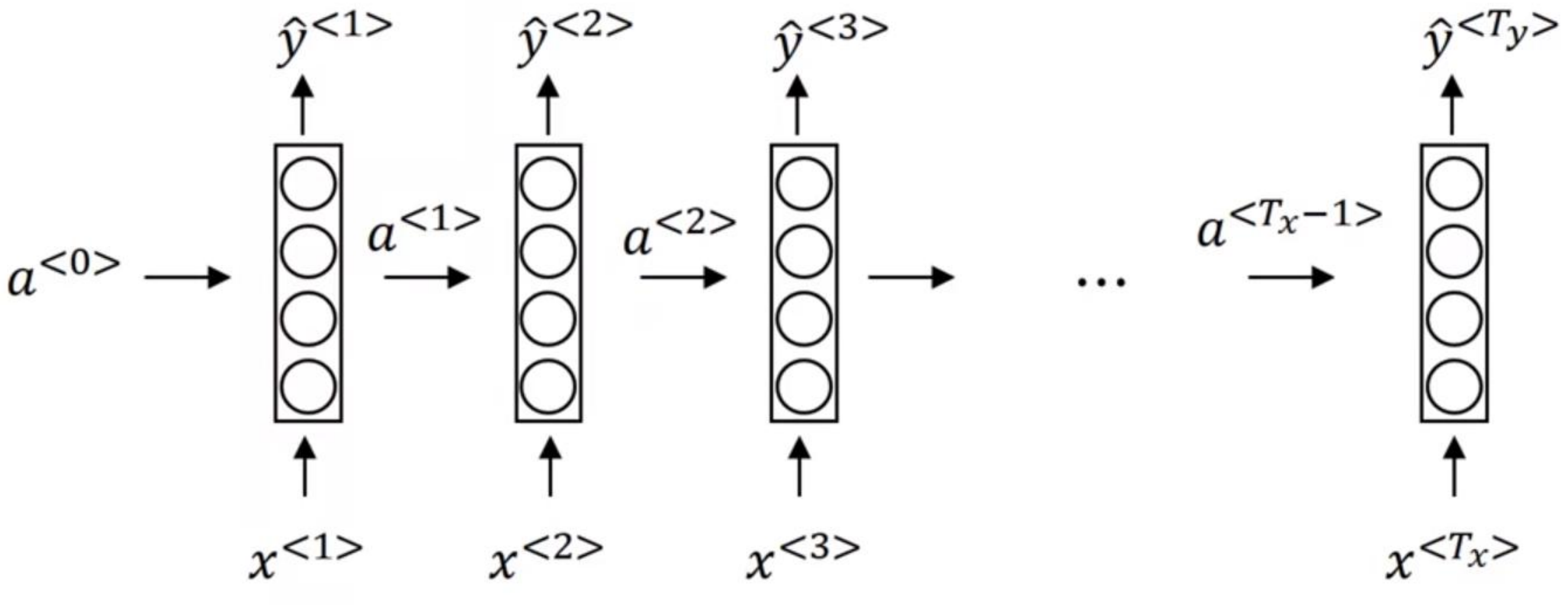
Attention

RNN

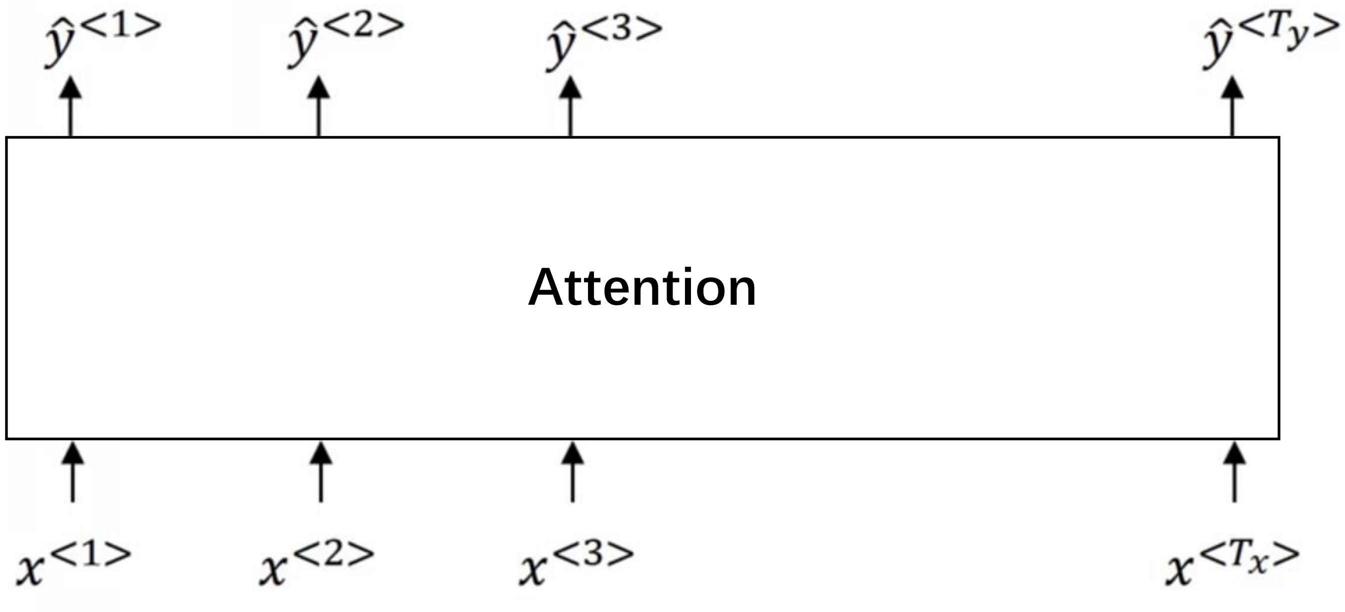
$$\boldsymbol{a}^{<t>} = \boldsymbol{g}(\boldsymbol{W}_{aa}\boldsymbol{a}^{<t-1>} + \boldsymbol{W}_{ax}\boldsymbol{x}^{<t>} + \boldsymbol{b}_a)$$

$$\hat{\boldsymbol{y}}^{<t>} = \boldsymbol{g}(\boldsymbol{W}_y\boldsymbol{a}^{<t>} + \boldsymbol{b}_y)$$



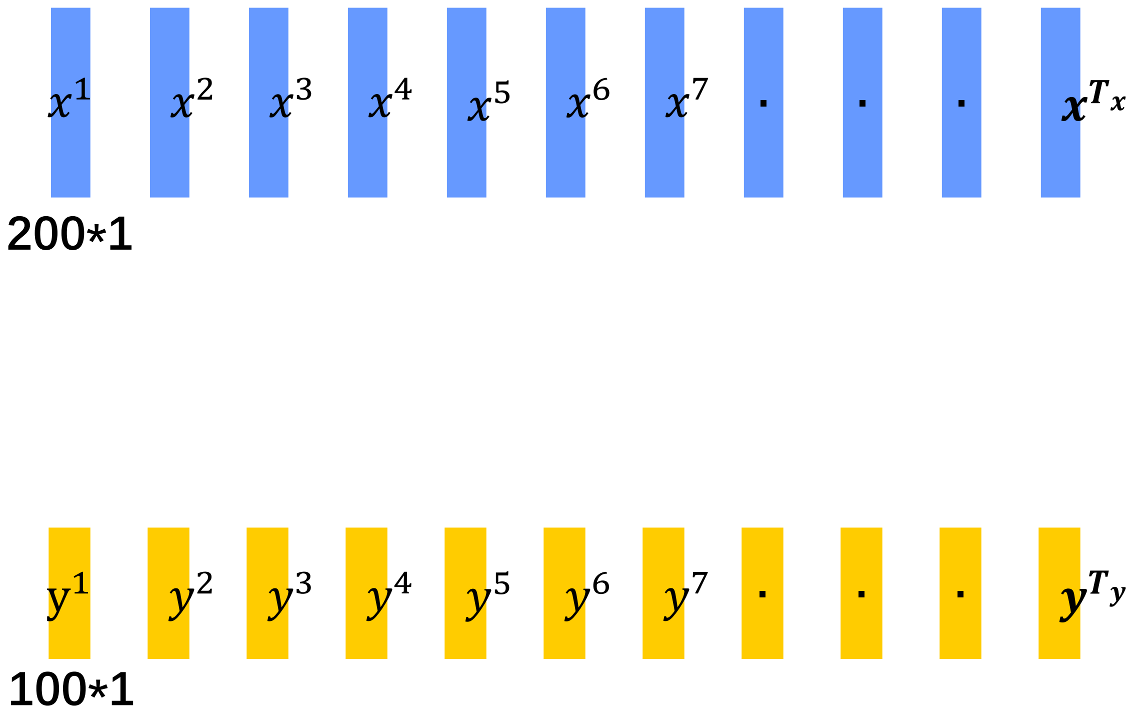
Attention

$$\boldsymbol{Q = W_q X}$$
$$\boldsymbol{K = W_k X}$$
$$\boldsymbol{V = W_v X}$$
$$\boldsymbol{A = K^T Q}$$
$$\boldsymbol{A' = softmax(A)}$$
$$\boldsymbol{Y = V A'}$$



