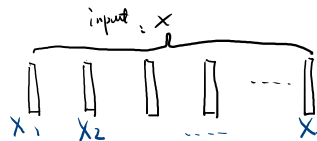


Self-Attention 讲解: 自注意力机制.



1. 首先,  $Q, K, V$  三个元素, 这个元素来自于哪里?  $Q$  输入  $x$  有什么联系呢?

2. 如何得到  $Q, K, V$  三个元素

3. 众所周知, Attention 机制是个加权平均. 那么在 Self-Attention 机制中, 加权对象是谁? Attention 权重又是如何计算出来的呢?  $Q, K, V$  又有什么关系呢?

Answer 1:  $Q, K, V$  是通过输入  $x$  经过个权重变换矩阵  $W_Q, W_K, W_V$  作用线性变换得来的.

$$\begin{cases} Q = x \cdot W_Q = \text{input} \times \underbrace{\text{input} \times \text{dim}}_{W_Q} \\ K = x \cdot W_K \\ V = x \cdot W_V \end{cases}$$

2.  $Q, K, V$  分别  $d, 1, d$  的维度.

3.

Answer 3: 3.1) 如何加权的问题. 加权对象是:  $V$ . 故:  $z = \sum z_i V_i$

3.2) 问题转化成了 2. 如何计算出来的.

$$z = \text{softmax}\left(\frac{Q^T K}{\sqrt{d}}\right)$$

故, 总结一下, 现在有个时刻  $t$  的

时序类型的输入.

$$z_i = \sum_j \text{softmax}\left(\frac{Q_i K_j}{\sqrt{d}}\right) V_j$$

并行计算能完成!

$$X_i = \begin{pmatrix} 1 \\ k \\ v \end{pmatrix}$$

$$W_Q, W_K, W_V$$

$$W_Q \cdot X_i = Q$$

$$W_K \cdot X_i = K$$

$$W_V \cdot X_i = V$$

$$\sum_i Q_i V_i = z$$

$$\text{softmax}\left(\frac{Q^T K}{\sqrt{d}}\right) = z$$

$z_1, z_2$  concat  $(z_1, z_2)$

$W_z$

$\downarrow$

$W_{z_1}$

$W_{z_2}$

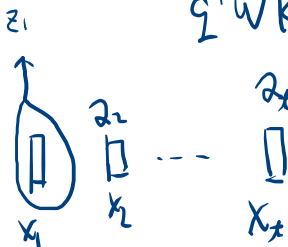
$W_{k_1}$

$W_{k_2}$

$(z_{new})$

$$\frac{Q^T \tanh(W_K + b)}{\sqrt{d}}$$

$$Q^T W_K$$



$$z_2 K + z_3 V$$











