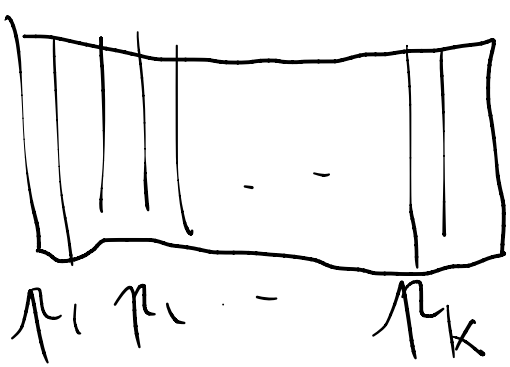
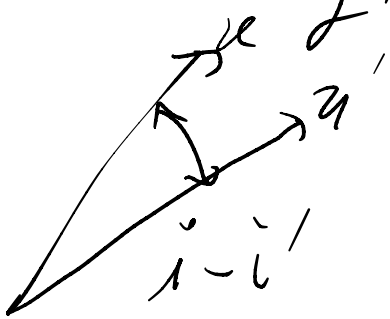
① Tokenizer

② Embeding. $(p, i) \to x \in \mathbb{R}^{d=512}$

$$x = \underbrace{\varphi(p)}_{\text{trained}} + \underbrace{\psi(i)}_{\text{pos. encoding}}$$





$$x = Rot(i/360) \cdot \rho(p)$$

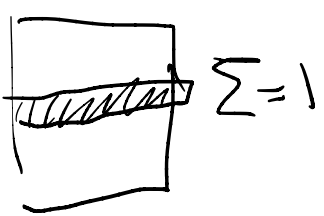
input $X = [x_1\ x_2\ \cdots\ x_N] \in \mathbb{R}^{d \times N}$

$\{x_i\} \longrightarrow \{y_i\}_i$

$X \longrightarrow \left\{ \sum_j w_{ij} \left(V x_j \right) \right\}_i$

$\tilde{w}_{ij} = e^{\langle Q x_i,\ K x_j \rangle_{\mathbb{R}^{16}}}$ (query, key, input)

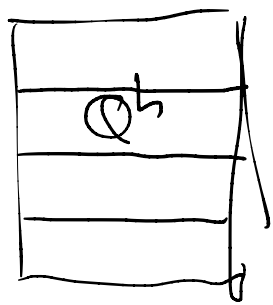
$w_{ij} = \dfrac{\tilde{w}_{ij}}{\sum_l w_{il}}$

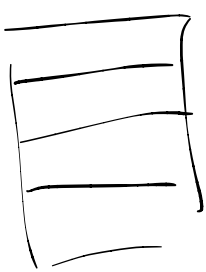


Param $= (Q, K, V)$

MHA : $\{x_i\} \rightarrow \sum_{h=1}^{H} \sum_{j} W_{ij}^{h} \cdot V^{h} x_j$

512

$Q =$  512

$K =$ 

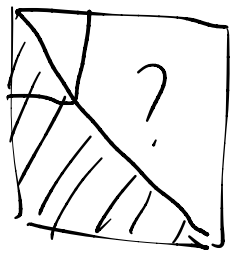
$x_i^0 \rightarrow \left(\frac{x_i}{\|x_i\|} \cdot \lambda \right) \in \mathbb{R}^d \xrightarrow{\text{AMHA}} \left(x_i + y_i^0 \right)_i =: z_i$

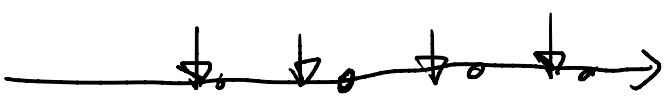
RMSNorm

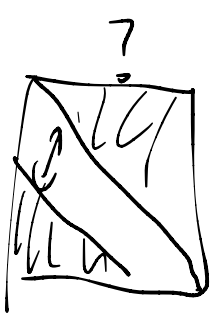
$z \cdot \lambda = \left(z[k] \cdot \lambda[k] \right)$ $\left(z_i + \text{MLP}(z_i)_i \right)$

Encoding $X \xrightarrow{L_1} L_2 \xrightarrow{L_3} Y$

$\tilde{W}_{ij} \cdot \mathbb{1}_{[i \le j]}$.







Input $\{x(t)\}$

output $\{y(t)\}$

$$\begin{cases} \tilde{y}(t+1) = A\tilde{y}(t) + Bx(t) \\ y(t) = C\tilde{y}(t) \end{cases}$$

AR

$$[CB,\ CAB,\ CAAB,\ CA^3B,\ \ldots]$$

SSSM

S^3

$$y(t+1) = \underbrace{A(n(t))}_{S_A} \times \underbrace{y(t)}_{} + \underbrace{B(n(t))}_{S_B} x(t)$$