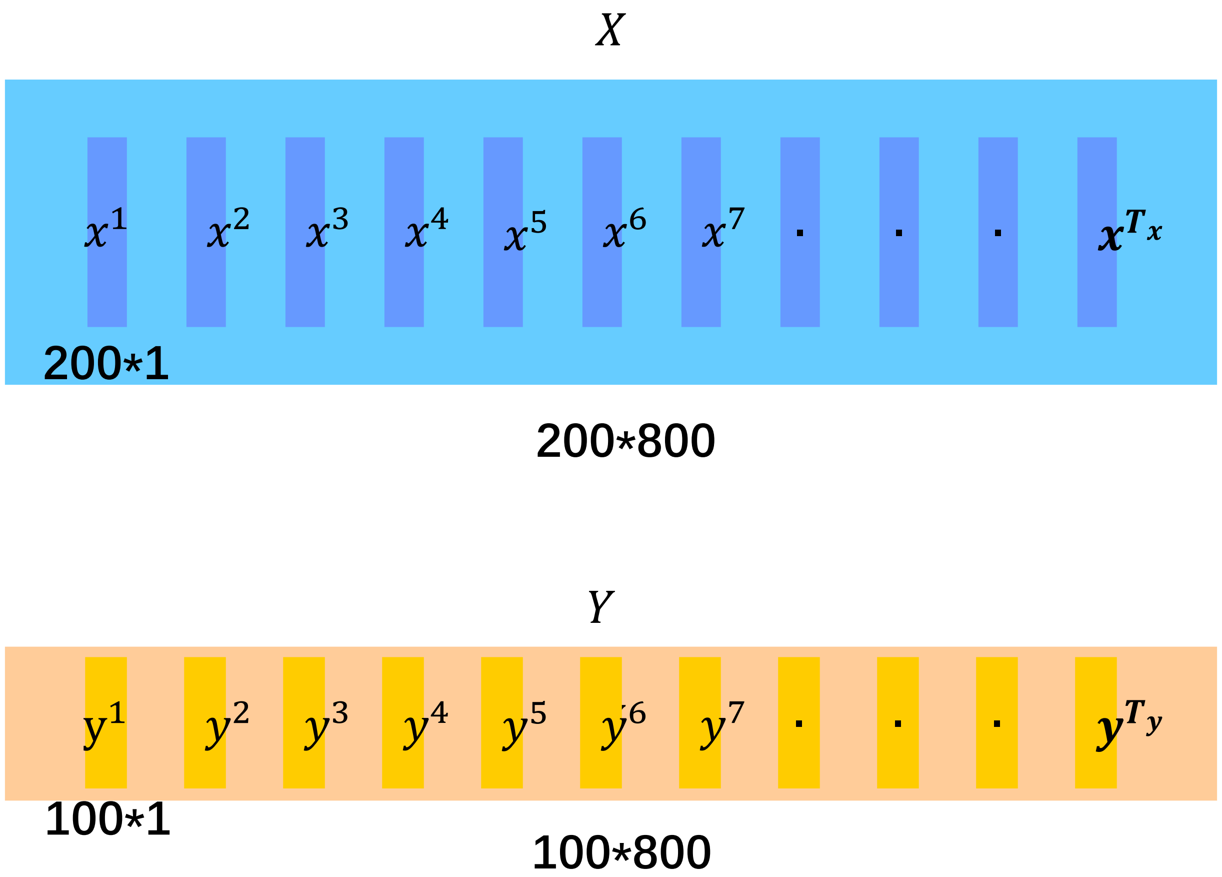
Attention

向量矩阵化



Attention

向量矩阵化



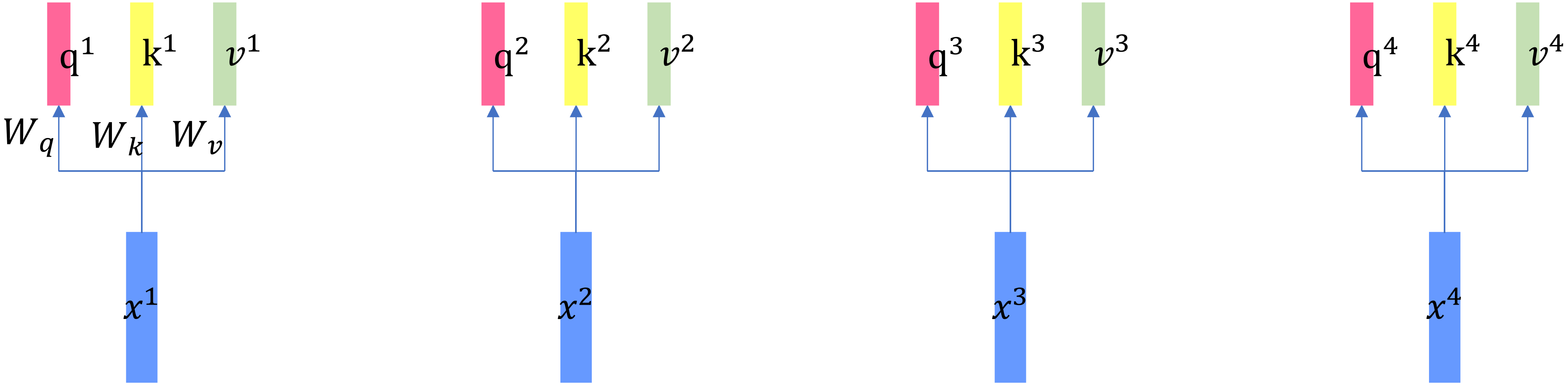
Attention

$$\boldsymbol{Q = W_q X}$$
$$\boldsymbol{K = W_k X}$$
$$\boldsymbol{V = W_v X}$$
$$\boldsymbol{A = K^T Q}$$
$$\boldsymbol{A' = softmax(A)}$$
$$\boldsymbol{Y = VA'}$$

Attention

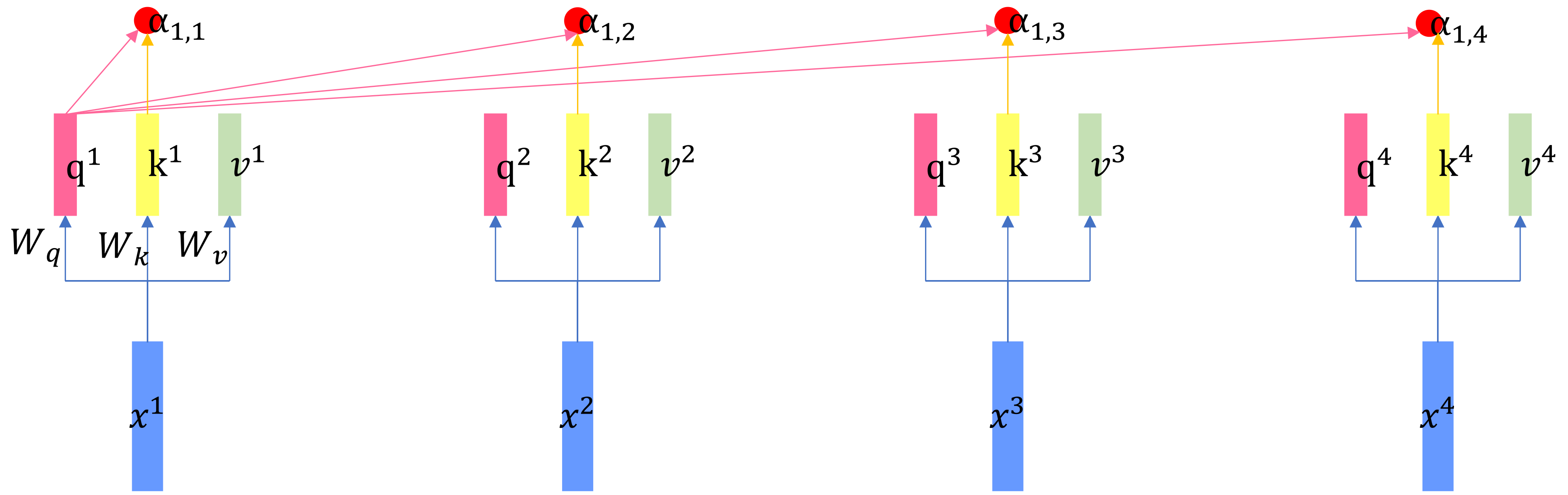
Attention

$$\boldsymbol{q^i = W_q x^i}$$
$$\boldsymbol{k^i = W_k x^i}$$
$$\boldsymbol{v^i = W_v x^i}$$



Attention

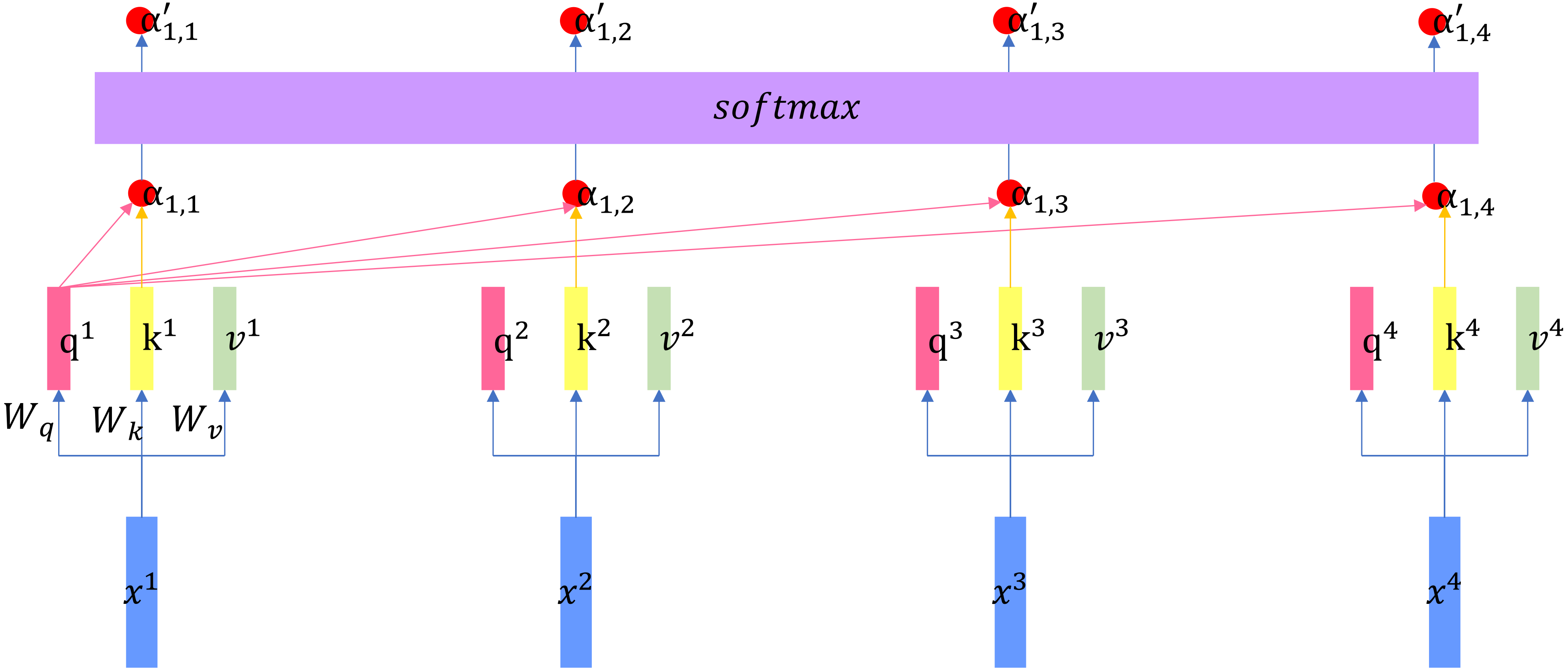
$$\boldsymbol{\alpha_{i,j} = q^i \cdot k^j}$$



