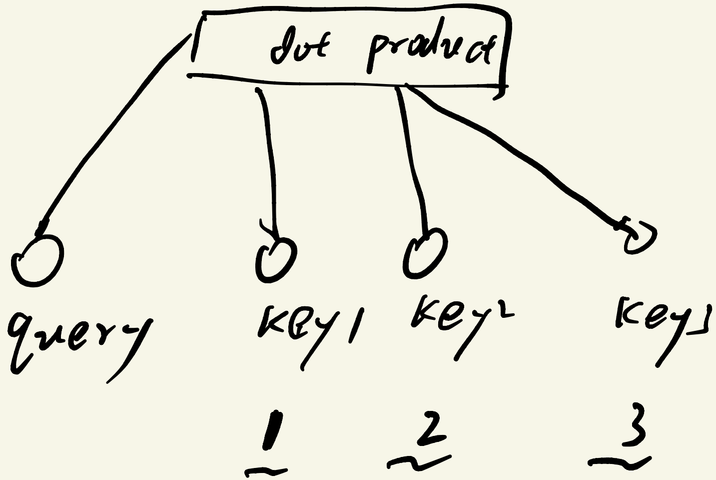
# Lecture 20 transformer & VAE

N X

$$Q \begin{bmatrix} q_1^T \\ q_2^T \\ q_3^T \\ \vdots \\ q_n^T \end{bmatrix} \times K^T \begin{bmatrix} k_1 & k_2 & \cdots & k_m \end{bmatrix}$$

$$n \times m \quad \begin{bmatrix} q_1^T k_1 & q_1^T k_2 & \cdots \\ q_2^T k_1 & q_2^T k_2 & \cdots \end{bmatrix}$$




$$\frac{\exp(1)}{\exp(1)+\exp(2)+\exp(3)}$$

1090.

$[x_1, x_2, x_3)$

$[y_1, y_2, y_3)$

$$\underbrace{x_1 y_1 + x_2 y_2 + x_3 y_3}_{J}$$

$\boxed{n \times m}$

$$\begin{array}{cc} & m \\ \text{attn weight} \quad n & \begin{bmatrix} 0 & & \\ & & \\ & & \end{bmatrix} \end{array}$$

$$[w_1 \quad w_2 \quad \cdots \quad w_m]$$

$$w_1 \vec{V_1} + w_2 \vec{V_2} + \cdots \; w_m \vec{V_m}$$

$W_Q$ $W_K$ $W_V$ parameters

$q_1$ $q_2$ $q_3$

$k_1$ $k_2$ $k_2$

$v_1$ $v_2$ $v_3$

$\boxed{q_1^T k_1}$ $\boxed{q_1^T k_2}$ $q_1^T k_3$

$q_2^T k_1$ $q_2^T k_2$

$$\text{Softmax}\left(\frac{Q^T K}{\sqrt{d_k}}\right) V$$

$$Q \in R^{n \times d}$$

$$K \in R^{m \times d}$$

$n = m$

$n$ = sequence length

$\boxed{n \times n}$

$\underline{\underline{h}}$

$z_0$ $z_1$ $z_2$

cross attention :

Q : from $x$

(K, V) from encoder output

output : [1 3 4 0.5]

after shuffle : [3 4 1 0.5]

