

Self Attention Numericals

Step 1

Date _____

$$Q = KW^T$$

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

3×4 4×3
 3×3

$$Q = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 2 & 2 \\ 2 & 1 & 3 \end{bmatrix} \quad \checkmark$$

$$K = KW^T$$

$$K = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

$$K = \begin{bmatrix} 0 & 1 & 1 \\ 4 & 4 & 0 \\ 2 & 3 & 1 \end{bmatrix}$$

Date _____

$$V = X W^V$$

$$V = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 0 & 2 & 0 & 7 \\ 0 & 3 & 0 & 0 \\ 1 & 0 & 3 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 8 & 0 \\ 2 & 6 & 3 \end{bmatrix}$$

Step 2

$$E = \emptyset K.T$$

$$E = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 2 & 2 \\ 2 & 1 & 3 \end{bmatrix} \times \begin{bmatrix} 0 & 4 & 2 \\ 1 & 4 & 3 \\ 1 & 0 & 1 \end{bmatrix}$$

$$E = \begin{bmatrix} 2 & 4 & 4 \\ 4 & 16 & 18 \\ 4 & 12 & 10 \end{bmatrix}$$

Step 3

$$A = \text{softmax}(E)$$

\Rightarrow ~~e^2, e^4, e^4~~

$$\left[\frac{e^2}{e^2 + e^4 + e^4}, \frac{e^4}{e^2 + e^4 + e^4}, \frac{e^4}{e^2 + e^4 + e^4} \right]$$

$$= [0.06, 0.468, 0.468] \rightarrow \text{Input}_1$$

$$\Rightarrow \left[\frac{e^4}{e^4 + e^{16} + e^{18}}, \frac{e^{16}}{e^4 + e^{16} + e^{18}}, \frac{e^{18}}{e^4 + e^{16} + e^{18}} \right]$$

$$= [0.000007, 0.119, 0.88] \rightarrow \text{Input}_2$$

$$\Rightarrow \left[\frac{e^4}{e^4 + e^{10} + e^{12}}, \frac{e^{12}}{e^4 + e^{12} + e^{10}}, \frac{e^{10}}{e^4 + e^{12} + e^{10}} \right]$$

$$= [0.00029, 0.88, 0.119] \rightarrow \text{Input}_3$$

∴

$$A = \begin{bmatrix} 0.06 & 0.468 & 0.468 \\ 0.000007 & 0.119 & 0.88 \\ 0.00029 & 0.88 & 0.119 \end{bmatrix}$$

✓

~~Step 4~~

Date _____

$$\text{Output} = A \times V$$

$$\text{Output} = \begin{bmatrix} 0.06 & 0.468 & 0.468 \\ 0.000007 & 0.119 & 0.88 \\ 0.00029 & 0.88 & 0.119 \end{bmatrix} X$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 8 & 0 \\ 2 & 6 & 3 \end{bmatrix}$$

$$\text{Output} = \begin{bmatrix} 1.0932 & 6.67 & 1.58 \\ 1.99 & 6.23 & 2.64 \\ 1.099 & 7.75 & 0.357 \end{bmatrix} \cancel{\text{X}}$$

~~1~~ ~~2~~ ~~7~~ ~~1.58~~

~~2~~

X — X — X

Transformer Numericals

1.) Masked Self-Attention:

$$\text{Step 1: } Q = X W^Q$$

$$Q = \begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.4 & 0.5 & 0.6 \\ 0.7 & 0.8 & 0.9 \\ 1.0 & 1.1 & 1.2 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 2 \\ 7 & 1 & 9 \end{bmatrix}$$

$$4 \times 3 \quad 3 \times 3$$

$$4 \times 3$$

$$Q = \begin{bmatrix} 3.15 & 3.4 \\ 6.6 & 3.9 & 7.6 \\ 10.2 & 6.3 & 11.8 \\ 13.8 & 8.7 & 16 \end{bmatrix}$$

$$K = X W^K$$

$$K = \begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.4 & 0.5 & 0.6 \\ 0.7 & 0.8 & 0.9 \\ 1.0 & 1.1 & 1.2 \end{bmatrix} \times \begin{bmatrix} 1 & 6 & 9 \\ 7 & 3 & 1 \\ 9 & 2 & 1 \end{bmatrix}$$

$$K = \begin{bmatrix} 4.2 & 1.8 & 1.4 \\ 9.3 & 5.1 & 4.7 \\ 14.4 & 8.4 & 8 \\ 19.5 & 11.7 & 11.3 \end{bmatrix}$$

$$V = X W^V$$

$$V = \begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.4 & 0.5 & 0.6 \\ 0.7 & 0.8 & 0.9 \\ 1.0 & 1.1 & 1.2 \end{bmatrix} \times \begin{bmatrix} 2 & 4 & 6 \\ 8 & 0 & 2 \\ 16 & 8 \end{bmatrix}$$

$$V = \begin{bmatrix} 2.1 & 2.2 & 3.4 \\ 5.4 & 5.2 & 8.2 \\ 8.7 & 8.2 & 13 \\ 12 & 11.2 & 17.8 \end{bmatrix}$$