Attention

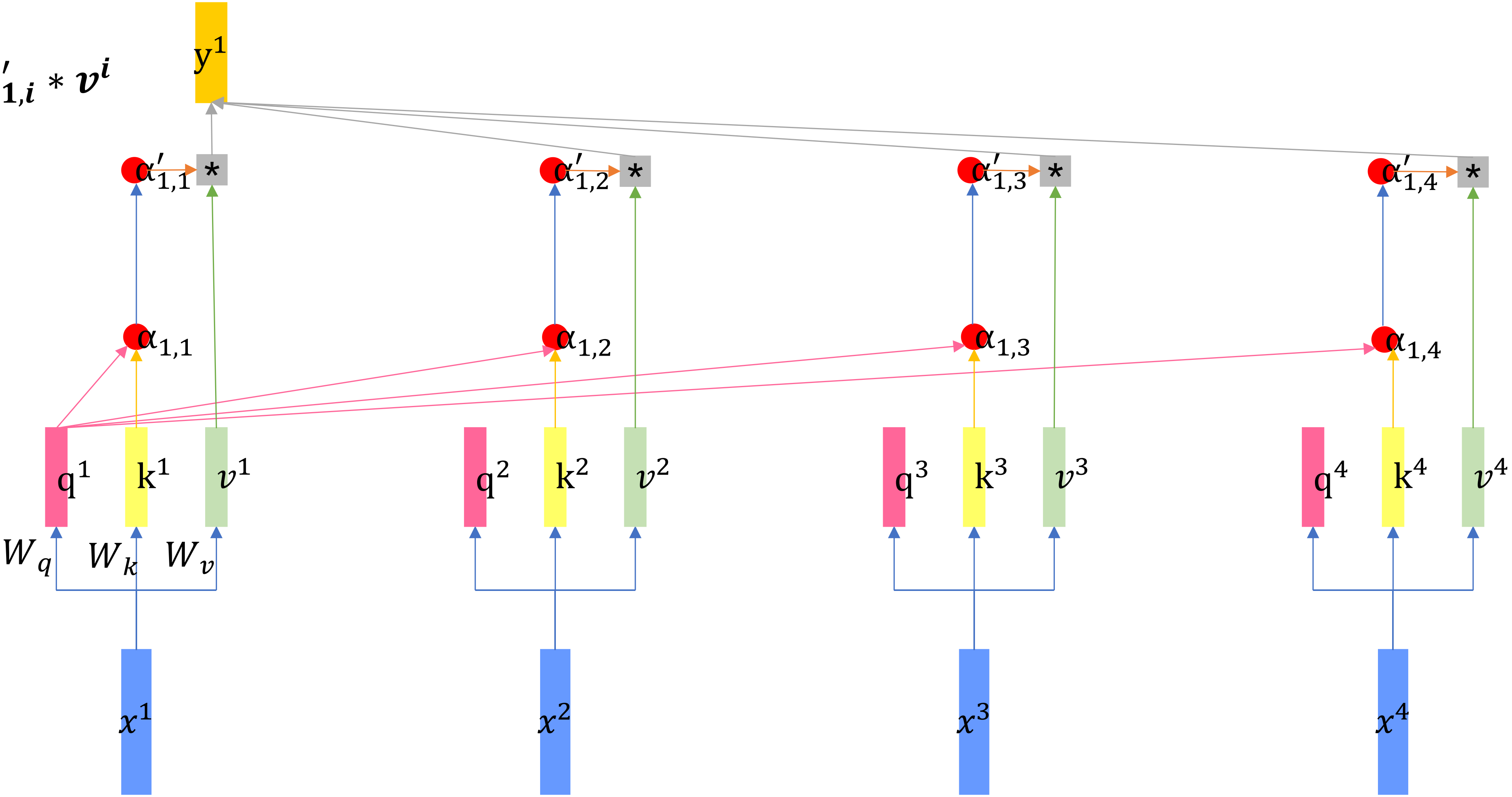
$$\boldsymbol{\alpha'_{1,i} = \frac{e^{\alpha_{1,i}}}{\sum_j e^{\alpha_{1,j}}}}$$

$$\boldsymbol{\alpha'_{1,2} = \frac{e^{\alpha_{1,2}}}{e^{\alpha_{1,1}} + e^{\alpha_{1,2}} + e^{\alpha_{1,3}} + e^{\alpha_{1,4}}}}$$



Attention

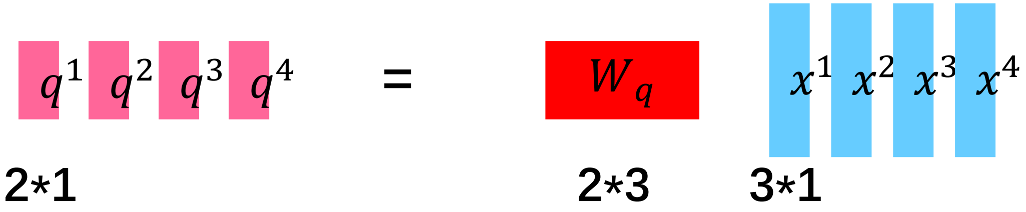
$$\boldsymbol{y^1 = \sum_i \alpha'_{1,i} * v^i}$$

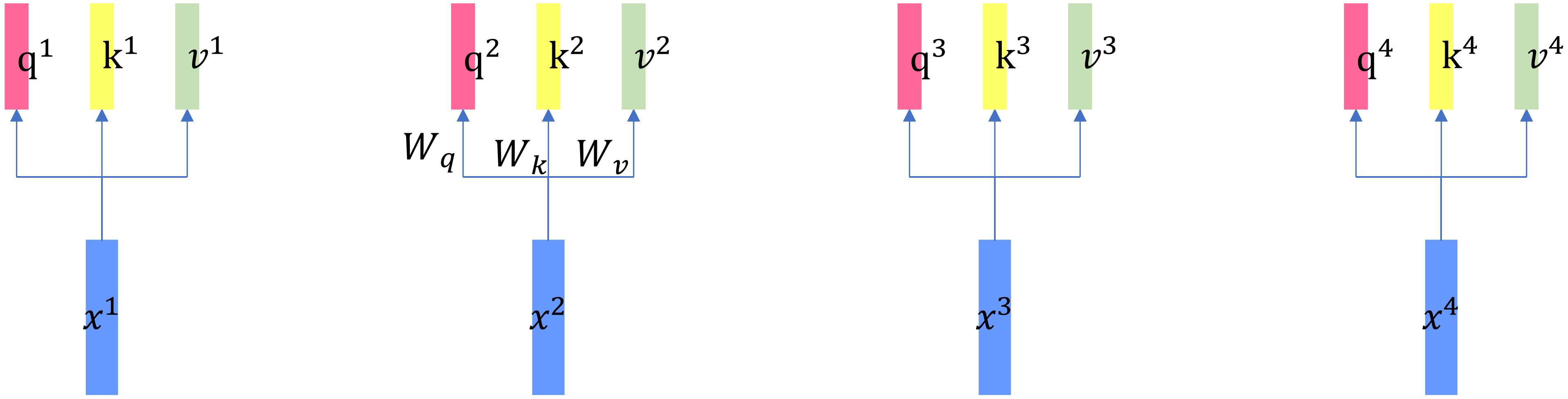


Attention

Attention

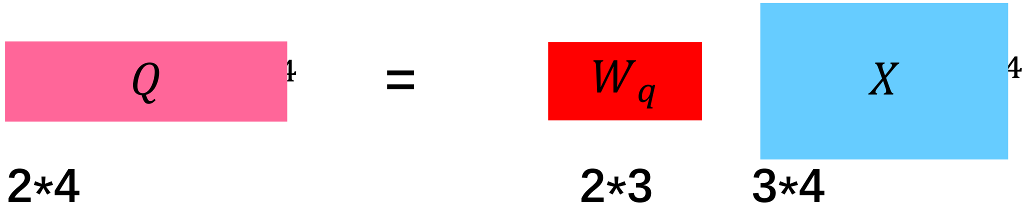
$$\boldsymbol{q}^i = \boldsymbol{W}_q \boldsymbol{x}^i$$
$$\boldsymbol{k}^i = \boldsymbol{W}_k \boldsymbol{x}^i$$
$$\boldsymbol{v}^i = \boldsymbol{W}_v \boldsymbol{x}^i$$




q^1 k^1 v^1
q^2 k^2 v^2
q^3 k^3 v^3
q^4 k^4 v^4
W_q W_k W_v
x^1
x^2
x^3
x^4

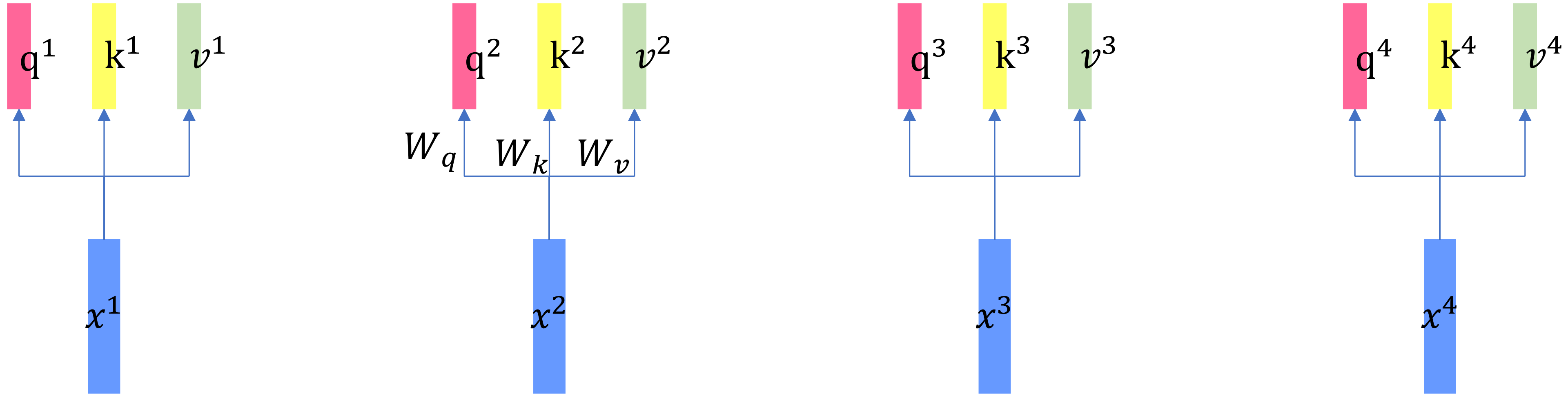
Attention

$$\boldsymbol{q^i = W_q x^i}$$
$$\boldsymbol{k^i = W_k x^i}$$
$$\boldsymbol{v^i = W_v x^i}$$



