



Diagram illustrating the matrix A used in the Softmax operation:

	トークン1	トークン2	トークン3	トークン4
トークン1				
トークン2				
トークン3				
トークン4				

$A = \text{Softmax}$

Diagram illustrating the matrix $Q^t K$ used in the matrix multiplication:

	トークン1	トークン2	トークン3	トークン4
トークン1				
トークン2				
トークン3				
トークン4				

$Q^t K$

Diagram illustrating the matrix S used in the matrix multiplication:

	トークン1	トークン2	トークン3	トークン4
トークン1				
トークン2				
トークン3				
トークン4				

S

\sqrt{d}