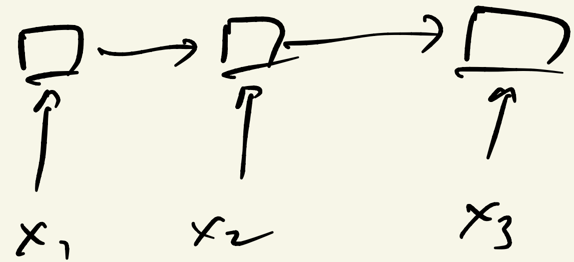
I give an apple to you

you give an

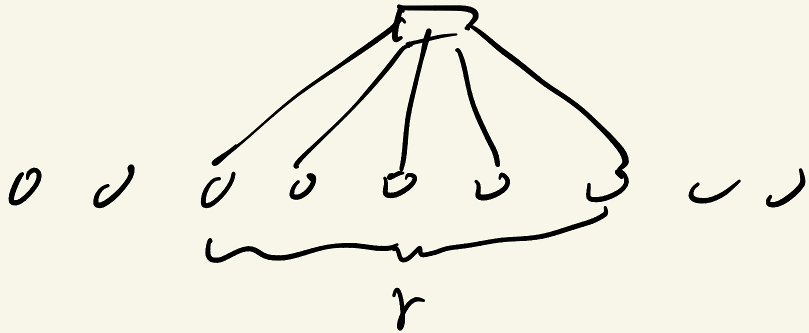
RNN

$d = 512$

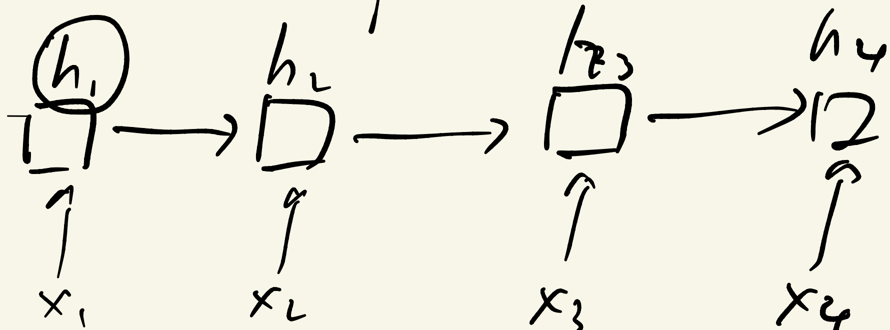
$[0 \quad ]$




$x_3 \quad x_2 \quad x_1$




Sequential operation $O(n)$

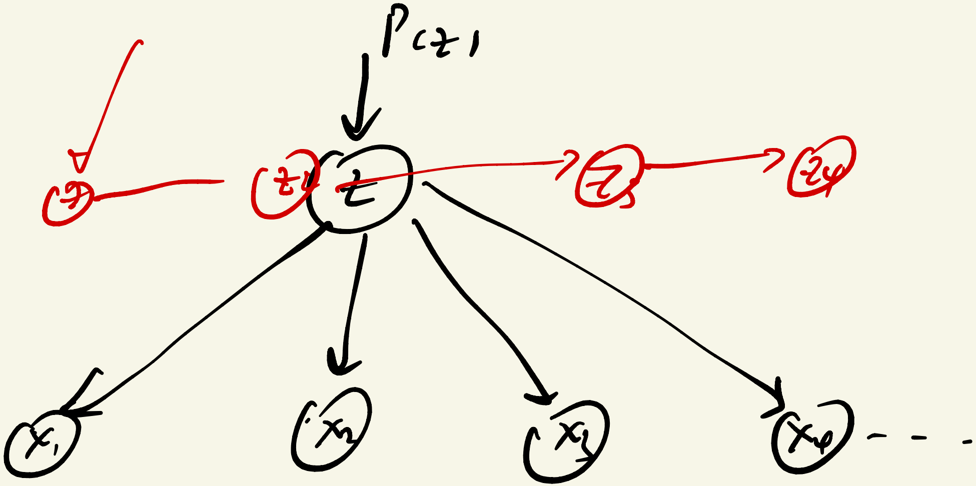


$$P(x_1, x_2, \cdots, x_n)$$

$$= P(x_1)\,P(x_2 \mid x_1)\,P(x_3 \mid x_1, x_2)\cdots P(x_n \mid x_1, \cdots, x_{n-1})$$

$$= P(x_1)\,P(x_2)\,P(x_3)\cdots P(x_n)$$

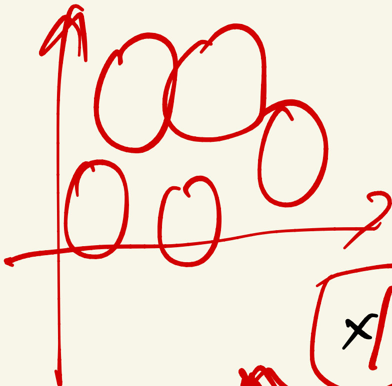
$$x_1 \perp x_2 \perp x_3$$




$$z \sim P(z)$$

$$P(x_1 \mid z)\, P(x_2 \mid z)\, P(x_3 \mid z) \cdots$$

$$\boxed{x_1 \perp x_2} \;?$$



$$x \sim P(X, f(z;\theta))$$

$P \to$ Gaussian

$$x|z \sim \text{Gaussian}(\mu = f(z;\theta), I)$$

NN

$N(0,1)$

$$P(x) = \int_z P(z)\, P(x|z)$$

$$P(x) = \int_z P(z) P(x|z)$$

not Gaussian

$$= E_{z \sim p(z)} P(x|z)$$

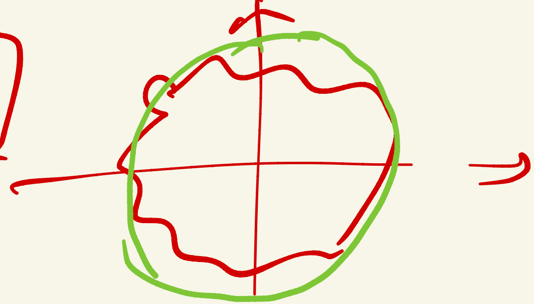
$$P(z|x) \propto P(z) P(x|z)$$

ELBO:

$$E_{z \sim p(z|x)} \log P(x, z; \theta)$$

sample $z \sim P(z|x)$

$\boxed{p(z|x)}$



simple $q(z|x)$  $q(z;\phi)$

easy to sample $z$ from $q(z|x)$

$d[q(z|x), p(z|x)]$

$\underline{\log P(x;\theta)} \geq ELBO$

$\downarrow$

unrelated to Q

when $\boxed{Q = P(z|\theta)}$

$\log P(x) = ELBO$

$$\boxed{ELBO = \log P(x) - KL(Q(z) \| P(z|x))}$$

$Q(z) = P(z|x)$

$\downarrow$

$KL = 0$

$\underset{\phi}{\arg\min}\ \underline{KL(q(z|x;\phi) \| P(z|x))}$

$p(z|x)$ exact inference

$q(z|x) \longrightarrow p(z|x)$, variational inference

MLE $\log p(\cdot)$

$q(z|x;\phi)$ simple

$z \sim q(z|x;\phi)$