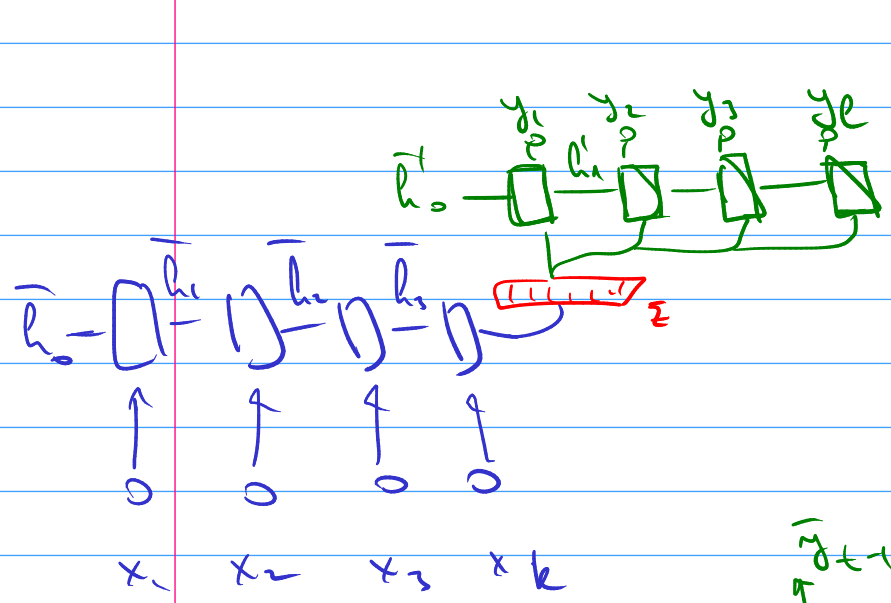
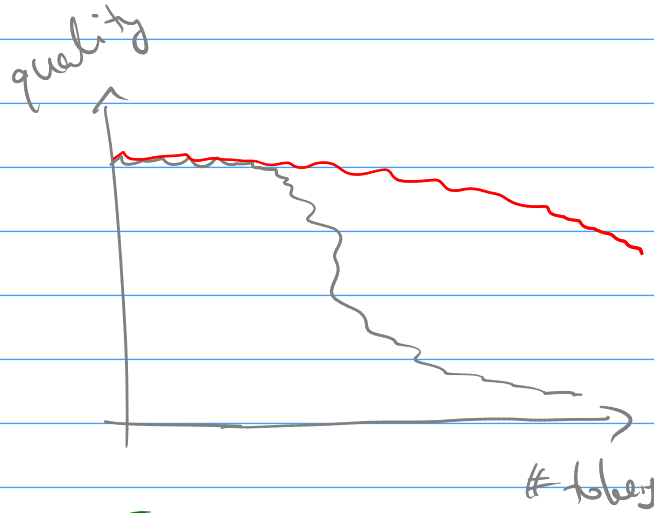
