神经网络第十七章

2022年7月23日 星期六 00:40

Transformer Network Intuition

RNN GRU LSTM.

Complex → 序列模型

Self - Attention

A(q, K, V) = attention - based vector representation of a word.

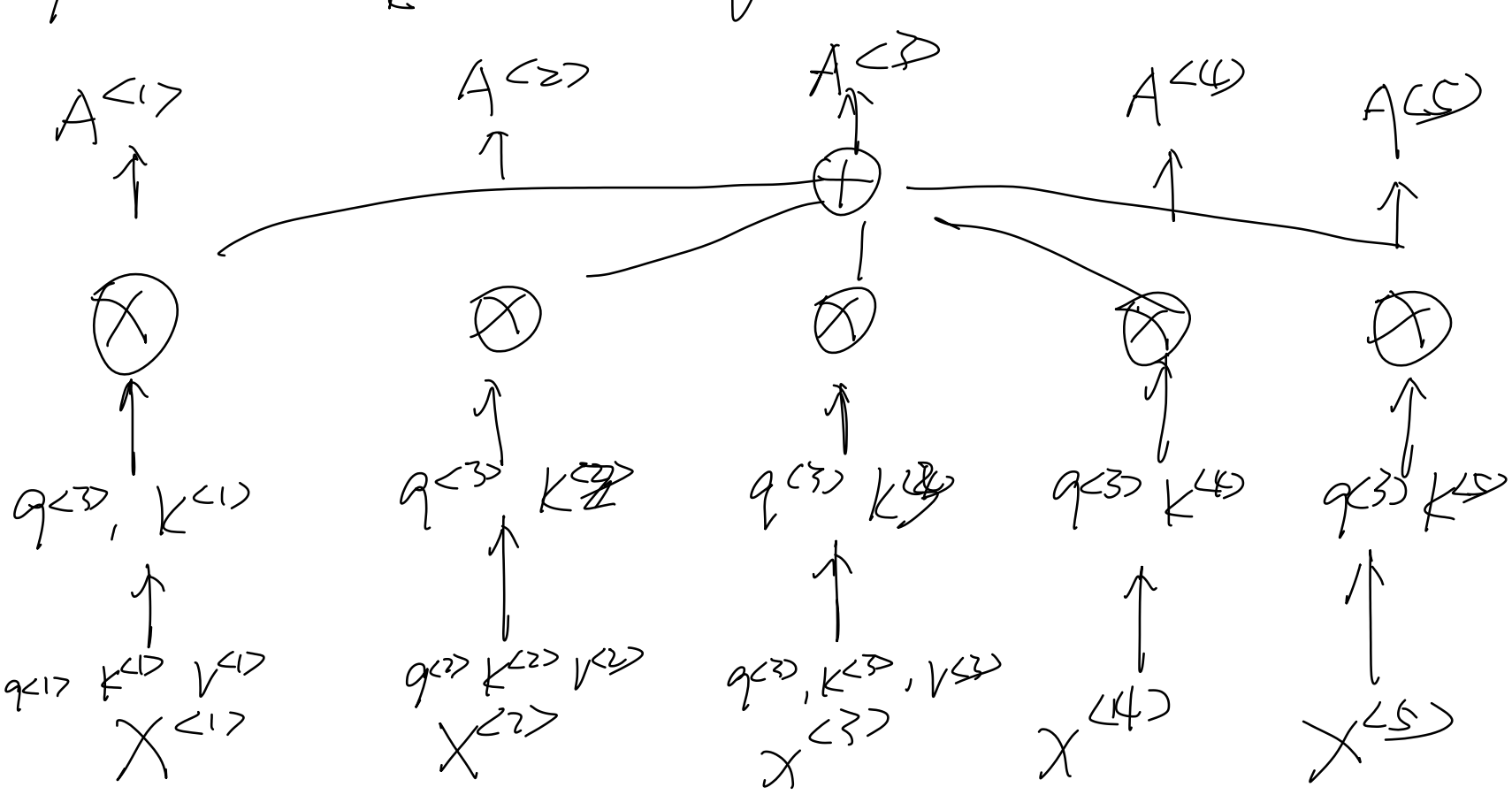
RNN:

$$a^{<t,t'>} = \frac{exple^{<t,t'>}}{\sum_{t'=1}^{Tx} exple^{<t,t'>})}$$

$$A(q,K,V) = \sum_i \frac{\exp(q \cdot k^{<i>})}{\sum_j \exp(q \cdot k^{<j>})} v^{<i>}$$

self - Attention

Query (Q)	Key (K)	Value (V)
$q^{<1>}$	$k^{<1>}$	$v^{<1>}$
$q^{<2>}$	$k^{<2>}$	$v^{<2>}$
$q^{<3>}$	$k^{<3>}$	$v^{<3>}$
$q^{<4>}$	$k^{<4>}$	$v^{<4>}$
$q^{<5>}$	$k^{<5>}$	$v^{<5>}$



Multi - Head Attention & Transformer Network