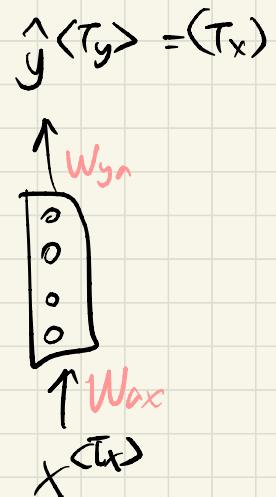
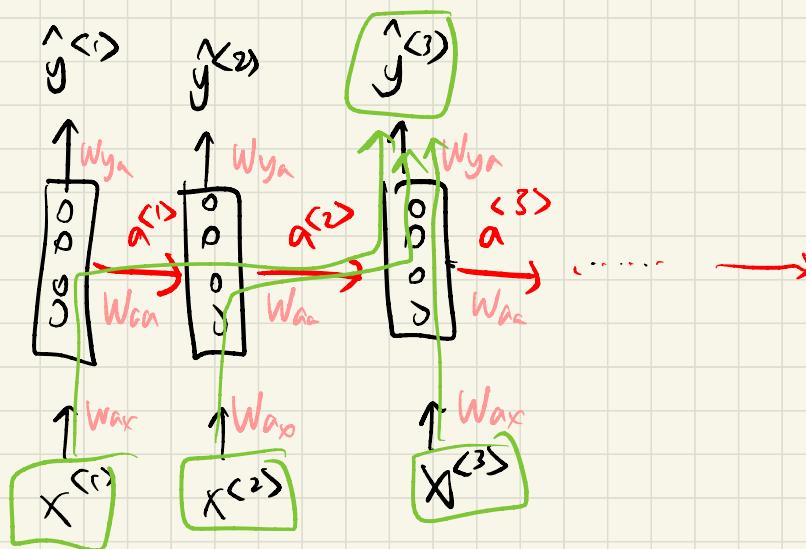
