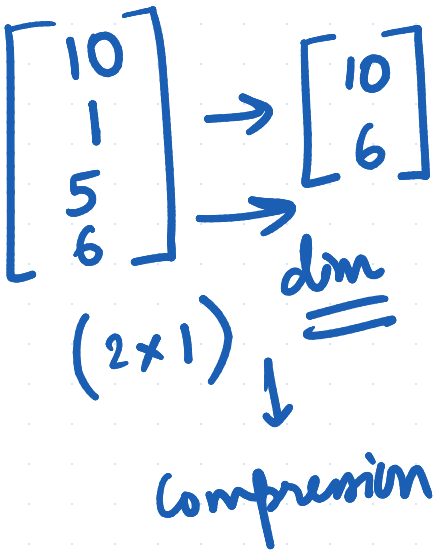
10/4/24 **Attention**

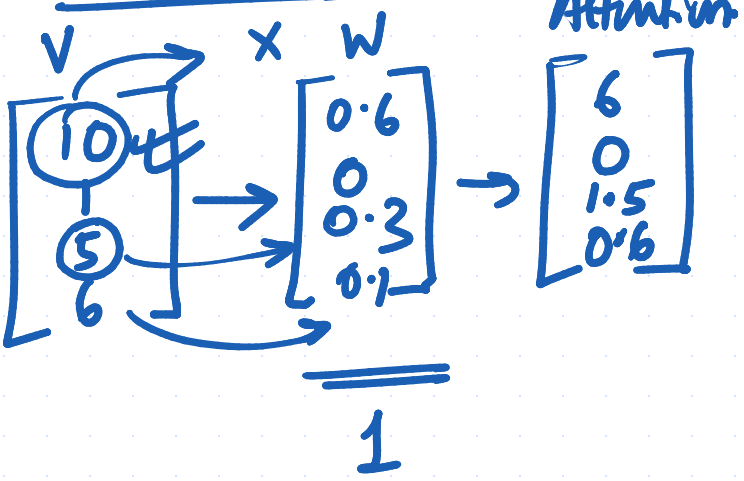
$$\begin{bmatrix} 15 \\ 2 \\ 10 \\ 3 \end{bmatrix} \Rightarrow v \Rightarrow \begin{bmatrix} x_1' \\ x_2' \\ x_3' \\ x_4' \end{bmatrix} \Rightarrow \frac{v}{\sqrt{(x_1')^2+(x_2')^2+(x_3')^2+(x_4')^2}}$$

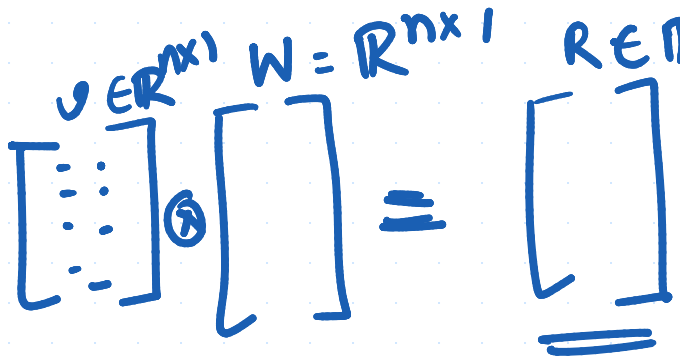
DL.

Attention is focusing on relevant information/ features and not focusing on others.

