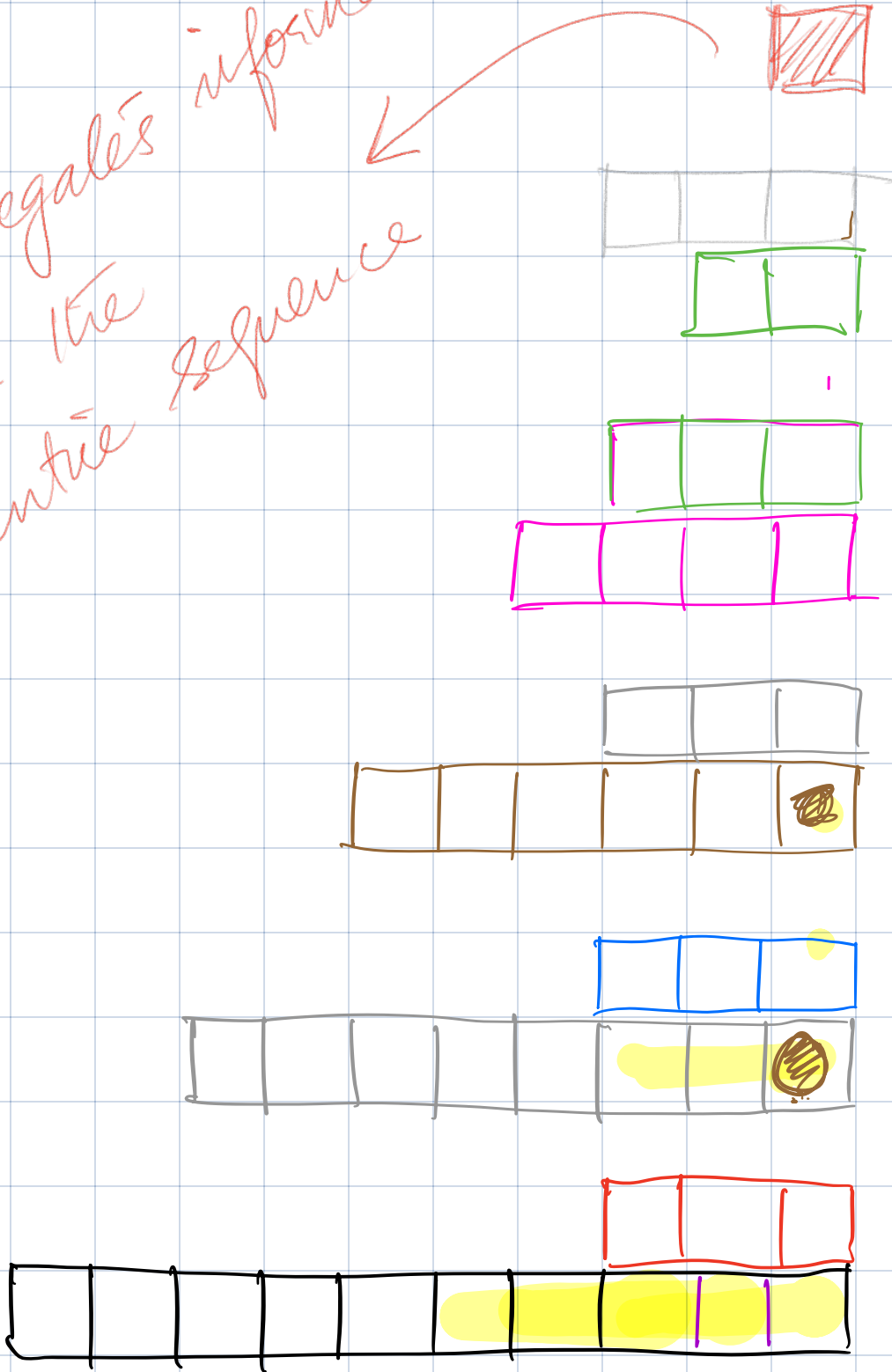
8. Temporal convolutional Network

$$\vec{x}_{t-k+1} \dots \vec{x}_{t-2} \vec{x}_{t-1} \vec{x}_t \in \mathbb{R}^d$$

Example: say $d=2$, $k=7$

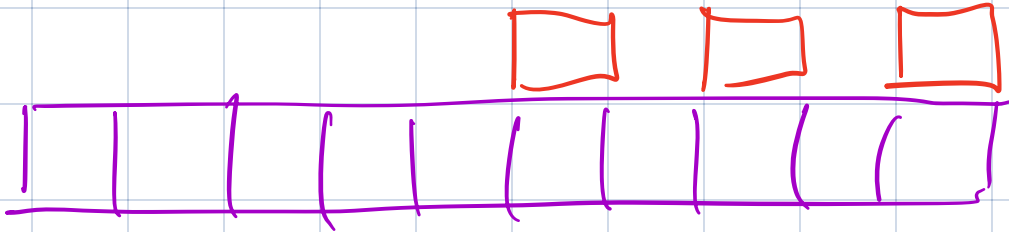


Aggregate's information
from the entire sequence



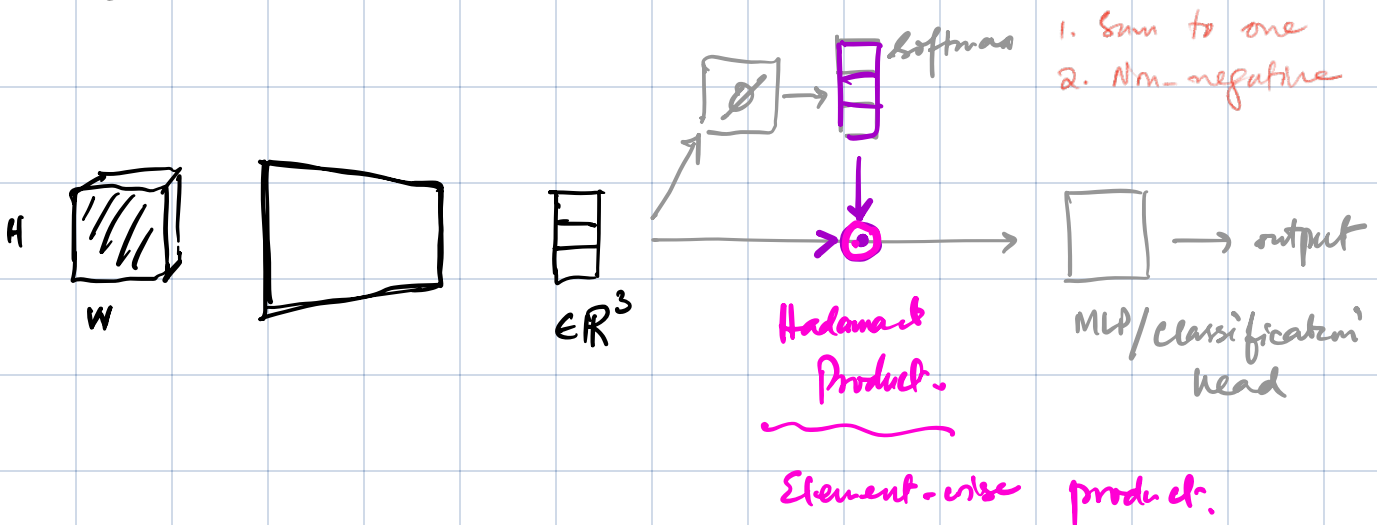
$\in \mathbb{R}^2$