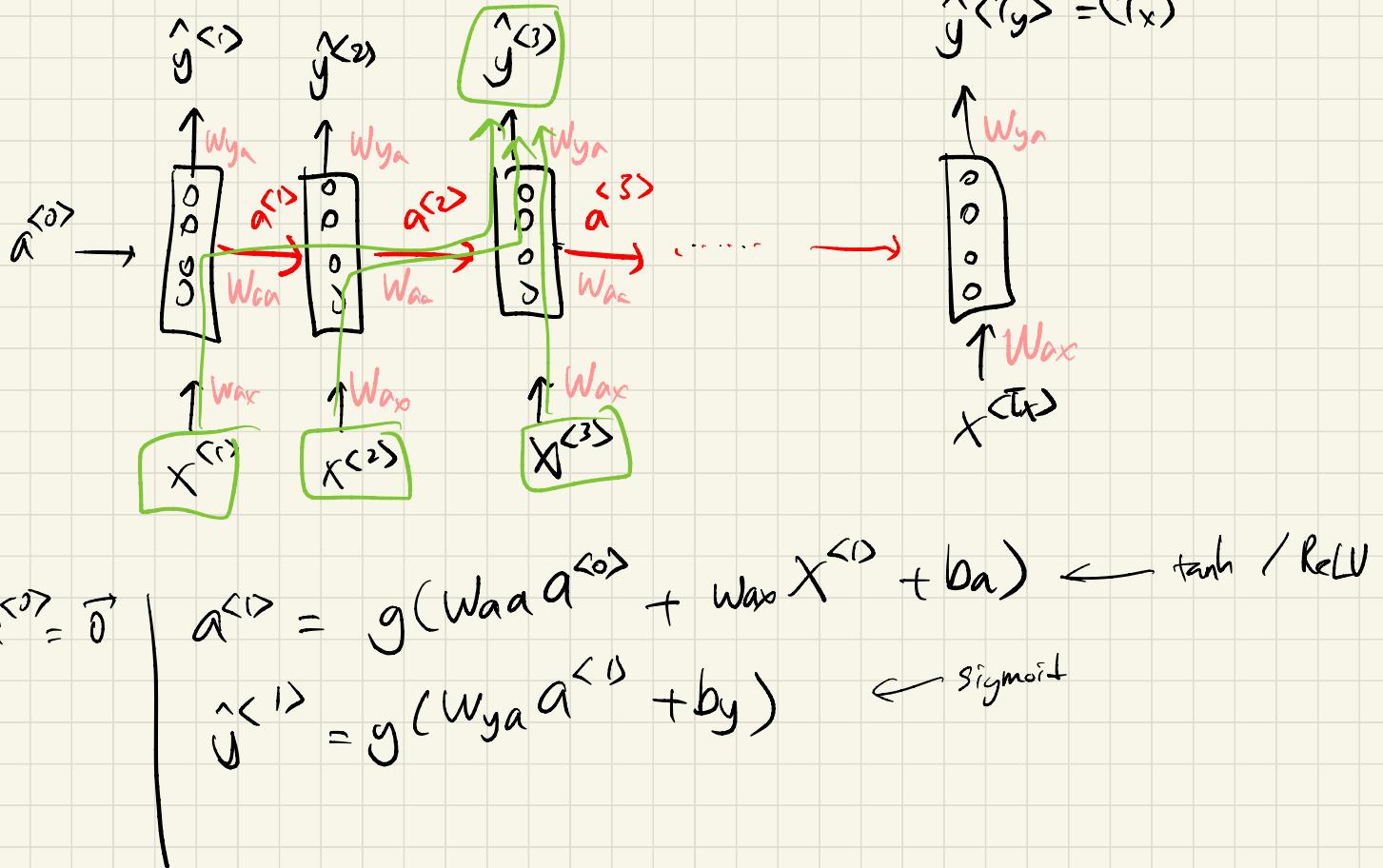


