

Self-Attention as Matrix Calculation

$$X \times W^Q = Q$$


$$X \times W^K = K$$


$$X \times W^V = V$$


$$\text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \times V = Z$$
