Self Attention

Features

x_1	x_2	x_3	x_4
2	0	0	2
0	1	0	0
0	2	1	0
0	0	1	1
2	0	0	0
1	0	1	1

MatMul
(K^TQ)

q_1	q_2	q_3	q_4

k_1^T			
k_2^T			
k_3^T			
k_4^T			

W_Q

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

q_1	q_2	q_3	q_4

W_K

0	0	1	0	0	0
0	1	0	0	0	0
1	0	0	0	0	-1

k_1	k_2	k_3	k_4

Scale

$$\frac{\square}{\sqrt{dk}}$$

Softmax

$e^{\square}$

○ Σ

Attention Weight Matrix (A)

MatMul

W_V

10	0	0	0	0	0
0	0	0	10	0	0
0	10	0	0	0	0

v_1	v_2	v_3	v_4

z_1	z_2	z_3	z_4

Attention Weighted Features

FFN

© 2024 Tom Yeh