

Single Head Attention

We are good friend

We=[0.5,0.4,0.8,0.1]

Are=[0.2,0.6,0.32,0.7]

Good=[0.65,0.23,0.54,0.66]

Friend=[0.32,0.11,0.74,0.92]

Query:

1	1	1	1
0	0	0	0
1	1	1	1

Key:

0	0	0	0
1	1	1	1
1	1	1	1

Value:

1	1	1	1
1	1	1	1
0	0	0	0

Input Matrix:

0.5	0.4	0.8	0.1
0.2	0.6	0.32	0.7
0.65	0.23	0.54	0.66
0.32	0.11	0.74	0.92

$Q = \text{Input Matrix} * \text{Query}_{(\text{Transpose})}$

1.8	0	1.8
1.82	0	1.82
2.08	0	2.08
2.09	0	2.09

$K = \text{Input Matrix} * \text{Key}_{(\text{Transpose})}$

0	1.8	1.8
0	1.82	1.82
0	2.08	2.08
0	2.09	2.09

$V = \text{Input Matrix} * \text{Value}_{(\text{Transpose})}$

1.8	1.8	0
1.82	1.82	0
2.08	2.08	0
2.09	2.09	0

$QK = Q * K_{(\text{Transpose})}$

3.24	3.27	3.744	3.762
3.276	3.3124	3.7856	3.8038
3.744	3.7856	4.3264	4.3472
3.762	3.8038	4.3472	4.3681

S=Softmax(QK):

*softmax example for the (0,0) index: $e^{3.24}/(e^{3.24}+e^{3.27}+e^{3.744}+e^{3.762})$
=25.533/137.146=0.816

0.186	0.1918	0.308	0.313
0.185	0.192	0.308	0.314
0.1766	0.184	0.316	0.3229
0.1763	0.1838	0.3165	0.3232

Result=S*V:

Y _{we}	1.9786	1.978686	0
Y _{are}	1.97934	1.97934	0
Y _{good}	1.99951	1.999531	0
Y _{friend}	1.985664	1.985664	0