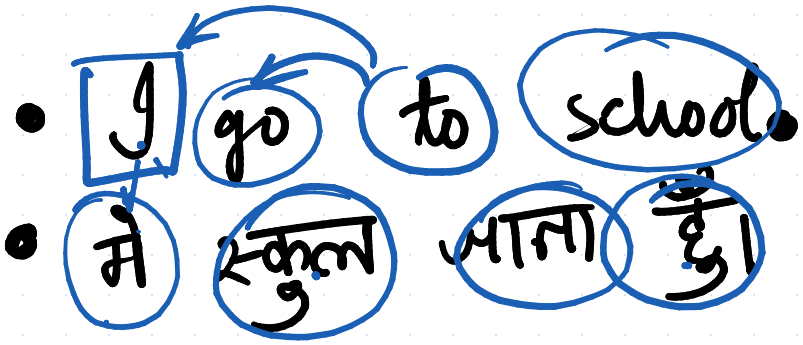
(Max pool etc.) **VS** **Attention**

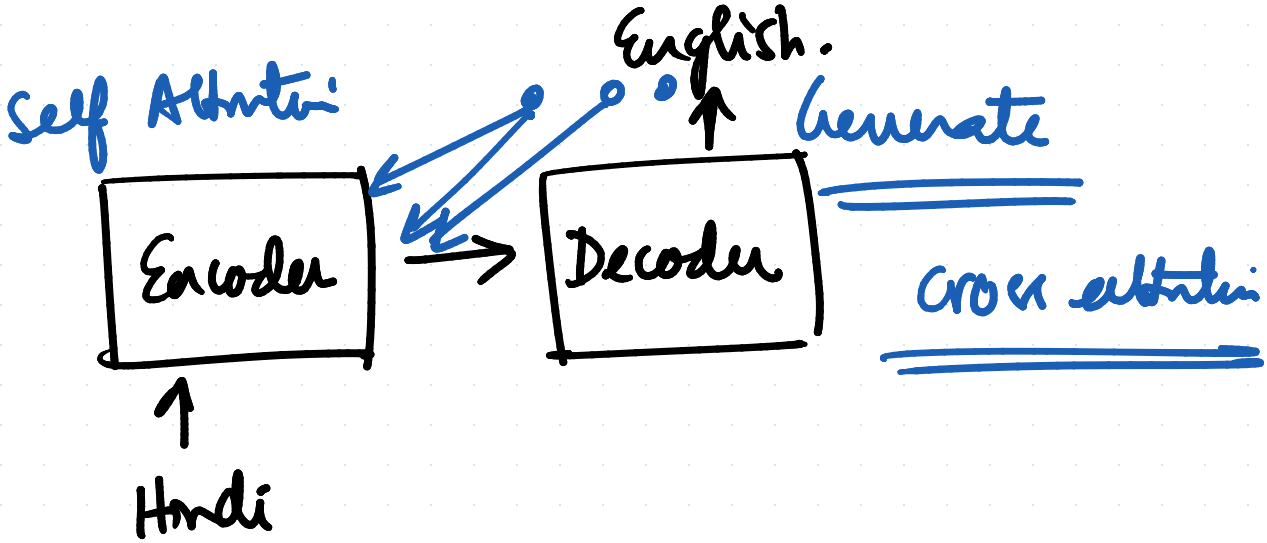




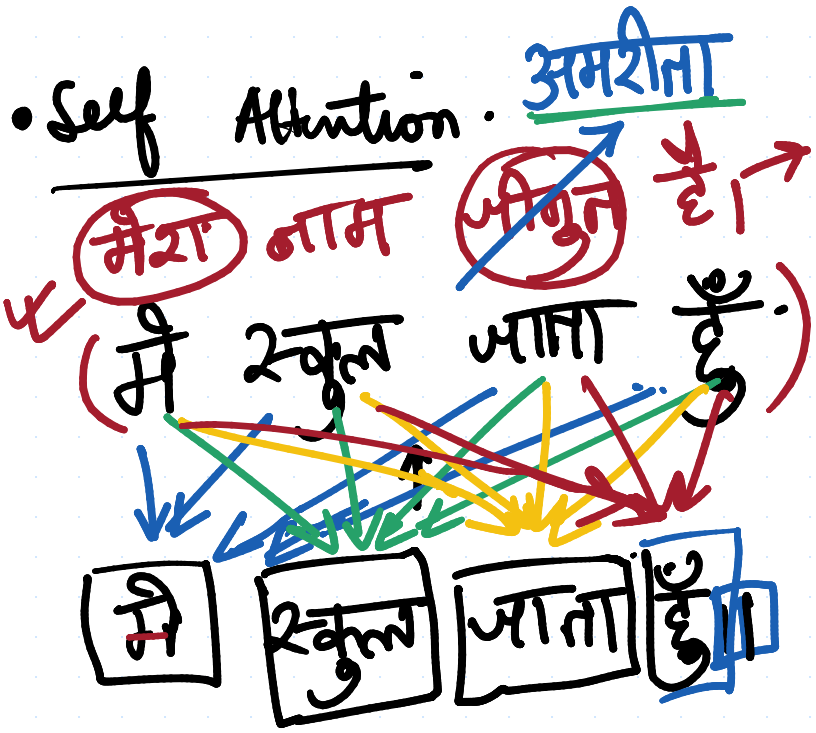


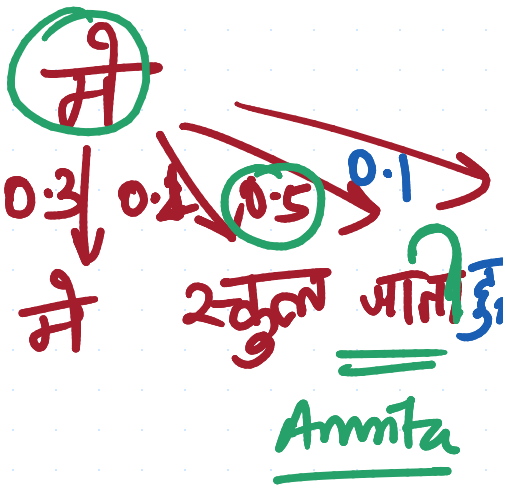
⇒ Find the weight matrix which will be used to multiply the initial feature volume so that the less important features gets less weight & the more imp feature gets more weight.



