

Self-Attention Mechanism Calculation

Given the input sequence

$$X = \begin{bmatrix} 0.8 & -0.4 \\ 1.2 & 0.3 \\ -0.5 & 1.1 \end{bmatrix}$$

and learnable weight matrices

$$W_Q = \begin{bmatrix} 0.6 & -0.2 \\ 0.1 & 0.5 \end{bmatrix}, \quad W_K = \begin{bmatrix} -0.3 & 0.7 \\ 0.4 & 0.9 \end{bmatrix}, \quad W_V = \begin{bmatrix} 1.0 & -0.5 \\ 0.2 & 0.8 \end{bmatrix}.$$

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

$$Q = \begin{bmatrix} 0.44 & -0.36 \\ 0.75 & -0.09 \\ -0.19 & 0.65 \end{bmatrix} \quad K = \begin{bmatrix} -0.4 & 0.20 \\ -0.24 & 1.11 \\ 0.59 & 0.64 \end{bmatrix} \quad V = \begin{bmatrix} 0.72 & -0.72 \\ 1.26 & -0.36 \\ -0.28 & 1.13 \end{bmatrix}$$

$$QK^T = \begin{bmatrix} -0.2480 & -0.5052 & 0.0292 \\ -0.3180 & -0.2799 & 0.3849 \\ 0.2060 & 0.7671 & 0.3039 \end{bmatrix}$$

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

$$\text{Softmax_Scores} = \begin{bmatrix} 0.3233 & 0.2500 & 0.4266 \\ 0.2464 & 0.2560 & 0.4976 \\ 0.2594 & 0.4546 & 0.2861 \end{bmatrix}$$

$$\text{attention scores} = \begin{bmatrix} 0.4284 & 0.1593 \\ 0.3606 & 0.2928 \\ 0.6794 & -0.0272 \end{bmatrix}$$

```
C:\Users\tch0905\PycharmProjects\ai3120\venv\Scripts\python.exe C:\Users\tch0905\PycharmProjects\ai3120\ass2\Assignment2_Q2.py
Scores shape: torch.Size([2, 3, 4])
Scores: tensor([[[ 0.5865,  0.5994, -0.3947, -0.1969],
                  [ 0.5851,  0.5983, -0.3935, -0.1972],
                  [ 0.5674,  0.5815, -0.3774, -0.1999]],
                [[ 0.5864,  0.1387, -0.0796, -0.0647],
                  [ 0.4677,  0.1198, -0.1733,  0.0150],
                  [ 0.5065,  0.1500,  0.0420, -0.1510]]], grad_fn=<UnsafeViewBackward0>)

Process finished with exit code 0
```

Figure 1: Enter Caption