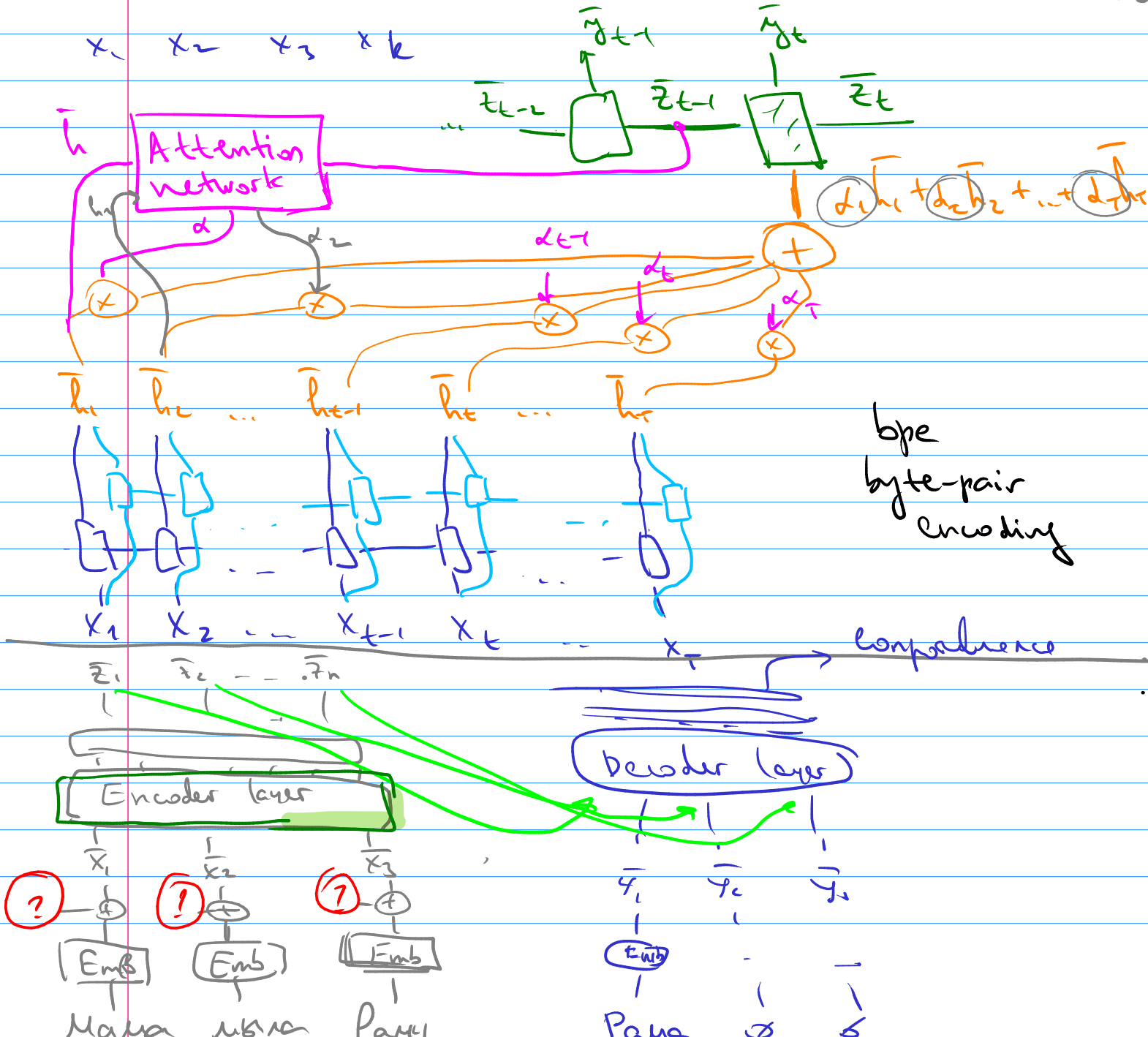
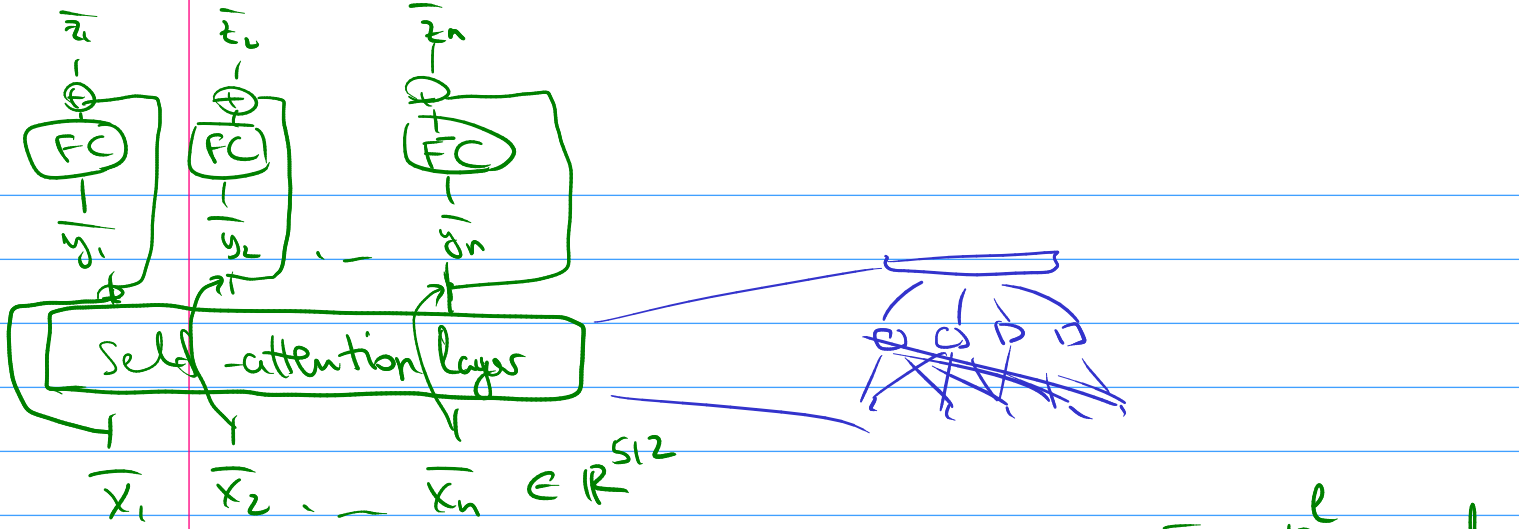
Мыла находится от раме на 60 см