[RNN]

Sentence "v"

" $x^{<1>} , x^{<2>} , \dots , x^{<T_x>}$ "





ex) (100, 100)

(100, 10000)

$$\hat{a}^{(t)} = g(W_a \cdot a^{(t-1)} + W_{ax} \cdot x^{(t)} + b_a)$$

$$\hat{y}^{(t)} = g(W_y \cdot a^{(t)} + b_y)$$

$$a^{(t)} = g(W_a [a^{(t-1)}, x^{(t)}] + b_a)$$

$$\begin{matrix} 100 \\ 100 \end{matrix} \left[\begin{matrix} W_{aa} & | & W_{ax} \end{matrix} \right] = W_a$$

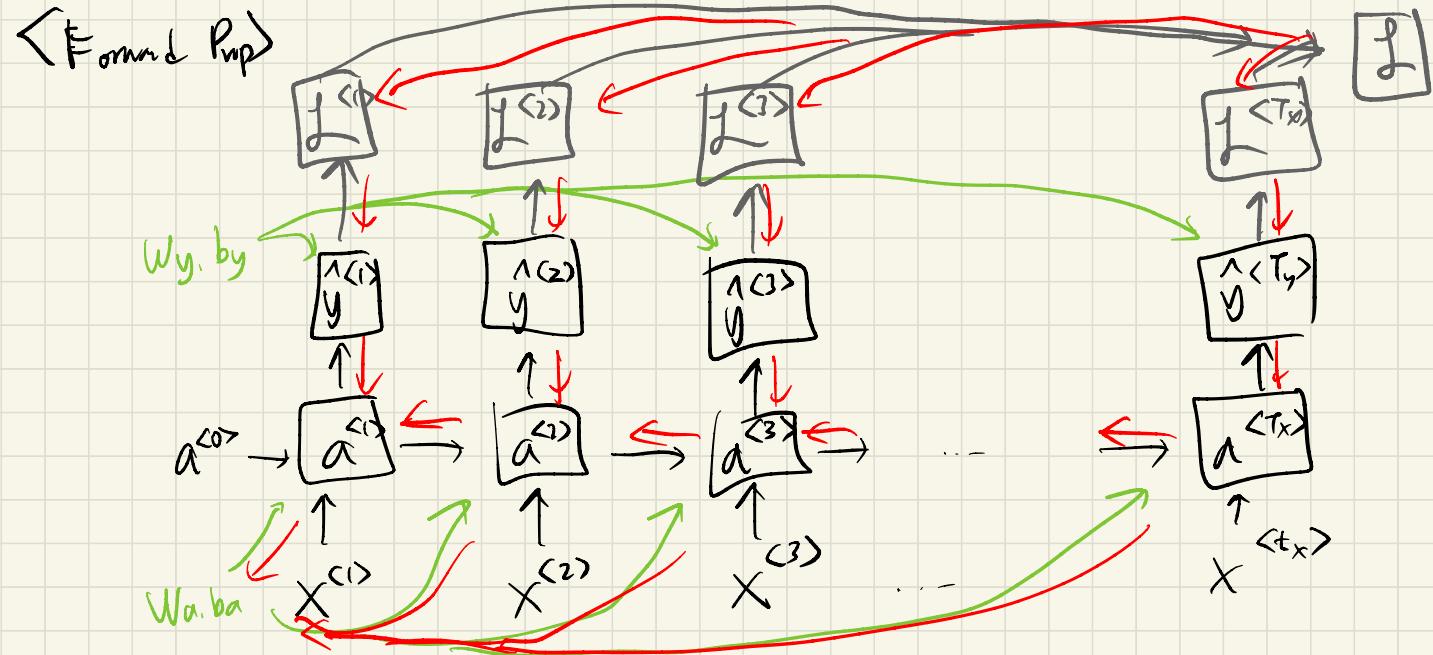
\longleftrightarrow

$$(100, 10000)$$

$$[a^{(t-1)}, x^{(t)}] = \begin{bmatrix} a^{(t-1)} \\ x^{(t)} \end{bmatrix}$$

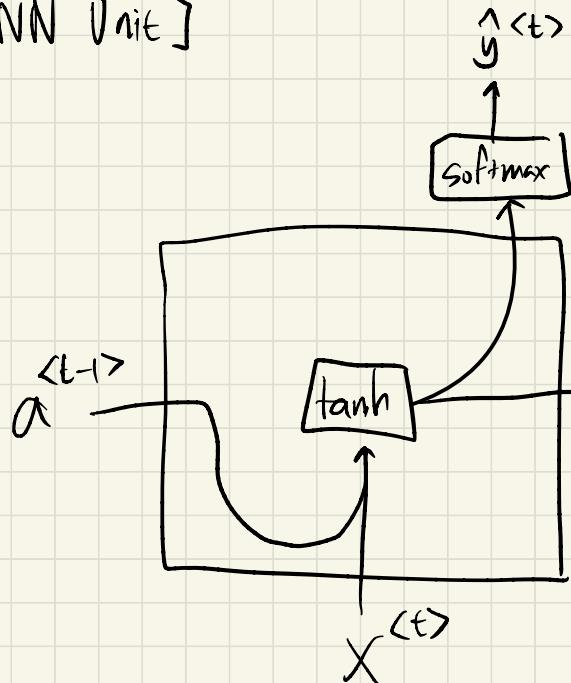
$\uparrow 100$
 $\downarrow 10,000$
 $\uparrow 10100$
 $\downarrow 10100$

$$[W_{aa} | W_{ax}] \begin{bmatrix} a^{(t-1)} \\ x^{(t)} \end{bmatrix} = W_{aa} a^{(t-1)} + W_{ax} x^{(t)}$$



