Home Task

Suppose, you have the following sequence: "This is my first embedding computation". You have to compute Multi-head Self-Attention on this sequence, that has an embedding of 4 dimensions, and divide into heads of 4.

$$H_1 = Softmax\left(\frac{Q^1 * K^1}{\sqrt{dk}}\right) * V^1$$

$$H_2 = Softmax\left(\frac{Q^2 * K^2}{\sqrt{dk}}\right) * V^2$$

$$H_3 = Softmax\left(\frac{Q^3 * K^3}{\sqrt{dk}}\right) * V^3$$

$$H_4 = Softmax\left(\frac{Q^4 * K^4}{\sqrt{dk}}\right) * V^4$$

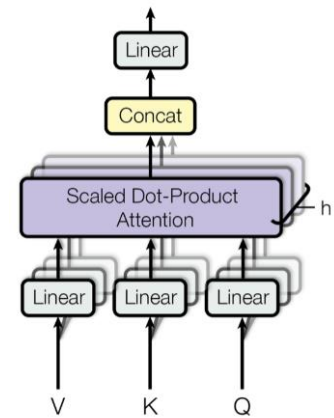$$Multi\ Head\ ATT(Q,K,V) = concat\ (H_1, H_2, H_3, H_4) * W^o$$

Note:

$Q^1 = Q * W^1,\ Q^2 = Q * W^2, Q^3 = Q * W^3, Q^4 = Q * W^4$

$K^1 = K * W^1,\ K^2 = K * W^2, K^3 = K * W^3, K^4 = K * W^4$

$V^1 = V * W^1, V^2 = V * W^2, V^3 = V * W^3,\ V^4 = V * W^4$

Assume the token embedding to be the following:

Note: Transform below data into $seq * Dmodel$ and consider X as $Q, K\ and\ V$

• This: 0.31, 0.22, 0.99, 0.04

• is: 0.21, 0.42, 0.09, 0.06

• my: 0.72, 0.41, 0.30, 0.39

• first: 0.51, 0.31, 0.87, 0.78

• embedding: 0.62, 0.73, 0.11, 0.12

• computation: 0.72, 0.11, 0.14, 0.15

Assume the weights to be:

W^1 [0.8 0.3 0.9 0.6]^T ➔ 4x1 ($d_{model} * dk$) Dimensions

W^2 [0.3 0.8 0.9 0.6]^T➔ 4x1 Dimensions

W^3 [0.8 0.3 0.6 0.9]^T➔ 4x1 Dimensions

W^4 [0.6 0.3 0.9 0.8]^T➔ 4x1 Dimensions

W^o [1 1 1 1; 0.5 0.5 0.5 0.5; 0.7 0.7 0.7 0.7; 0.3 0.3 0.3 0.3] ➔ 4x4 Dimensions

Compute the Multi-head Self-Attention for this sequence, given the above weight matrix.