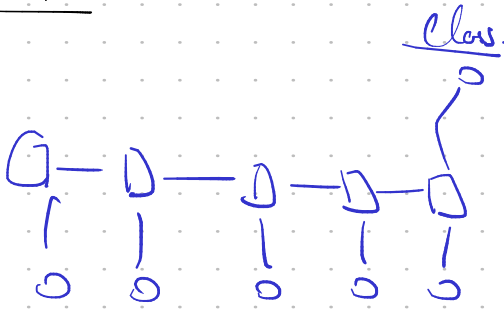
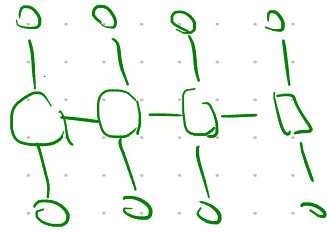


RNNs

1)

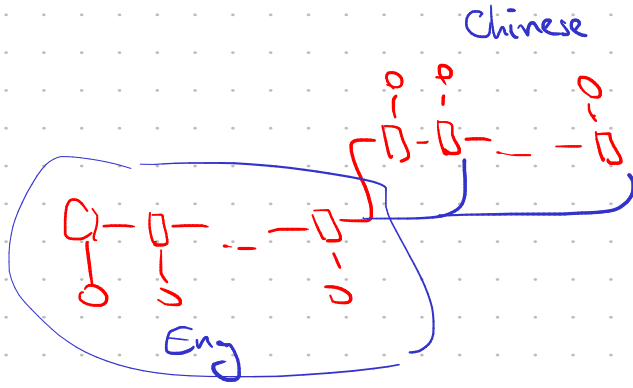


2)