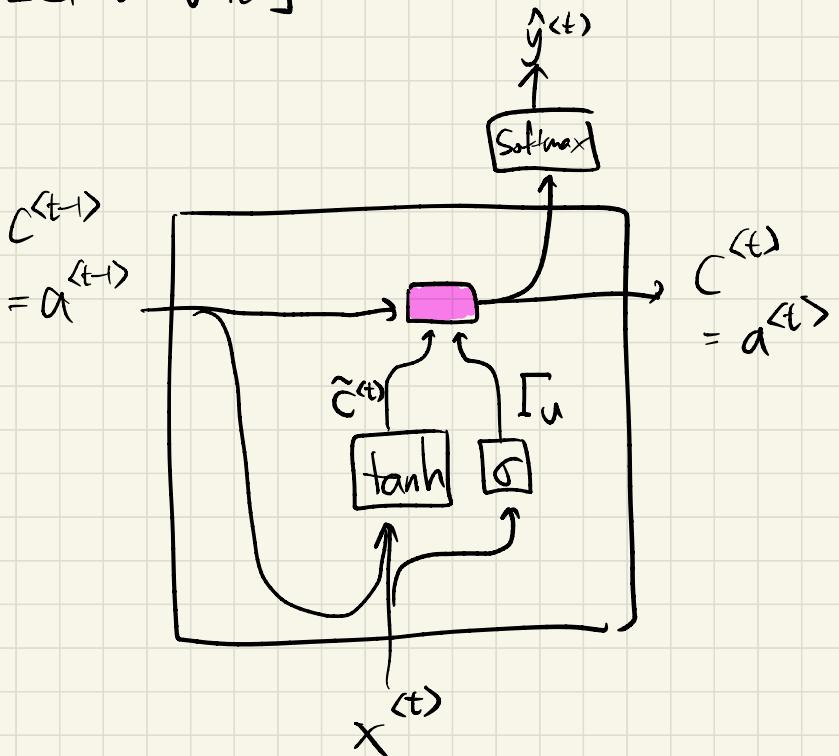
$$J^{(t)} (y^{(t)}, \hat{y}^{(t)}) = -y^{(t)} \log \hat{y}^{(t)} - (1-y^{(t)}) \log (1-\hat{y}^{(t)})$$

[RNN Unit]



$$a^{(t)} = \tanh(g(W_a[a^{(t-1)}, x^{(t)}] + b_a))$$

[GRU Unit]



$C = \text{memory cell}$

$$\rightarrow C^{(t)} = a^{(t)}$$

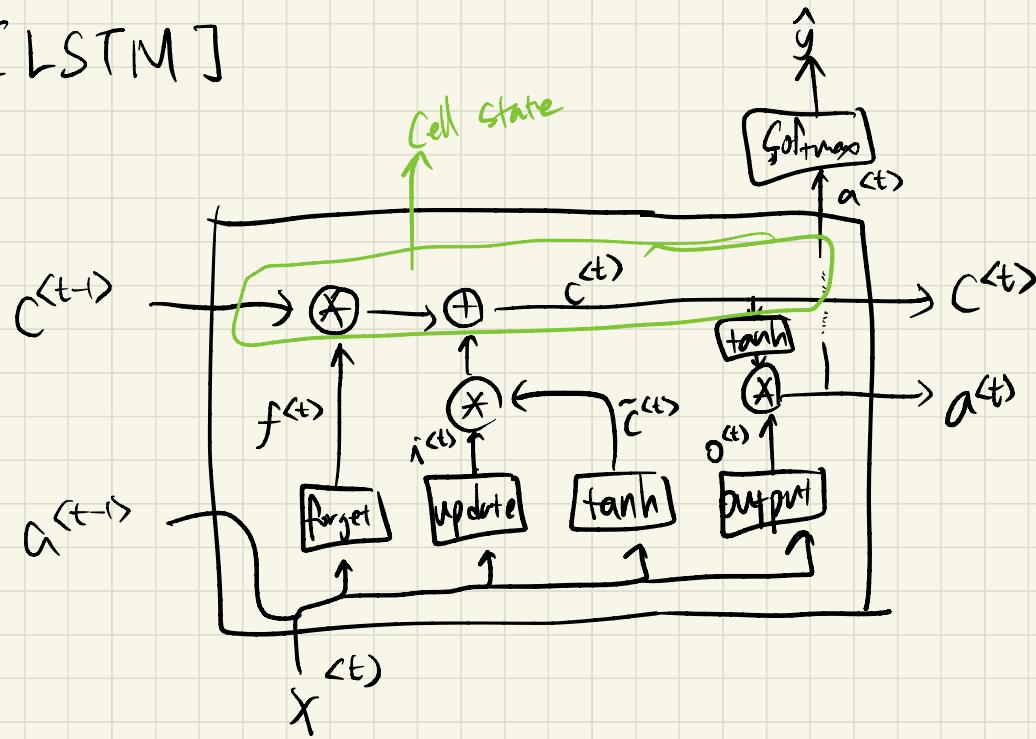
$$\rightarrow \tilde{C}^{(t)} = \tanh (\quad)$$

$$\rightarrow \Gamma_u = \sigma (\quad)$$

\uparrow
"update"