$\bar{x}_1 \quad \bar{x}_2 \quad \dots \quad \bar{x}_n \in \mathbb{R}^{512}$

$\mathbb{R}^d$

$\bar{q}_i \in \mathbb{R}^m$ — query

$\bar{k}_i \in \mathbb{R}^m$ — key

$\bar{v}_i \in \mathbb{R}^l$ — value

$\bar{x}_1 \to (\bar{q}_1, \bar{k}_1, \bar{v}_1) \to (\bar{q}_i^T \bar{k}_1)$

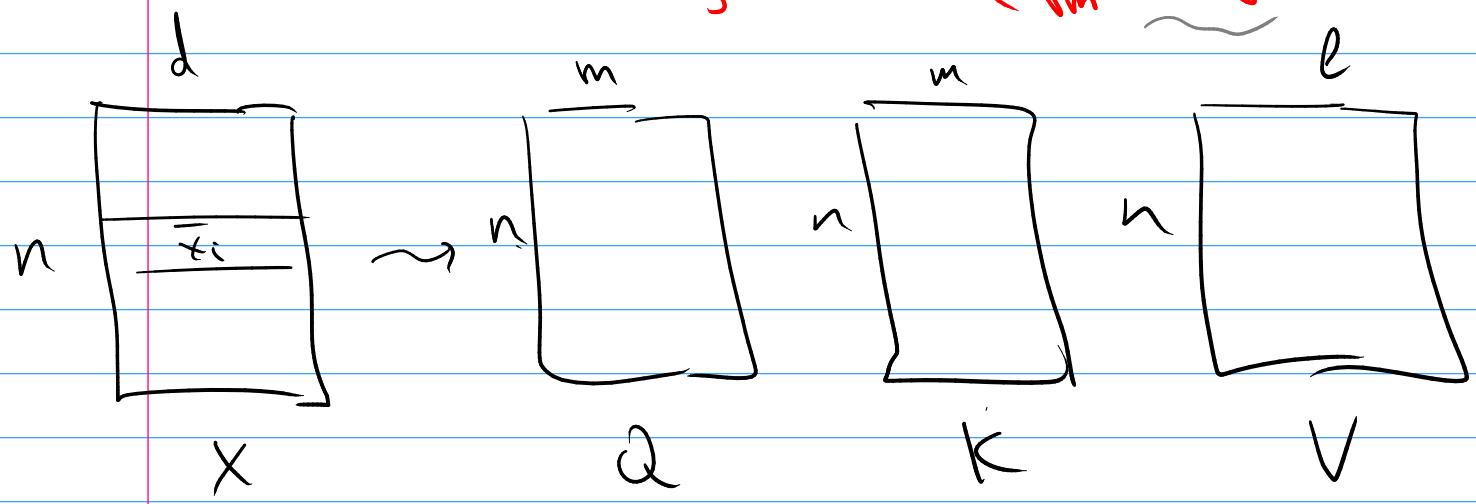
$\bar{x}_2 \to (\bar{q}_2, \bar{k}_2, \bar{v}_2) \to (\bar{q}_i^T \bar{k}_2)$