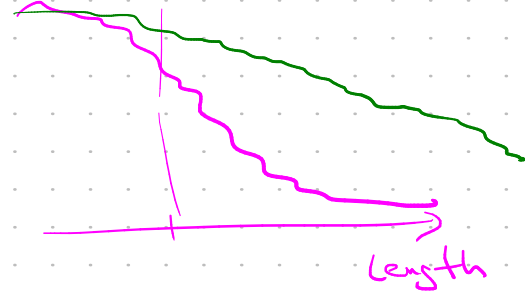


NER

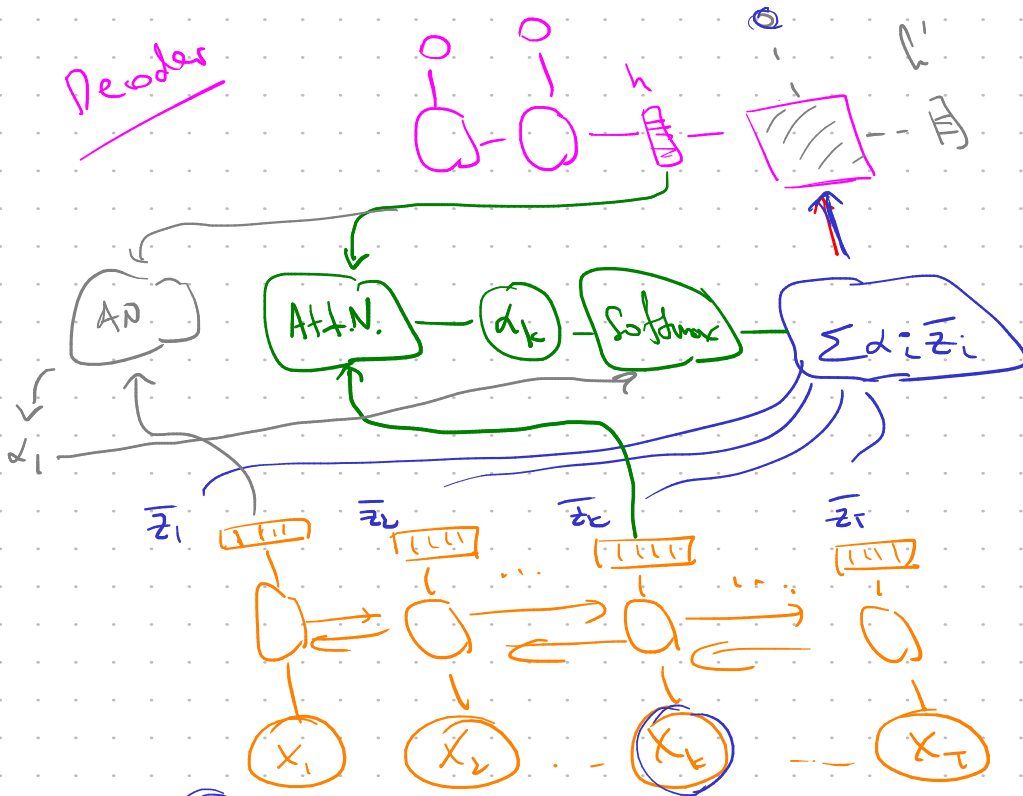
3) Seq2seq



Score



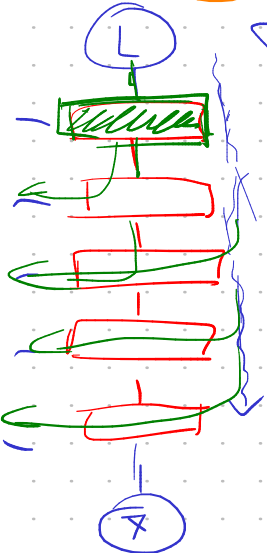
Decoder



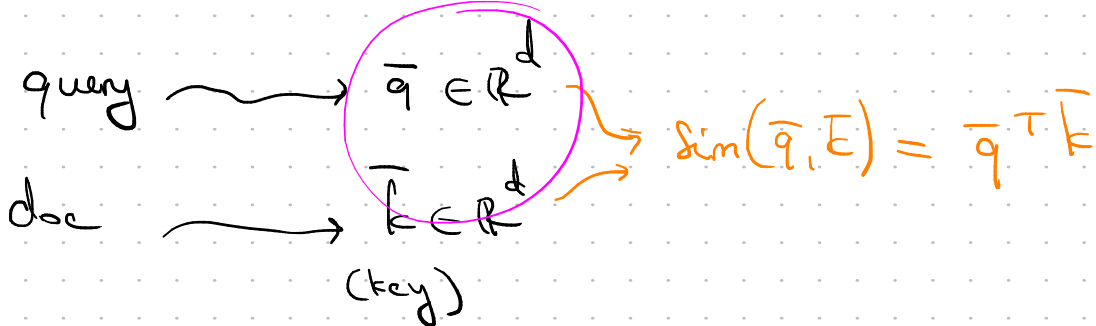
$$\Delta_{\bar{y}} L \rightarrow \min$$

$$\bar{y} = F(x) + \gamma$$

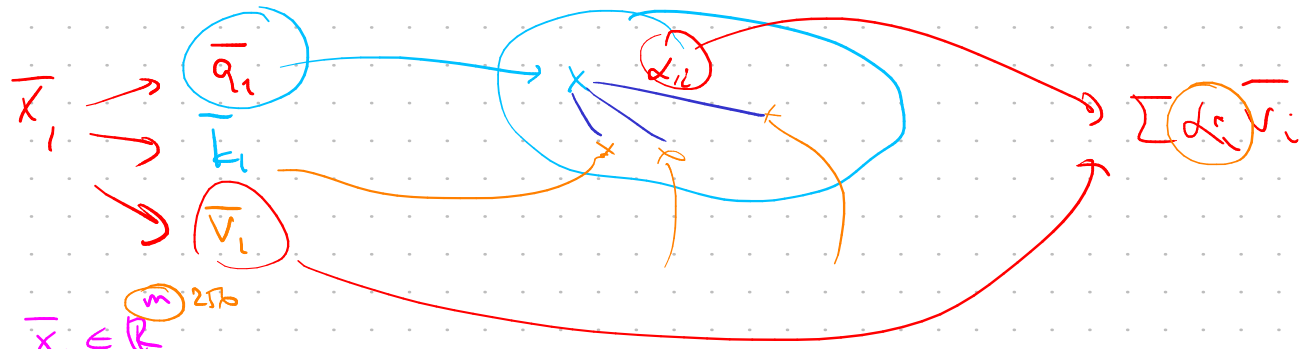
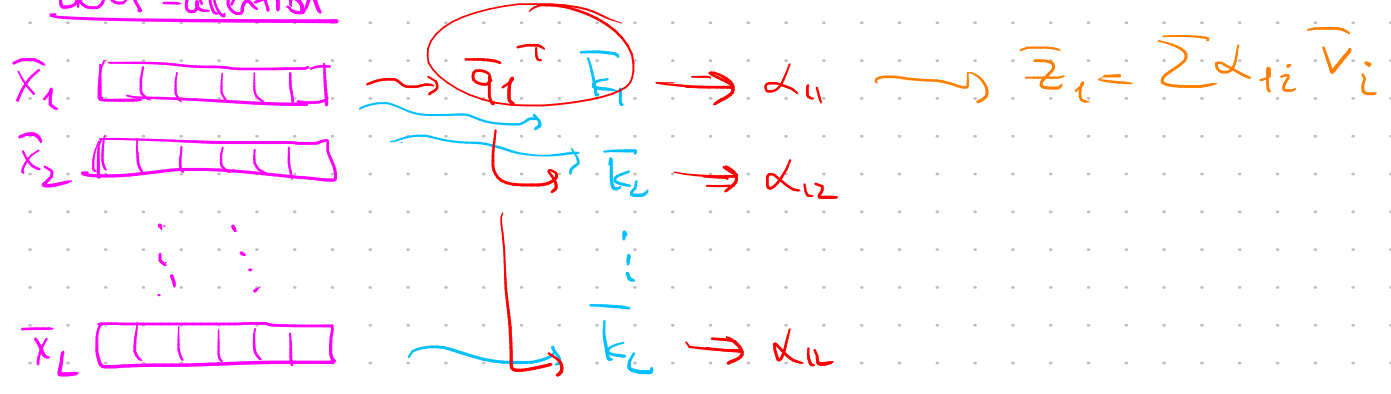
$$\Delta_x \bar{y} = \nabla_x F + I$$



12
12
12
12



Self-attention



$\bar{x}_1 \in \mathbb{R}^m$

$\bar{q}_1 = W^Q \cdot \bar{x}_1 \in \mathbb{R}^d$ $W^Q : d \times m$

$\bar{k}_1 = W^K \cdot \bar{x}_1$ $W^K : d \times m$

$\bar{v}_1 = W^V \cdot \bar{x}_1$ $W^V : v \times m$

$\frac{1}{\sqrt{d}} \bar{q}_1^T \bar{k}_1$ $\frac{1}{\sqrt{d}} \bar{q}_1^T \bar{k}_2$... $\frac{1}{\sqrt{d}} \bar{q}_1^T \bar{k}_L$

Softmax

α_{11} α_{12} α_{1L}

$\alpha_{ij} = \frac{e^{\frac{1}{\sqrt{d}} \bar{q}_i^T \bar{k}_j}}{\sum_k e^{\frac{1}{\sqrt{d}} \bar{q}_i^T \bar{k}_k}}$

$$\bar{z}_1 = \sum \alpha_{1e} \bar{v}_e = \sum_e \text{softmax}\left(\frac{1}{\sigma} \bar{q}_1^T k_e\right) \cdot \bar{v}_e$$

$$\bar{z}_2 = \sum \alpha_{2e} \bar{v}_e$$

' - -

$$\bar{z}_L = \sum \alpha_{Le} \bar{v}_e$$

$$Z = \text{softmax}\left(\frac{1}{\sigma} Q K^T\right) \cdot V$$