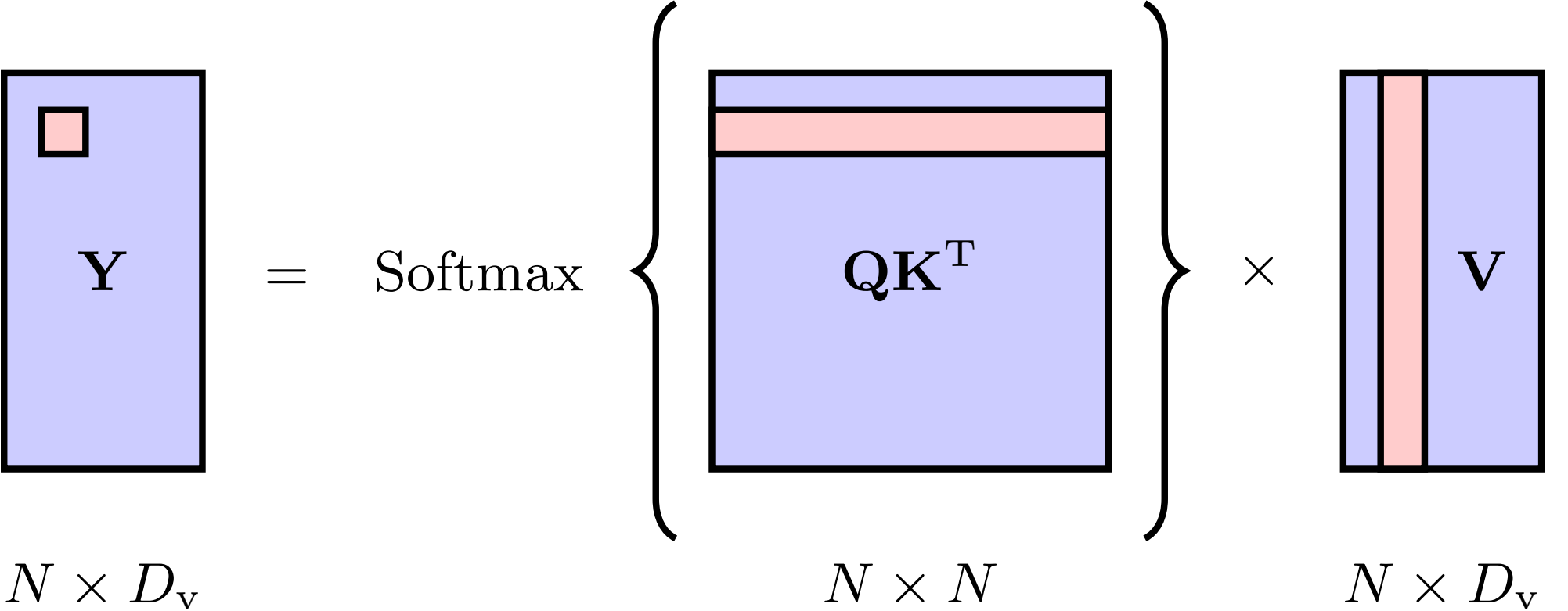
Y
=
Softmax
QK^T
×
V
N × D_v
N × N
N × D_v