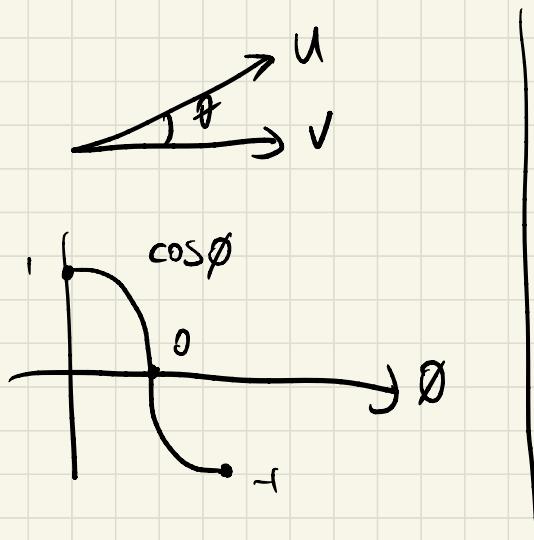
$$C^{(t)} = \Gamma_u * \tilde{C}^{(t)} + (1 - \Gamma_u) * \tilde{C}^{(t-1)}$$

[LSTM]

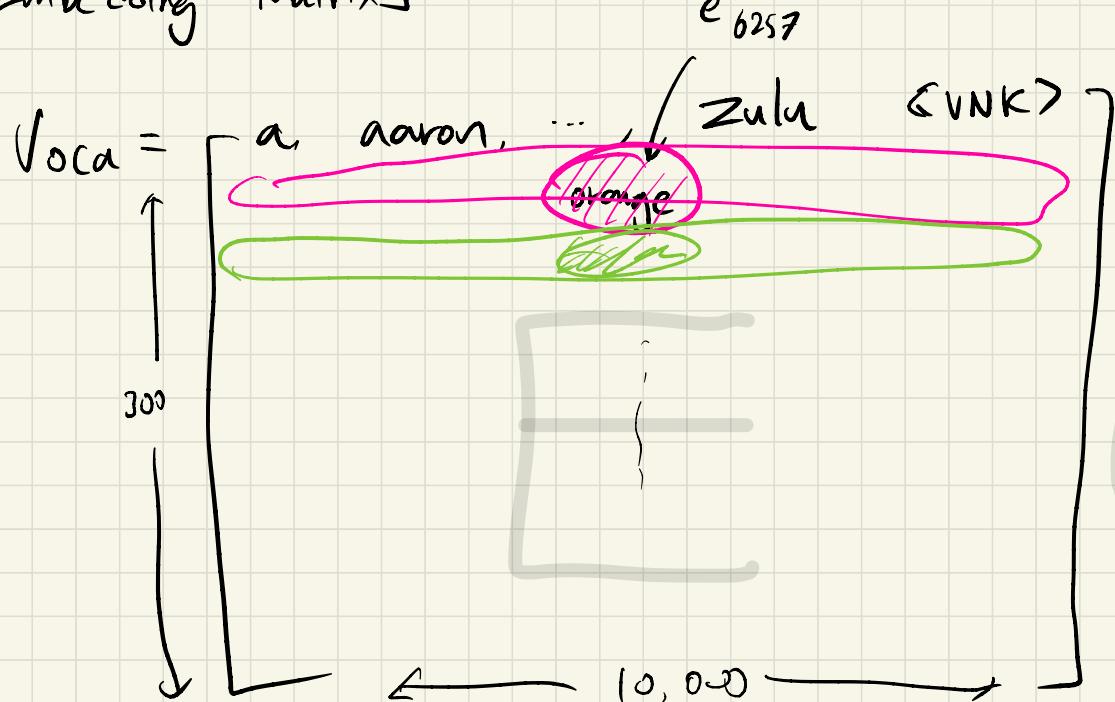


[Cosine Similarity]

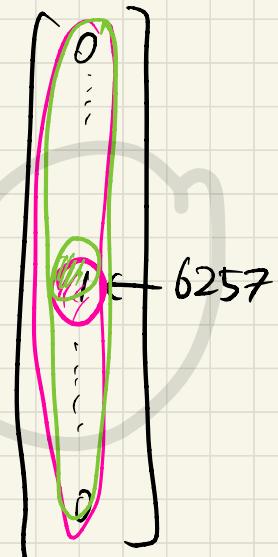
$$\text{sim}(u, v) = \frac{u^T v}{\|u\|_2 \cdot \|v\|_2}$$