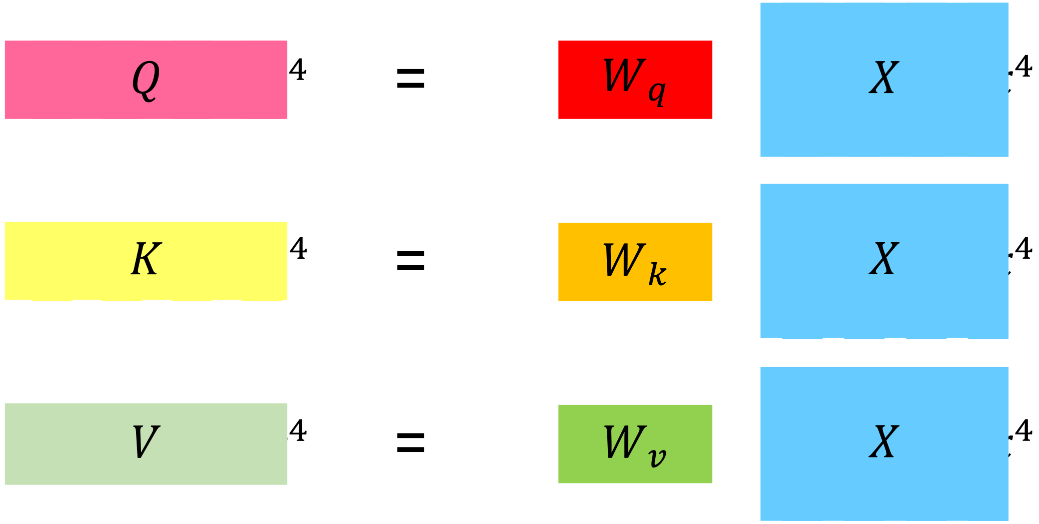
$$\boldsymbol{Q = W_q X}$$

$$\begin{pmatrix}1\\2\end{pmatrix}\begin{pmatrix}3\\1\end{pmatrix}\begin{pmatrix}3\\0\end{pmatrix}\begin{pmatrix}2\\6\end{pmatrix} = \begin{pmatrix}1&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}1\\0\\2\end{pmatrix}\begin{pmatrix}3\\2\\1\end{pmatrix}\begin{pmatrix}3\\0\\0\end{pmatrix}\begin{pmatrix}2\\9\\6\end{pmatrix}$$

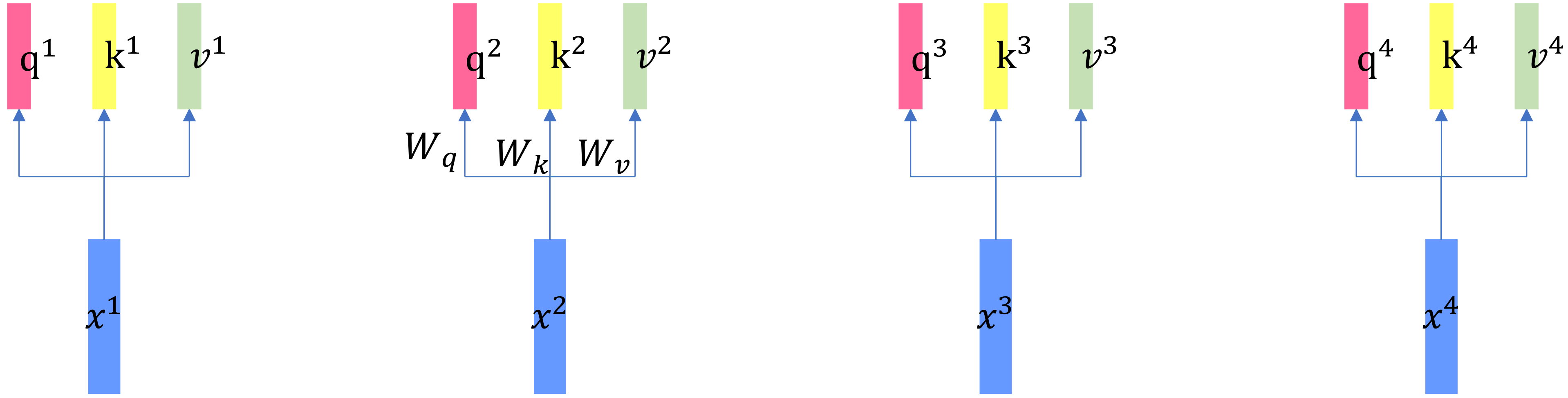


Attention

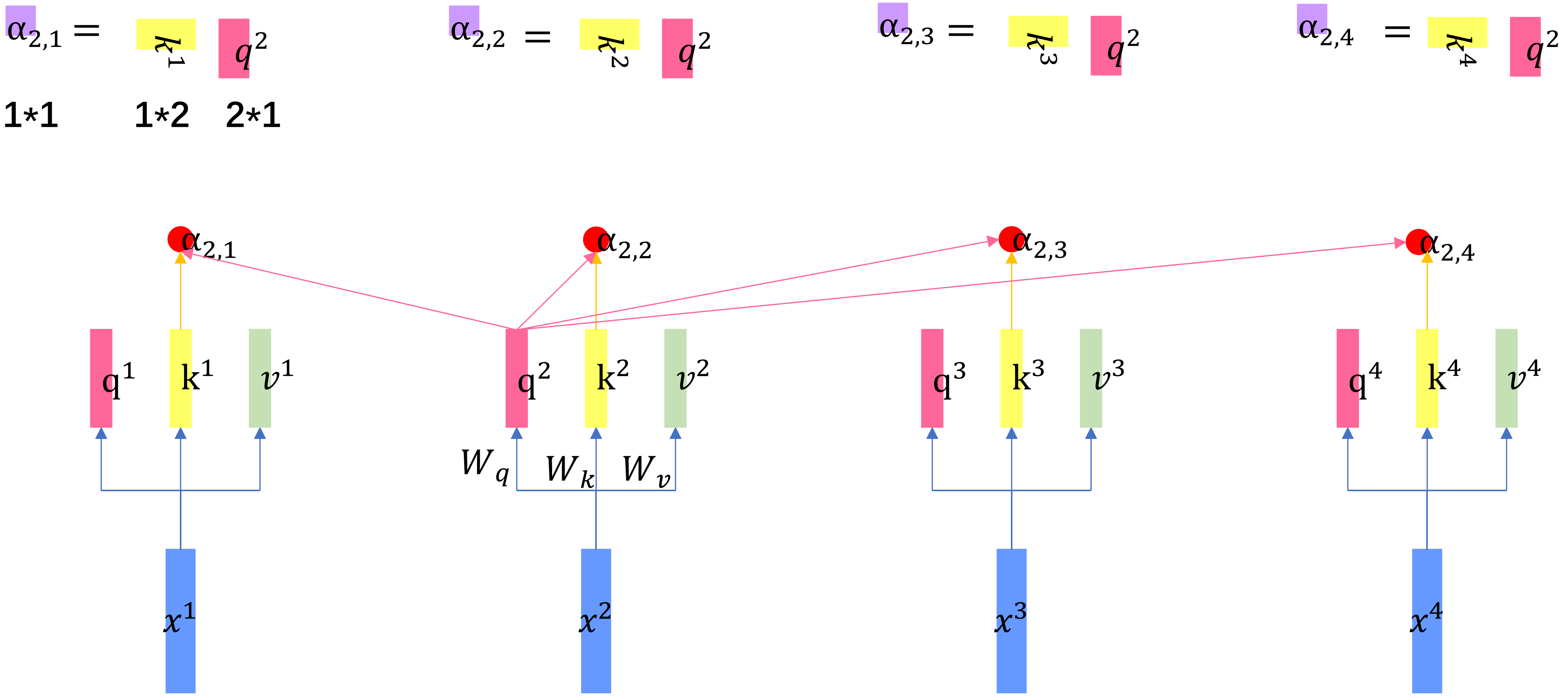
$$\boldsymbol{q^i = W_q x^i}$$
$$\boldsymbol{k^i = W_k x^i}$$
$$\boldsymbol{v^i = W_v x^i}$$



$$\boldsymbol{Q = W_q X}$$
$$\boldsymbol{K = W_k X}$$
$$\boldsymbol{V = W_v X}$$



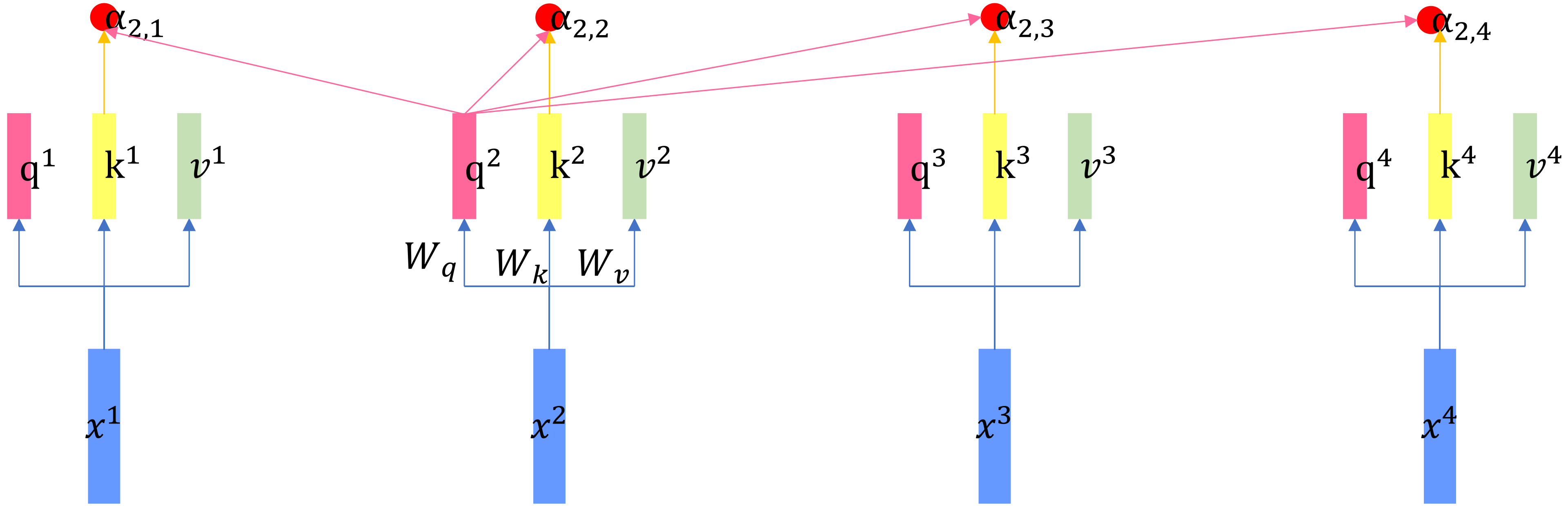
Attention



$\alpha_{2,1}$ = k^1 q^2
$\alpha_{2,2}$ = k^2 q^2
$\alpha_{2,3}$ = k^3 q^2
$\alpha_{2,4}$ = k^4 q^2
1*1
1*2
2*1
$\alpha_{2,1}$
$\alpha_{2,2}$
$\alpha_{2,3}$
$\alpha_{2,4}$
q^1 k^1 v^1
q^2 k^2 v^2
q^3 k^3 v^3
q^4 k^4 v^4
W_q W_k W_v
x^1
x^2
x^3
x^4

