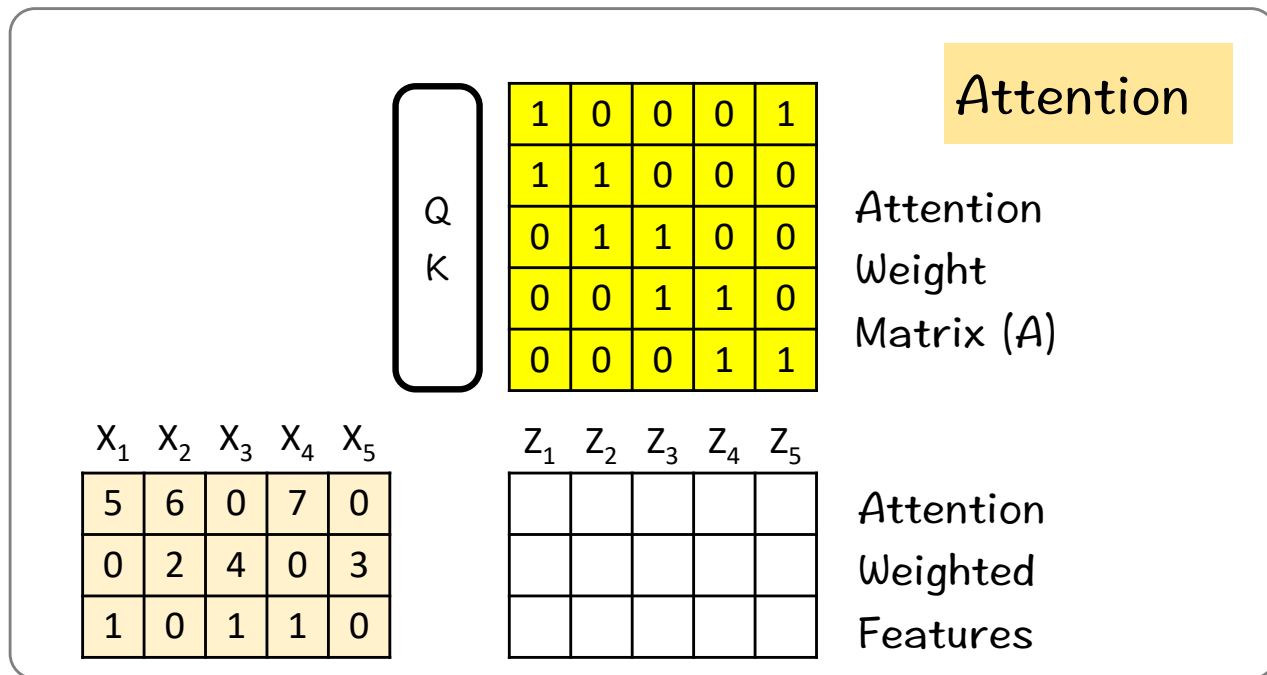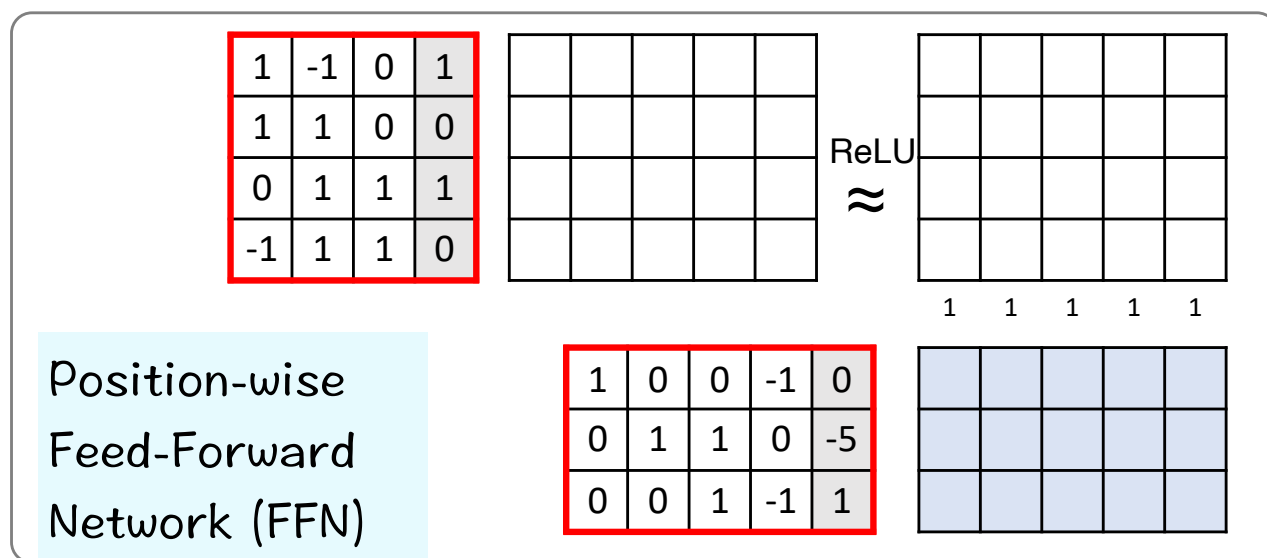# Transformer

Features from the
Previous Block
↓ ↓ ↓ ↓ ↓

Q
K

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

Attention
Weight
Matrix (A)

X_1 X_2 X_3 X_4 X_5

5	6	0	7	0
0	2	4	0	3
1	0	1	1	0

Z_1 Z_2 Z_3 Z_4 Z_5

Attention
Weighted
Features

1 1 1 1 1

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

ReLU
≈

1 1 1 1 1

Position-wise
Feed-Forward
Network (FFN)

1	0	0	-1	0
0	1	1	0	-5
0	0	1	-1	1

↓ ↓ ↓ ↓ ↓
Next Block