$\bar{x}_n \to (\bar{q}_n, \bar{k}_n, \bar{v}_n) \to (\bar{q}_i^T \bar{k}_n)$

$a_{ij} \approx \bar{q}_i^T \bar{k}_j$ „насколько $\bar{x}_i$ хочет узнать $\bar{x}_j$"

$\bar{x}_i \to (\bar{q}_i, \bar{k}_i, \bar{v}_i) \longrightarrow \bar{y}_i = \sum_{j=1}^{n} a_{ij} \bar{v}_j$,

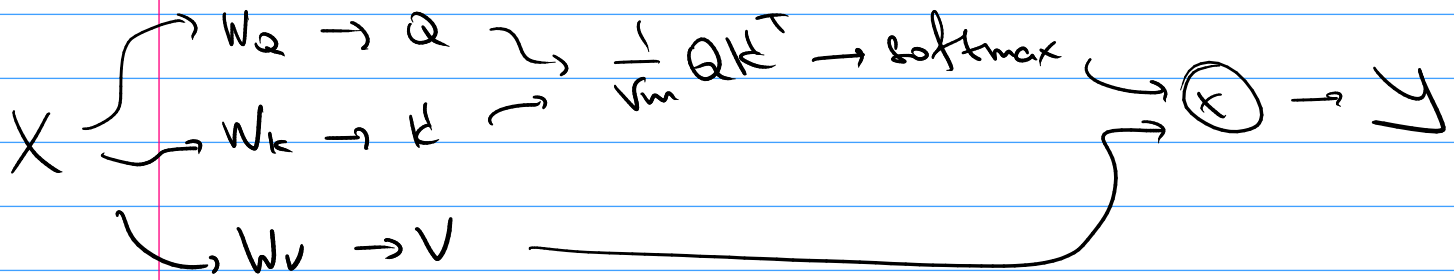
$a_{ij} = \mathrm{softmax}\left(\frac{1}{\sqrt{m}} \bar{q}_i^T \bar{k}_j\right)$



W_Q: $q_i = W_Q x_i$ $\quad Q = W_Q X$

W_K: $k_i = W_K x_i$ $\quad K = W_K X$

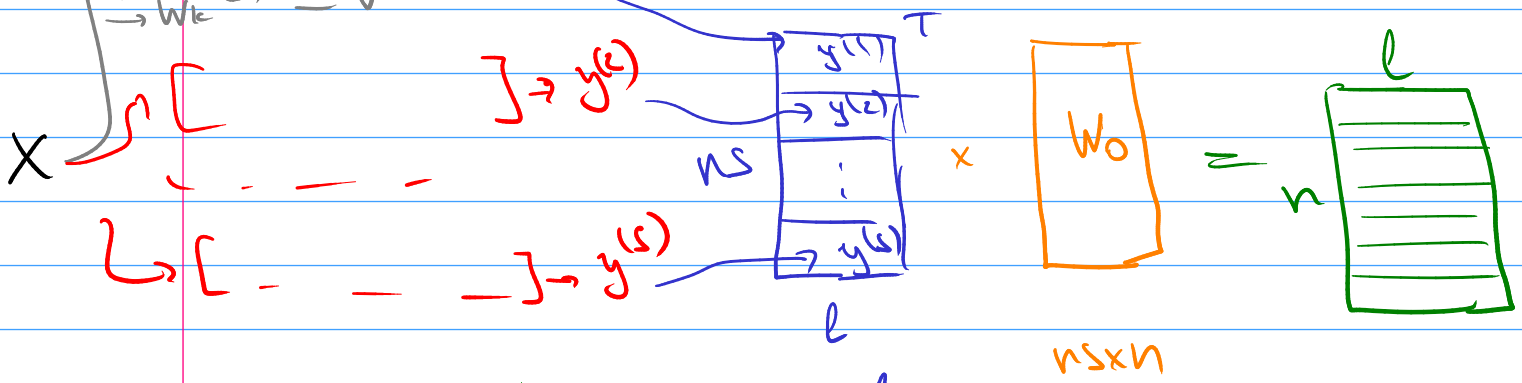
W_V: $v_i = W_V x_i$ $\quad V = W_V X$



Multi-head attention

$(W_Q^{(1)}, W_K^{(1)}, W_V^{(1)}), \ldots, (W_Q^{(s)}, W_K^{(s)}, W_V^{(s)})$

$W_Q^{(1)} = Q^{(1)}$, $W_K^{(1)} = K^{(1)}$, $W_V^{(1)} = V^{(1)}$ $\Rightarrow y^{(1)} = \text{softmax}\left(\frac{1}{\sqrt{m}} Q^{(1)} K^{(1)T}\right) V^{(1)}$