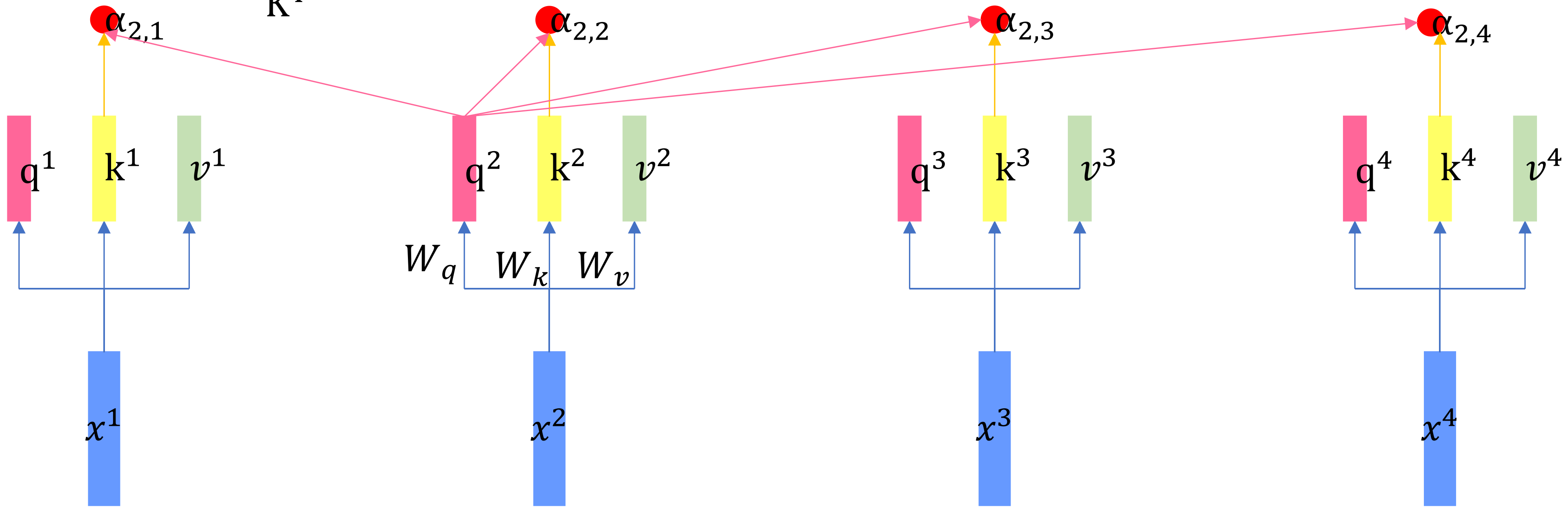
Attention

$$\begin{matrix}\alpha_{2,1}\\ \alpha_{2,2}\\ \alpha_{2,3}\\ \alpha_{2,4}\\ \mathbf{4*1}\end{matrix} = \begin{matrix}k^1\\ k^2\\ k^3\\ k^4\\ \mathbf{4*2}\end{matrix} \quad \begin{matrix}q^2\\ \mathbf{2*1}\end{matrix}$$

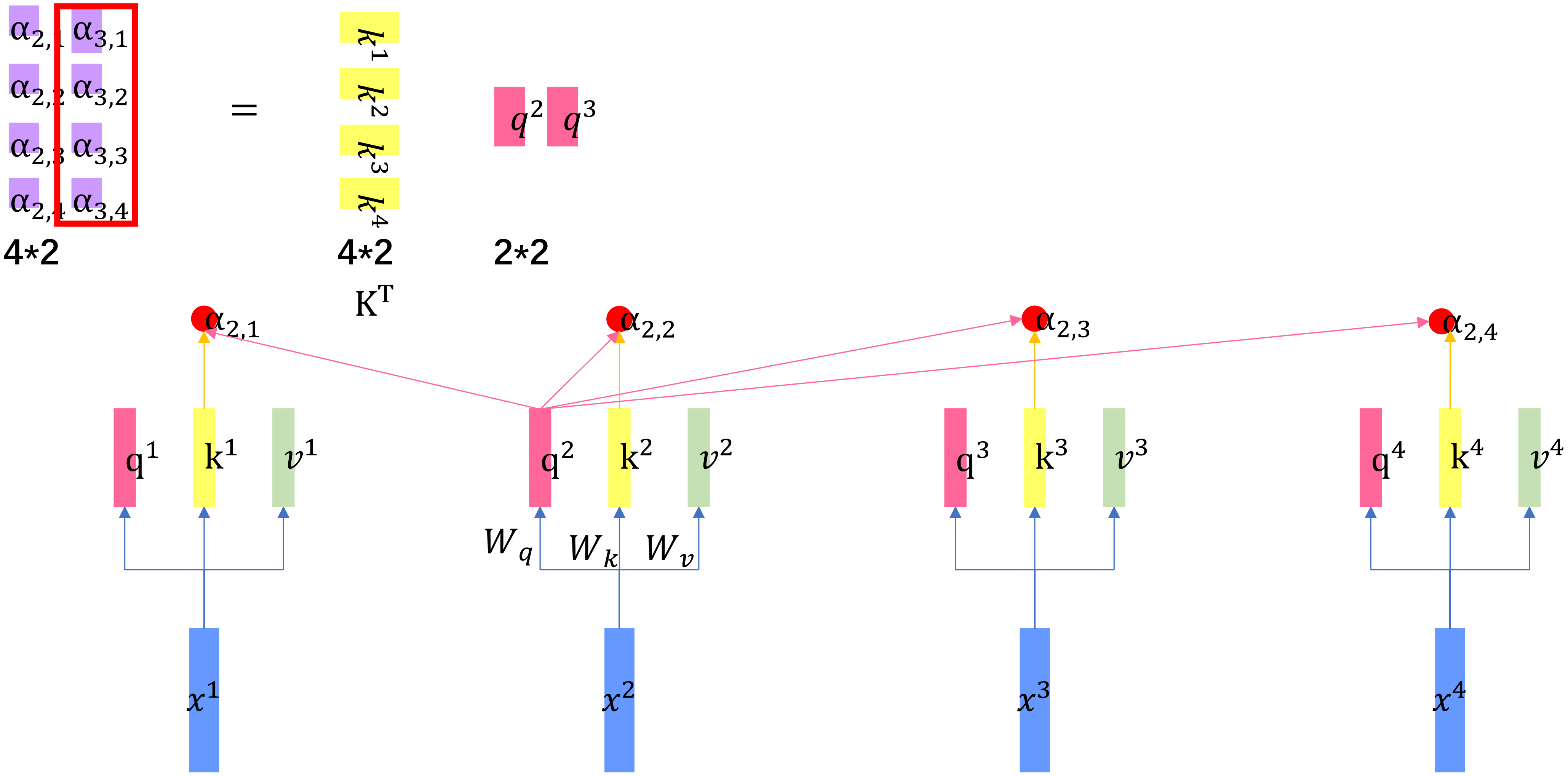


Attention

$$\begin{array}{c}\alpha_{2,1}\\ \alpha_{2,2}\\ \alpha_{2,3}\\ \alpha_{2,4}\\ \mathbf{4*1}\end{array} = \begin{array}{c}k^1\\ k^2\\ k^3\\ k^4\\ \mathbf{4*2}\\ \mathrm{K}^{\mathrm{T}}\end{array} \begin{array}{c}q^2\\ \\ \\ \\ \mathbf{2*1}\\ \\ \end{array}$$

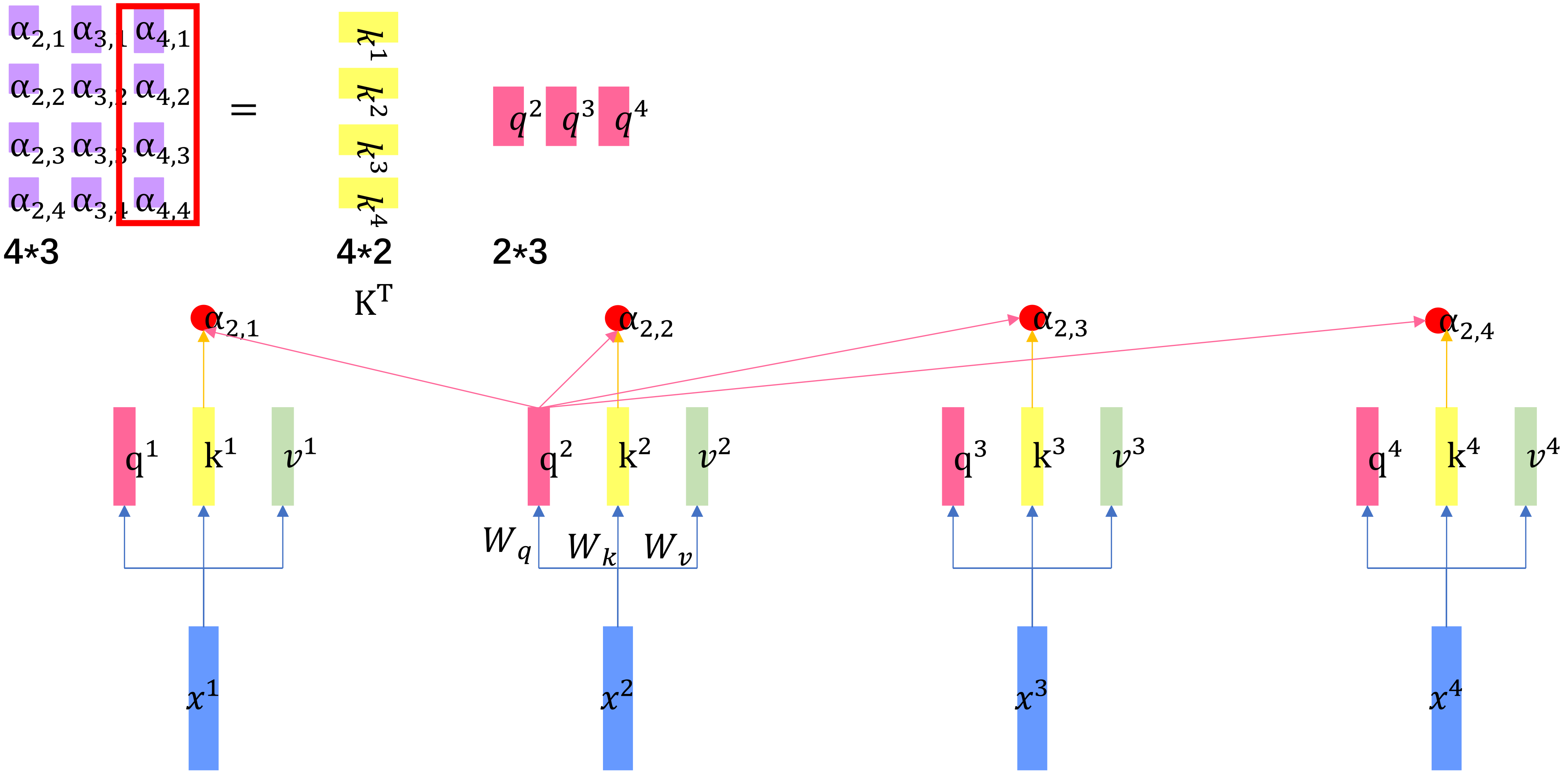


Attention



$\alpha_{2,1}$ $\alpha_{3,1}$
$\alpha_{2,2}$ $\alpha_{3,2}$
$\alpha_{2,3}$ $\alpha_{3,3}$
$\alpha_{2,4}$ $\alpha_{3,4}$
4*2
=
k^1 k^2 k^3 k^4
4*2
K^T
q^2 q^3
2*2
$\alpha_{2,1}$
$\alpha_{2,2}$
$\alpha_{2,3}$
$\alpha_{2,4}$
q^1 k^1 v^1
q^2 k^2 v^2
q^3 k^3 v^3
q^4 k^4 v^4
W_q W_k W_v
x^1
x^2
x^3
x^4