increase the receptive field - dilation:



Attention:

"All you need is attention" → transformers.



input

$$\begin{bmatrix} 3 \\ 7 \\ -1 \end{bmatrix} \vec{a}$$

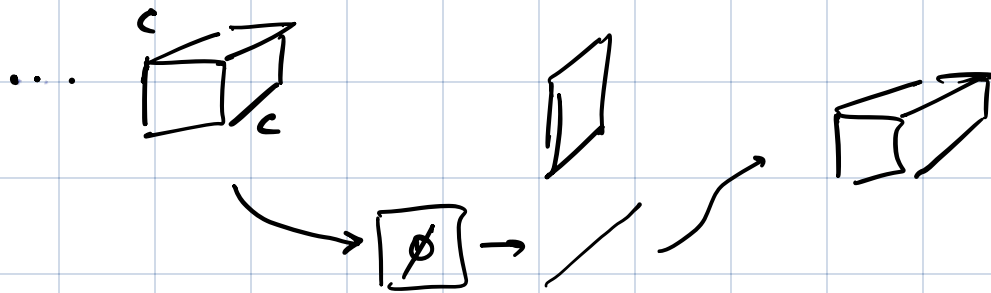
output

$$\begin{bmatrix} 9 \\ 8 \\ -3 \end{bmatrix} \vec{b} \xrightarrow{\text{softmax}} \begin{bmatrix} 0.7 \\ 0.2 \\ 0.1 \end{bmatrix}$$

