Self Attention Numerical

```
x = [
  [1, 0, 1, 0], # Input 1
  [0, 2, 0, 2], # Input 2
  [1, 1, 1, 1]  # Input 3
 ]
```

```
w_key = [
  [0, 0, 1],
  [1, 1, 0],
  [0, 1, 0],
  [1, 1, 0] ]
w_query = [
  [1, 0, 1],
  [1, 0, 0],
  [0, 0, 1],
  [0, 1, 1] ]
w_value = [
  [0, 2, 0],
  [0, 3, 0],
  [1, 0, 3],
  [1, 1, 0] ]
```

Solve the above example. For simplicity dk=1.

Round off answer =

```
[[2.        7.        1.5       ]
 [2.        8.        0.        ]
 [2.        7.7999997 0.3       ]]
```

Transformer Numerical

Input sentence: "Transformers transforming our lives"
Output sentence: “For sure”

The embedding matrix for the words is represented as follows:

"Transformers": [0.1, 0.2, 0.3]
"transforming": [0.4, 0.5, 0.6]
"our": [0.7, 0.8, 0.9]
"lives": [1.0, 1.1, 1.2]
"For": [0.4, 0.1, 0.8]
"sure": [0.9, 0.7, 0.2]

$$W_q = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 2 \\ 7 & 1 & 9 \end{bmatrix}$$

$$W_k = \begin{bmatrix} 1 & 6 & 9 \\ 7 & 3 & 1 \\ 9 & 2 & 1 \end{bmatrix}$$

$$W_v = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 0 & 2 \\ 1 & 6 & 8 \end{bmatrix}$$

Assuming the weight matrices:

For simplicity, assume dk is 1.

1. Calculate masked self-attention for the above input sentence using the provided weight matrices. Show all steps clearly.
2. Calculate cross-attention for the above input and output sentence using the provided weight matrices. Show all steps clearly.