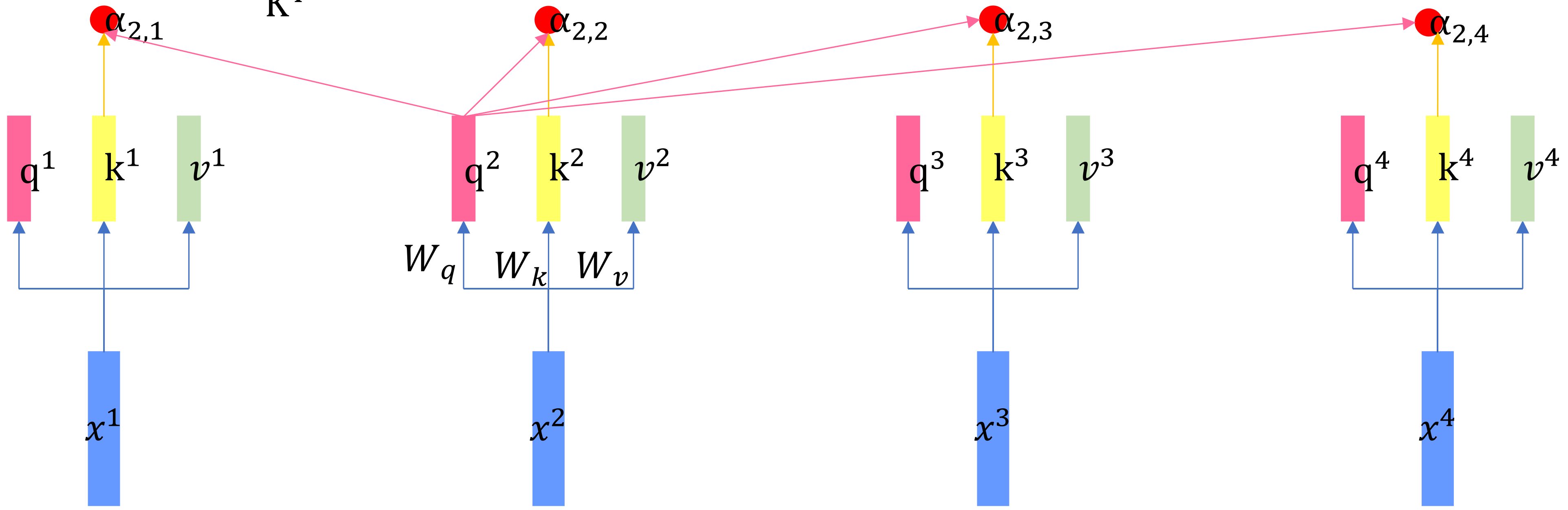
Attention



$\alpha_{2,1}$ $\alpha_{3,1}$ $\alpha_{4,1}$
$\alpha_{2,2}$ $\alpha_{3,2}$ $\alpha_{4,2}$
$\alpha_{2,3}$ $\alpha_{3,3}$ $\alpha_{4,3}$
$\alpha_{2,4}$ $\alpha_{3,4}$ $\alpha_{4,4}$
=
k^1 k^2 k^3 k^4
q^2 q^3 q^4
4*3
4*2
2*3
K^T
$\alpha_{2,1}$
$\alpha_{2,2}$
$\alpha_{2,3}$
$\alpha_{2,4}$
q^1 k^1 v^1
q^2 k^2 v^2
q^3 k^3 v^3
q^4 k^4 v^4
W_q W_k W_v
x^1
x^2
x^3
x^4

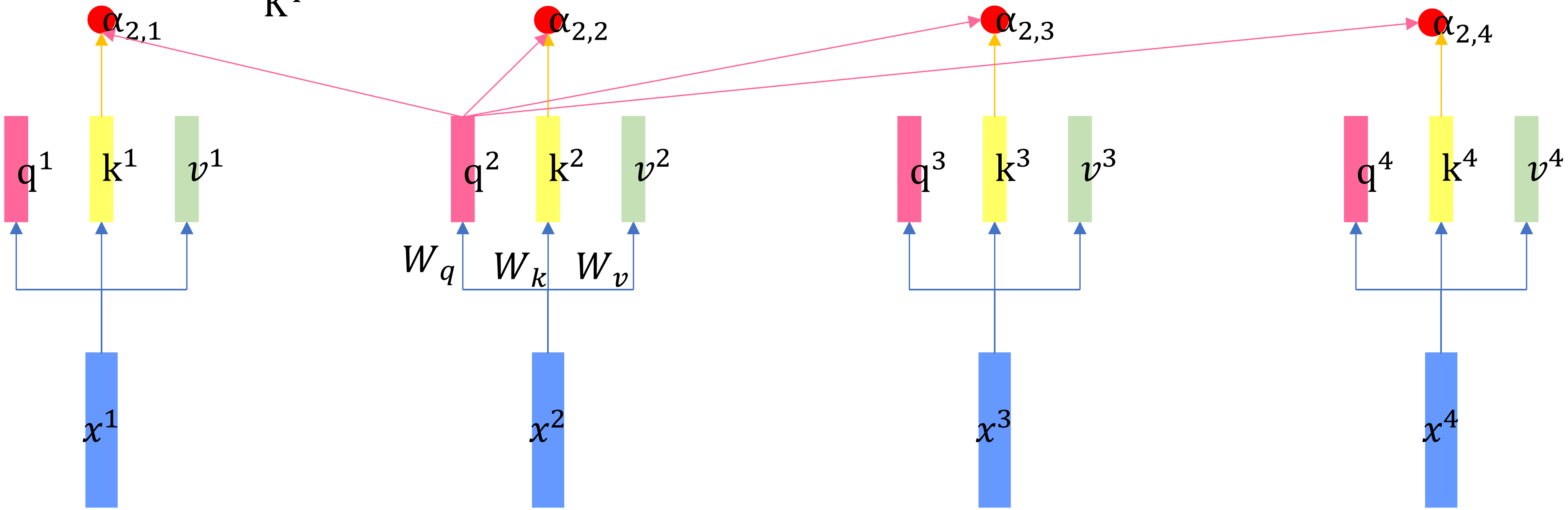
Attention

$$\underset{4*4}{\begin{bmatrix} \alpha_{1,1} & \alpha_{2,1} & \alpha_{3,1} & \alpha_{4,1} \\ \alpha_{1,2} & \alpha_{2,2} & \alpha_{3,2} & \alpha_{4,2} \\ \alpha_{1,3} & \alpha_{2,3} & \alpha_{3,3} & \alpha_{4,3} \\ \alpha_{1,4} & \alpha_{2,4} & \alpha_{3,4} & \alpha_{4,4} \end{bmatrix}} = \underset{\substack{4*2 \\ \mathrm{K}^{\mathrm{T}}}}{\begin{bmatrix} k^1 \\ k^2 \\ k^3 \\ k^4 \end{bmatrix}} \underset{2*4}{\begin{bmatrix} q^1 & q^2 & q^3 & q^4 \end{bmatrix}}$$

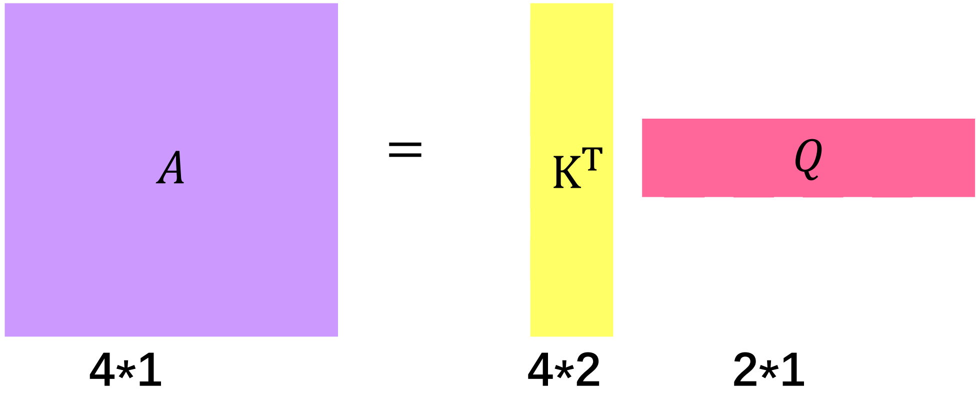


Attention

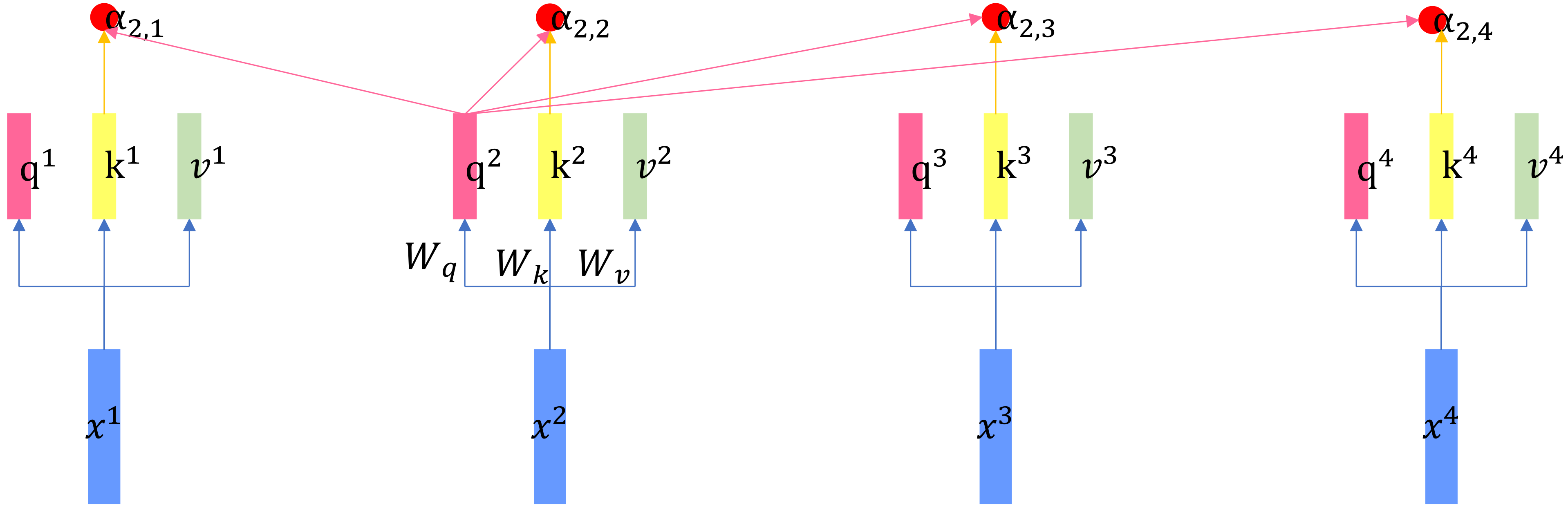
$$\underset{\textbf{4*4}}{\begin{matrix} \alpha_{1,1} & \alpha_{2,1} & \alpha_{3,1} & \alpha_{4,1} \\ \alpha_{1,2} & \alpha_{2,2} & \alpha_{3,2} & \alpha_{4,2} \\ \alpha_{1,3} & \alpha_{2,3} & \alpha_{3,3} & \alpha_{4,3} \\ \alpha_{1,4} & \alpha_{2,4} & \alpha_{3,4} & \alpha_{4,4} \end{matrix}} = \underset{\substack{\textbf{4*2} \\ \mathrm{K}^{\mathrm{T}}}}{\begin{matrix} k^1 \\ k^2 \\ k^3 \\ k^4 \end{matrix}} \quad \underset{\textbf{2*4}}{\begin{matrix} q^1 & q^2 & q^3 & q^4 \end{matrix}}$$