[Embedding Matrix]

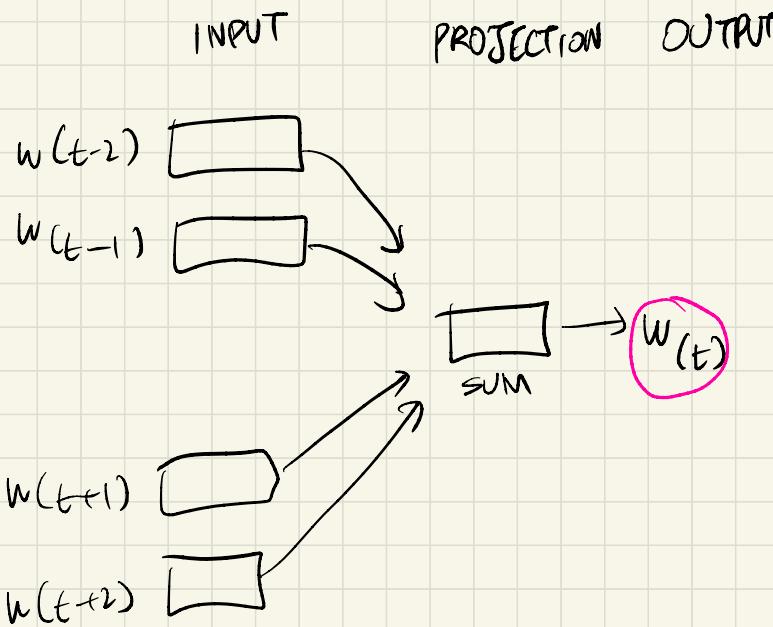


Orange (O_{6257})

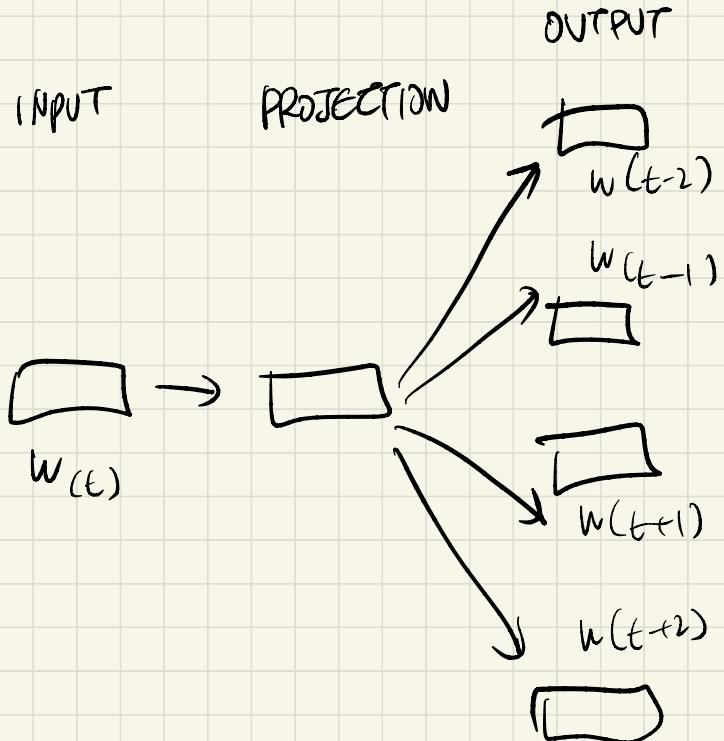


$$E \cdot O_{6257}^{(300, 1x)} \stackrel{(10x, 1)}{=} \begin{bmatrix} \vdots \\ \vdots \end{bmatrix}_{(300, 1)} = e_{6257}$$

[Word 2 Vec]



CBOW



Skip-Gram

[Attention]

RNN Attention

$$\alpha^{(t,t')} = \frac{\exp(e^{(t,t')})}{\sum_{t'=1}^{T_x} \exp(e^{(t,t')})}$$