Step 2 :

$$\textcircled{2} E = \emptyset K T$$

$$E = \begin{bmatrix} 3 & 1.5 & 3.4 \\ 6.4 & 3.9 & 7.6 \\ 10.2 & 6.3 & 11.8 \\ 13.8 & 8.7 & 16 \end{bmatrix} \times \begin{bmatrix} 4.2 & 9.3 & 14.4 & 19.5 \\ 1.8 & 5.1 & 8.4 & 11.7 \\ 1.4 & 4.7 & 8 & 11.3 \end{bmatrix}$$

$4 \times 3 \quad 4 \times 4 \quad 3 \times 4$

	20.06	51.53	83	114.47
E =	45.38	116.99	188.6	260.21
	70.7	182.45	294.2	405.95
	96.02	247.91	399.8	551.69

After applying marking :-

E =	20.06	-∞	-∞	-∞
	45.38	116.99	-∞	-∞
	70.7	182.45	294.2	-∞
	96.02	247.91	399.8	551.69

Step 3

$$A = \text{Sthmn}(E)$$

A =	1	0	0	0	1
	7.9×10^{-32}	1	0	0	
	8.6×10^{-98}	2.9×10^{-49}	1	0	
	1.7×10^{-72}	1.02×10^{-76}			
	1.27×10^{592}	1.17×10^{32}	1.08×10^{66}	1	

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Step 4

$$\text{Output} = A \times V$$

$$\text{Output} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 2 \cdot 1 & 2 \cdot 2 & 3 \cdot 4 \\ 5 \cdot 4 & 5 \cdot 2 & 8 \cdot 2 \\ 8 \cdot 7 & 8 \cdot 2 & 13 \\ 12 & 11 \cdot 2 & 17 \cdot 8 \end{bmatrix}$$

$$\text{Output} = \begin{bmatrix} 2 \cdot 1 & 2 \cdot 2 & 3 \cdot 4 \\ 5 \cdot 4 & 5 \cdot 2 & 8 \cdot 2 \\ 8 \cdot 7 & 8 \cdot 2 & 13 \\ 12 & 11 \cdot 2 & 17 \cdot 8 \end{bmatrix} \quad \text{Ans}$$

X — X — X

2.) Cross-Attention:

Step 1: $Q = XW^Q$ (Computed using output sentence)

$$Q = \begin{bmatrix} 0.4 & 0.1 & 0.8 \\ 0.9 & 0.7 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 2 \\ 7 & 1 & 9 \end{bmatrix}$$

$2 \times 3 \quad 3 \times 3 \quad 3 \times 3$

$$Q = \begin{bmatrix} 6.4 & 2.1 & 8.6 \\ 5.1 & 5.5 & 5.9 \end{bmatrix}$$

Now K & V will be computed using input Sentence.

\therefore Already Computed K & V in part 1

$$\therefore K = \begin{bmatrix} 4.2 & 1.8 & 1.4 \\ 9.3 & 5.1 & 4.7 \\ 14.4 & 8.4 & 8 \\ 19.5 & 11.7 & 11.3 \end{bmatrix}$$

$$\therefore V = \begin{bmatrix} 2.1 & 2.2 & 3.4 \\ 5.4 & 5.2 & 8.2 \\ 8.7 & 8.2 & 13 \\ 12 & 11.2 & 17.8 \end{bmatrix}$$

Step 2: $E = QK^T$

$$E = \begin{bmatrix} 6.4 & 2.1 & 8.6 \\ 5.1 & 5.5 & 5.9 \end{bmatrix} \begin{bmatrix} 4.2 & 9.3 & 14.4 & 19.5 \\ 1.8 & 5.1 & 8.4 & 11.7 \\ 1.4 & 4.7 & 8.1 & 11.3 \end{bmatrix}$$

$2 \times \cancel{3}$ $\cancel{3} \times 4$ 2×4

$$E = \begin{bmatrix} 42.7 & 110.65 & 178.6 & 246.55 \\ 39.58 & 103.21 & 166.84 & 230.47 \end{bmatrix}$$

Step 3: $A = \text{softmax}(E)$

$$A = \begin{bmatrix} 2.9 \times 10^{-89} & 9.5 \times 10^{-60} & 3.08 \times 10^{-30} & 1 \\ 1.2 \times 10^{-83} & 5.3 \times 10^{-56} & 2.3 \times 10^{-28} & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Date _____

Step 4: Output = A × V

$$\text{Output} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 2.1 & 2.2 & 3.4 \\ 5.4 & 5.2 & 8.2 \\ 8.7 & 8.2 & 13 \\ 12 & 11.2 & 17.8 \end{bmatrix}$$

2×4 4×3 2×3

$$\text{Output} = \begin{bmatrix} 12 & 11.2 & 17.8 \\ 12 & 11.2 & 17.8 \end{bmatrix}$$

Ans

X ————— X ————— X