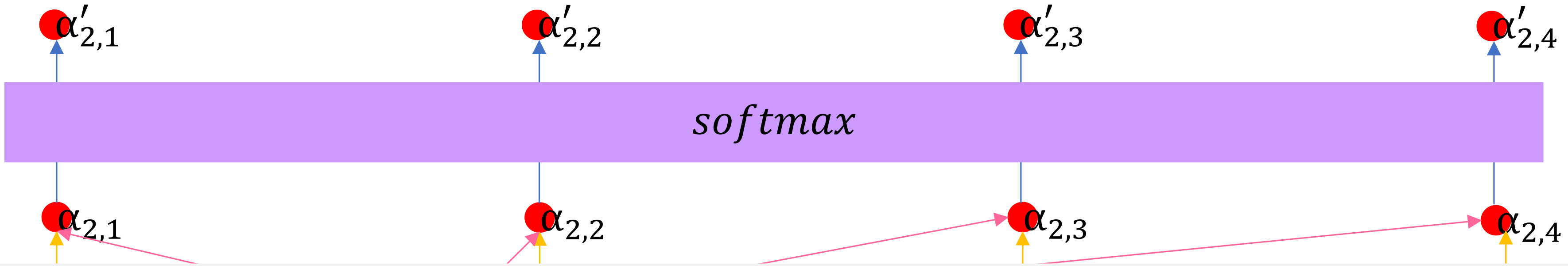
Attention



$$A = K^T Q$$

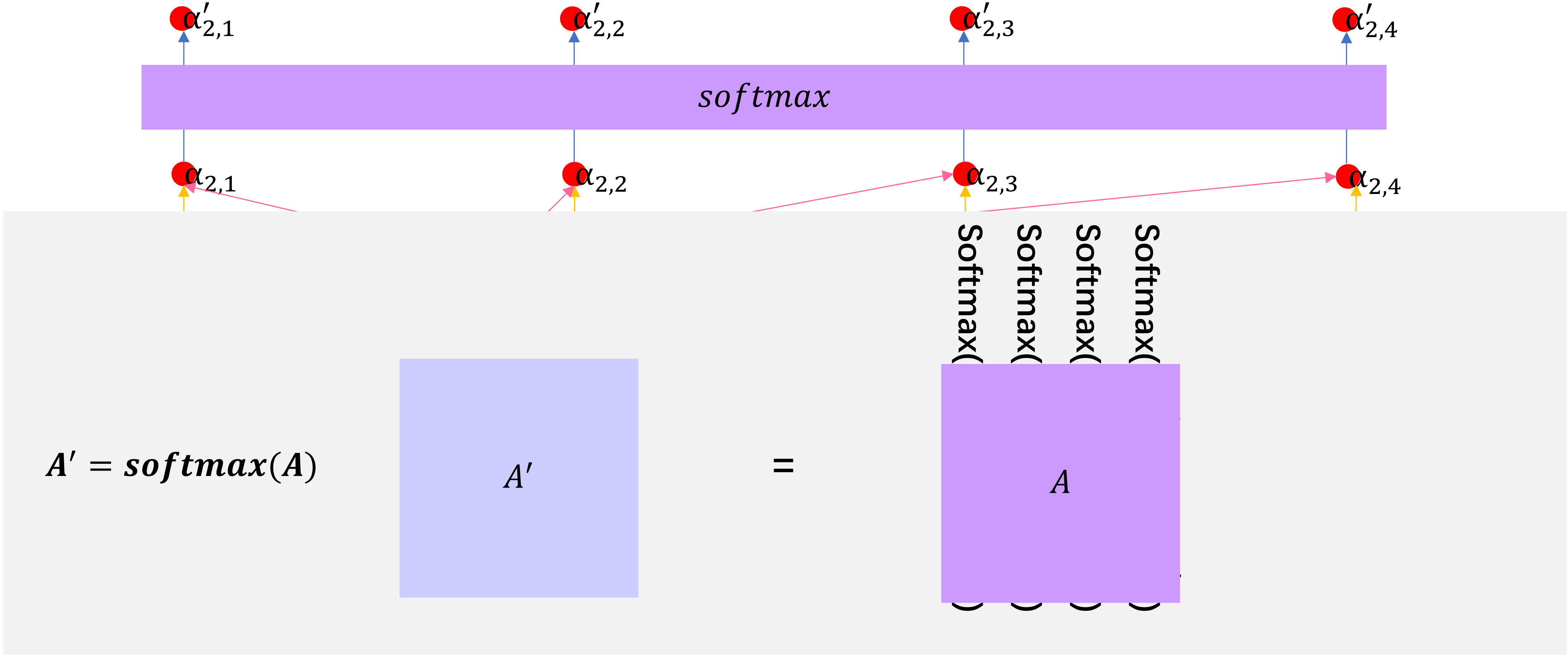


Attention



$$\begin{matrix} \alpha'_{1,1} & \alpha'_{2,1} & \alpha'_{3,1} & \alpha'_{4,1} \\ \alpha'_{1,2} & \alpha'_{2,2} & \alpha'_{3,2} & \alpha'_{4,2} \\ \alpha'_{1,3} & \alpha'_{2,3} & \alpha'_{3,3} & \alpha'_{4,3} \\ \alpha'_{1,4} & \alpha'_{2,4} & \alpha'_{3,4} & \alpha'_{4,4} \end{matrix} = \begin{matrix} \mathrm{Softmax}(& \mathrm{Softmax}(& \mathrm{Softmax}(& \mathrm{Softmax}(\\ \alpha_{1,1} & \alpha_{2,1} & \alpha_{3,1} & \alpha_{4,1} \\ \alpha_{1,2} & \alpha_{2,2} & \alpha_{3,2} & \alpha_{4,2} \\ \alpha_{1,3} & \alpha_{2,3} & \alpha_{3,3} & \alpha_{4,3} \\ \alpha_{1,4} & \alpha_{2,4} & \alpha_{3,4} & \alpha_{4,4} \\ ) &) &) &) \end{matrix}$$