$$\frac{e^{-9}}{e^{-9-8+3}}$$

$$\vec{b} = \text{relu} \left(\begin{bmatrix} 3 \times 3 \end{bmatrix} \vec{a} \right) \in \mathbb{R}^3$$

$$\begin{bmatrix} (3) & (0.7) \\ (7) & (0.2) \\ (-1) & (0.1) \end{bmatrix} = \begin{bmatrix} 2.1 \\ 1.4 \\ 0.01 \end{bmatrix}$$



1	3
1	4

1	2
3	4

$10.9/0.11$
attention
channel-wise

0.1	0.3
0.1	0.4

0.9	1.8
3.6	4.5

1	3
1	4

1	2
3	4

0.1	0.2
0.3	0.4

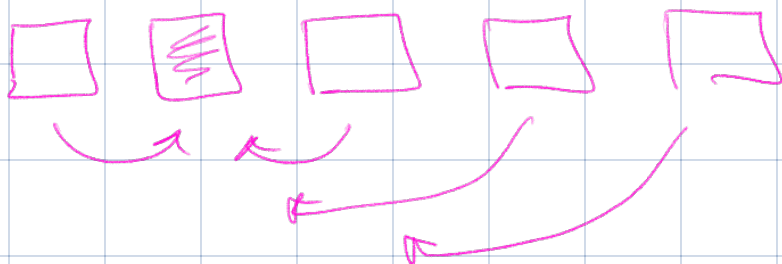
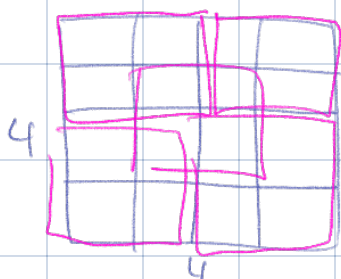
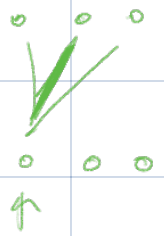
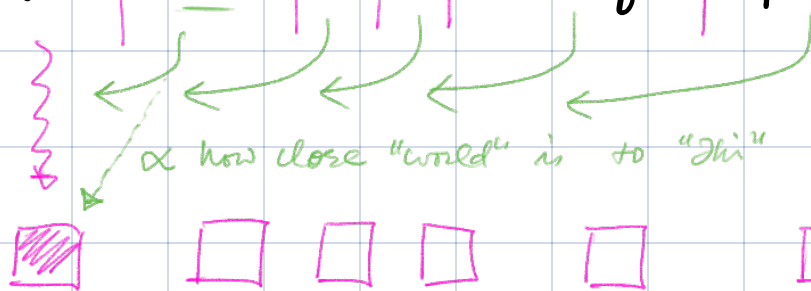
attention
spatial

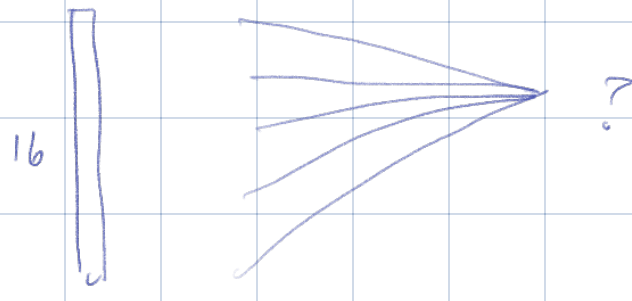
0.1	0.6
0.3	1.6

0.1	0.4
0.9	1.6

Self-attention:

this | word | is | a | beautiful | place!





positional
encoding.