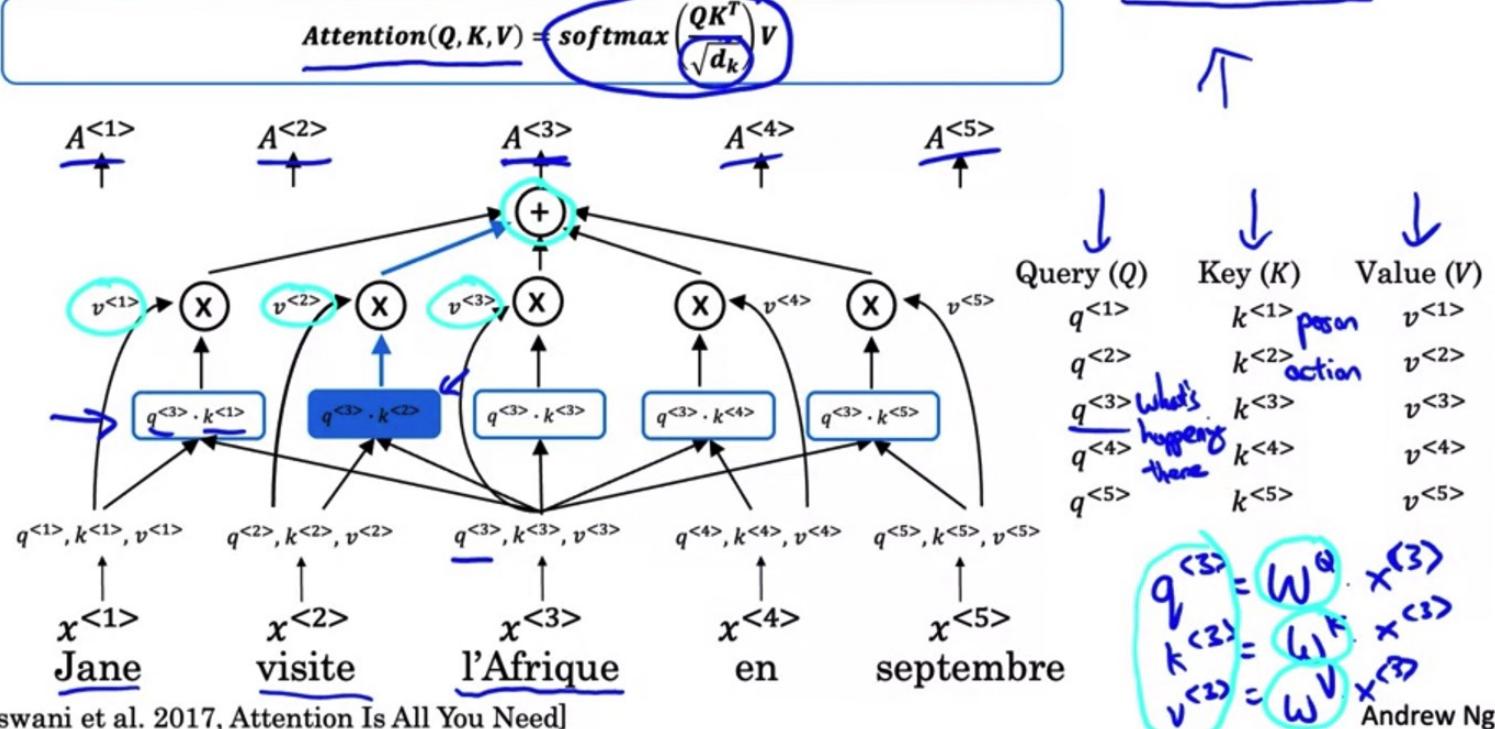
Transformer Attention

$$A(q, k, v) = \sum_i \frac{\exp(q \cdot k^{(i)})}{\sum_j \exp(q \cdot k^{(j)})} v^{(i)}$$

Self-Attention

$$A(q, K, V) = \sum_i \frac{\exp(q \cdot k^{<i>})}{\sum_j \exp(q \cdot k^{<j>})} v^{<i>}$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$



[Vaswani et al. 2017, Attention Is All You Need]

Andrew Ng