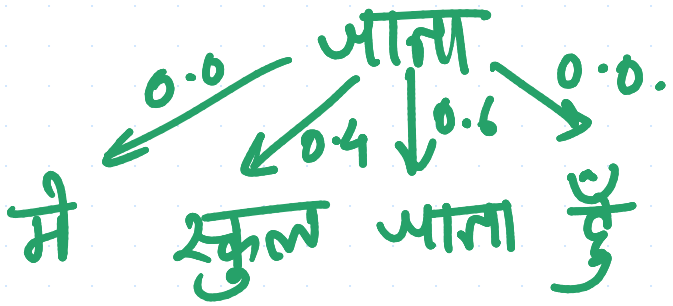


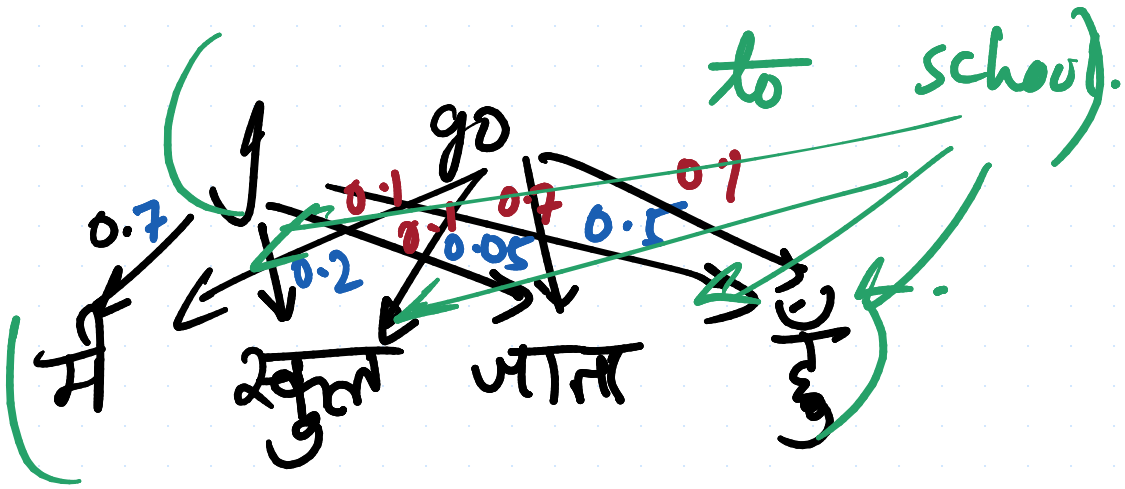
• Self Attention. अमरीता





Self Attention





Self Attention using matrix

X (input) $\times$ W^Q (learn) $= Q$

X (input) $\times$ W^K (learn) $= K$

X (input) $\times$ W^V (learn) $= V$

→ Weight matrix.

$$\text{Softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \times V = Z \quad \text{(learn)}$$