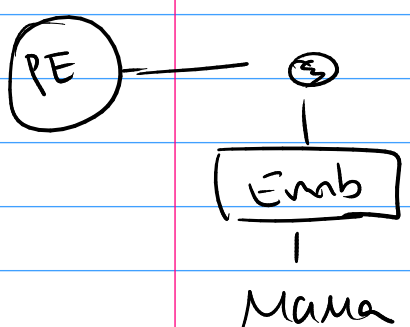


W_O $\quad ns \times n$

W_K, W_Q $\quad d \times m$

W_V $\quad d \times l$

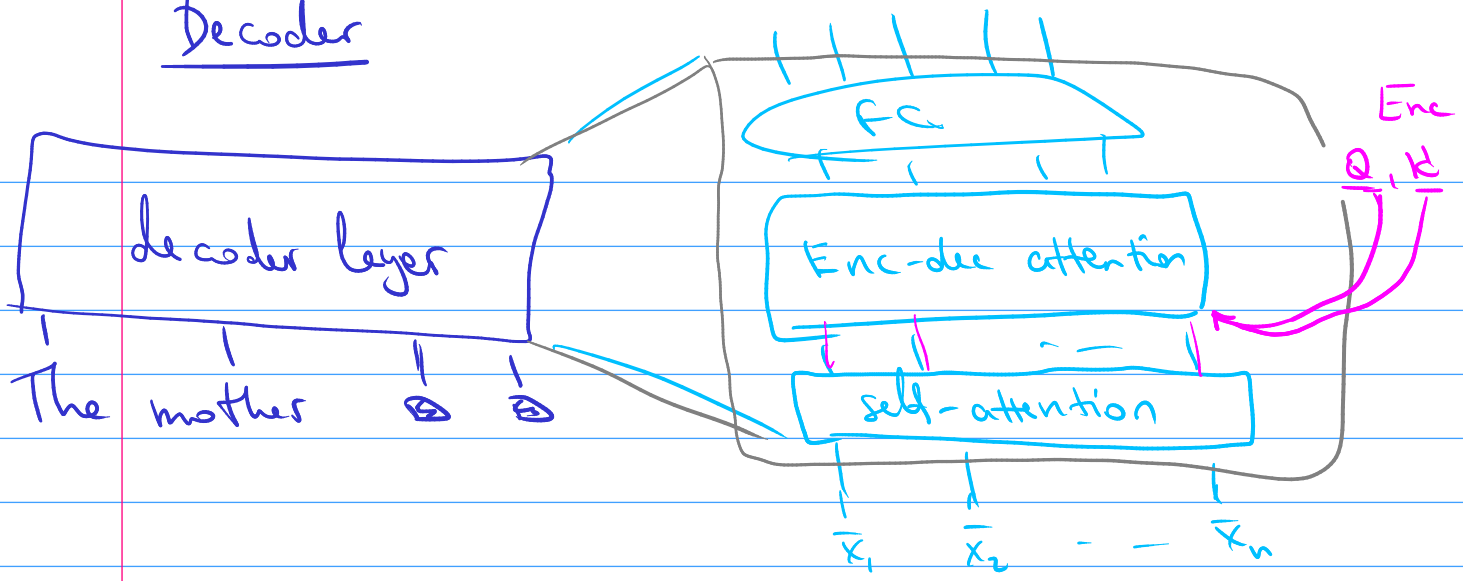
$s(n^2 + dm + dl)$ bcuz



$$PE(pos, 2i) = \sin\left(\left(\frac{pos}{10000}\right)^{2i/d}\right)$$

$$PE(pos, 2i+1) = \cos\left(\left(\frac{pos}{10000}\right)^{2i/d}\right)$$

Decoder


decoder layer
The
mother
FC
Enc-dec attention
self-attention
Enc
Q, K
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_n$