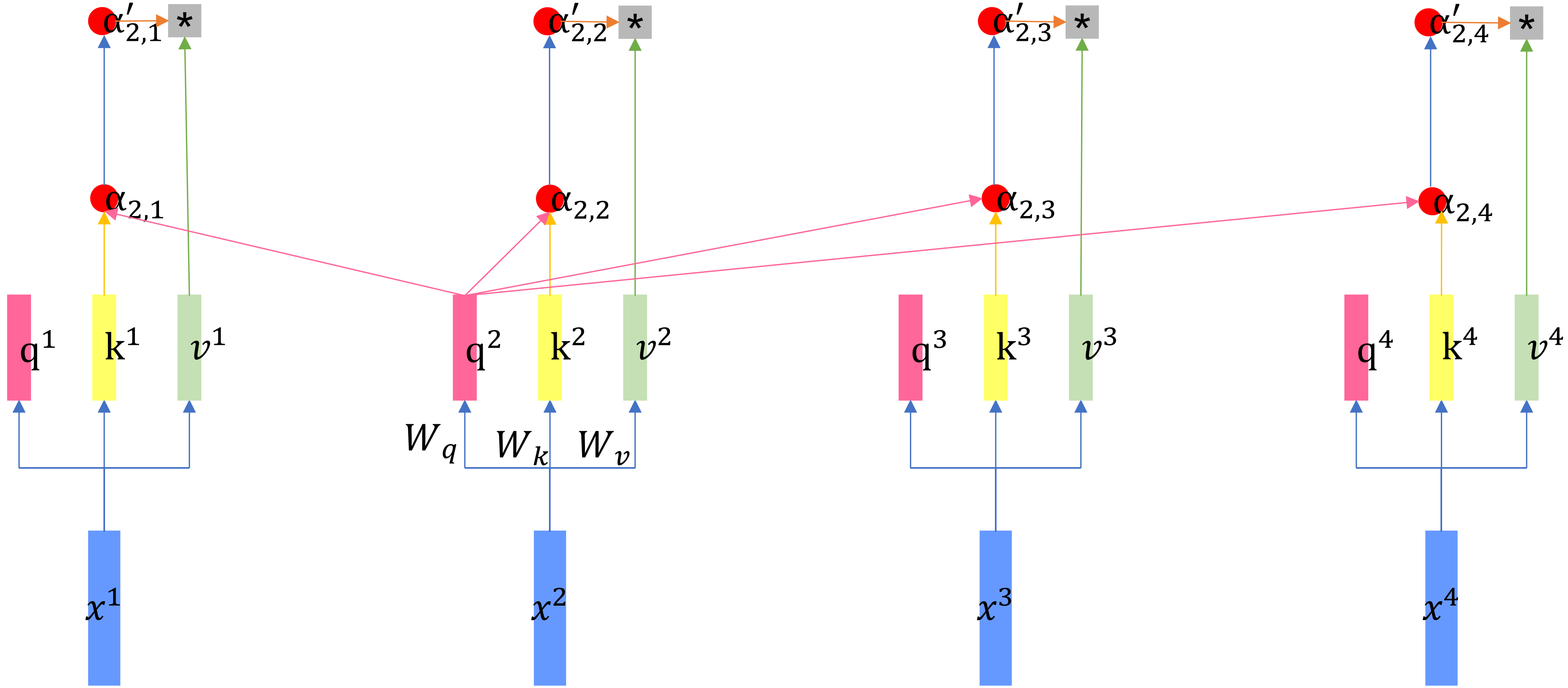
Attention



$$\boldsymbol{A' = softmax(A)}$$

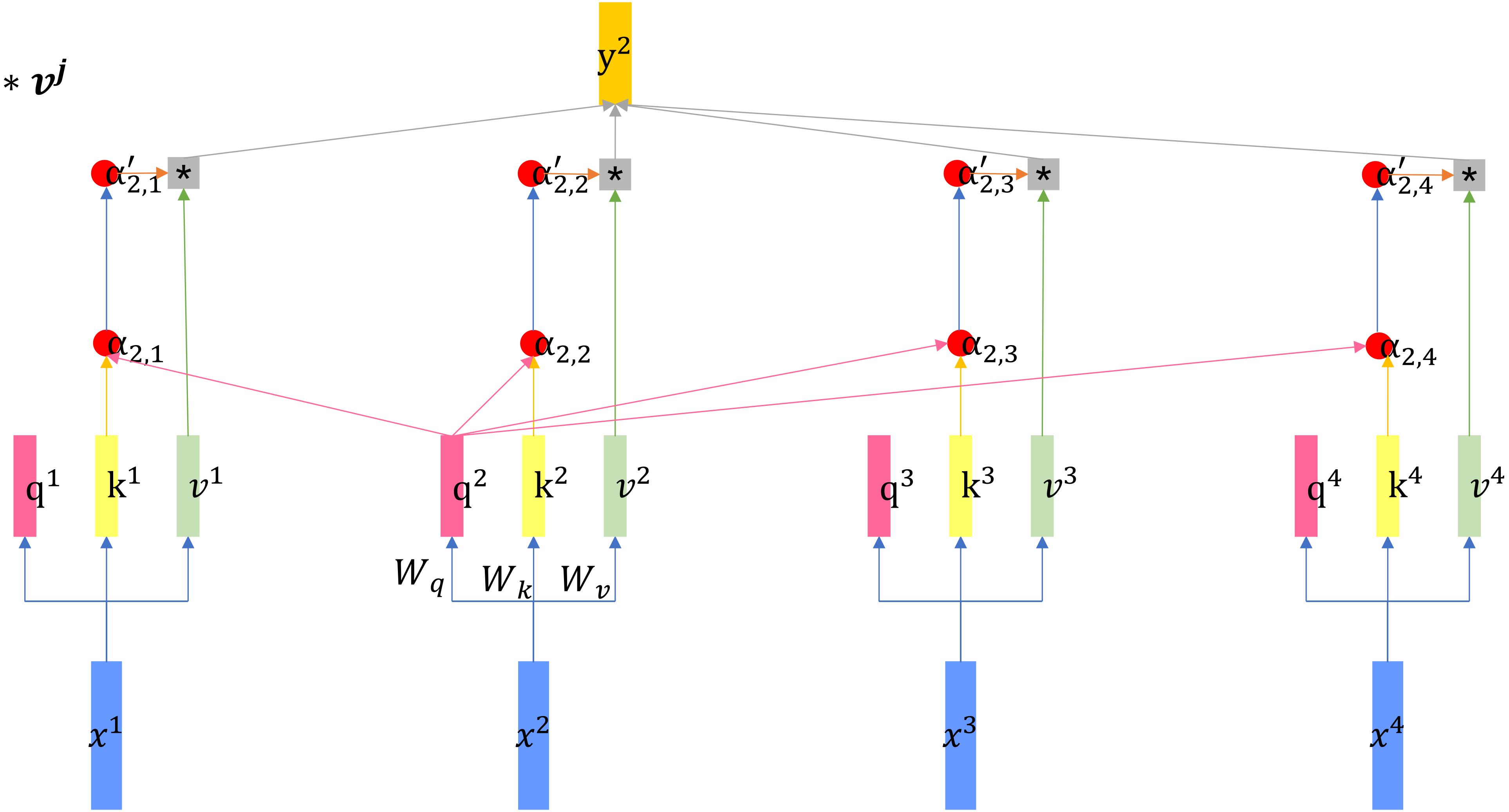
Attention

$$\boldsymbol{y^i = \sum_j \alpha'_{i,j} * v^j}$$



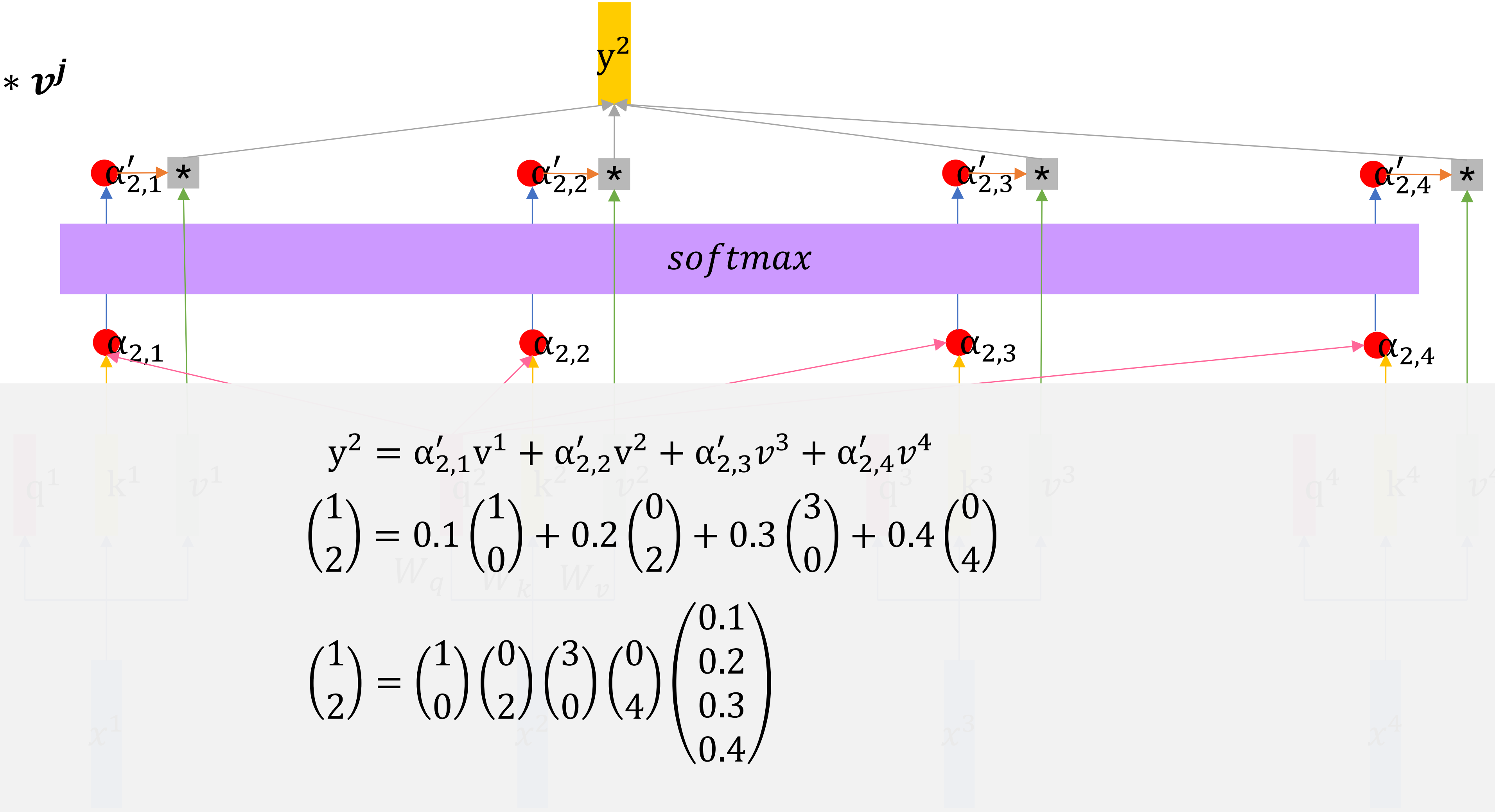
Attention

$$\boldsymbol{y^i = \sum_j \alpha'_{i,j} * v^j}$$



Attention

$$y^i = \sum_j \alpha'_{i,j} * v^j$$



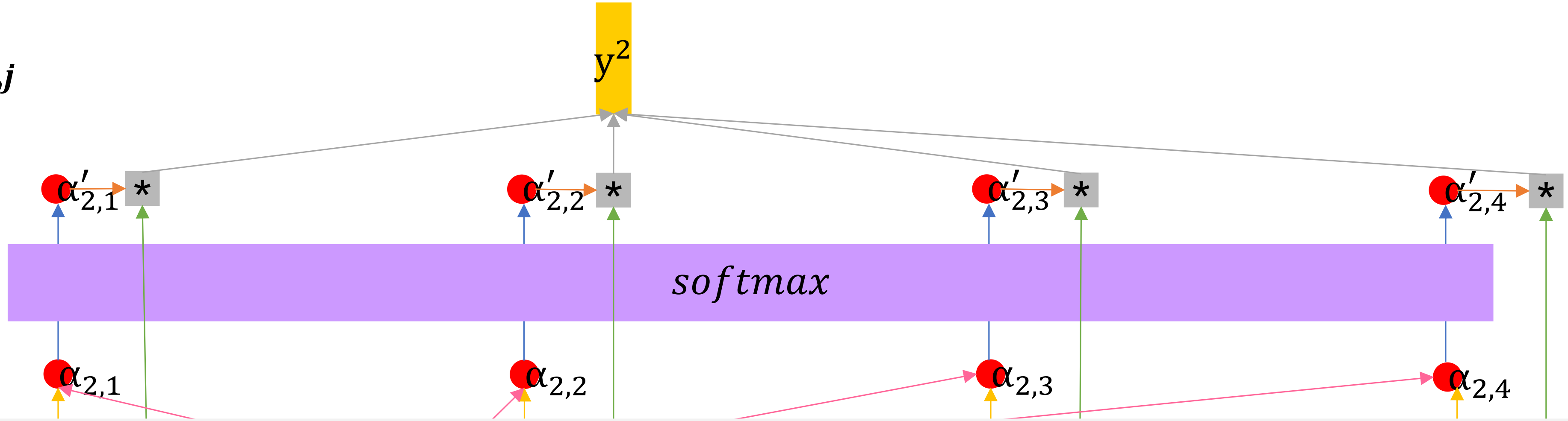
$$y^2 = \alpha'_{2,1}v^1 + \alpha'_{2,2}v^2 + \alpha'_{2,3}v^3 + \alpha'_{2,4}v^4$$

$$\begin{pmatrix}1\\2\end{pmatrix} = 0.1\begin{pmatrix}1\\0\end{pmatrix} + 0.2\begin{pmatrix}0\\2\end{pmatrix} + 0.3\begin{pmatrix}3\\0\end{pmatrix} + 0.4\begin{pmatrix}0\\4\end{pmatrix}$$

$$\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}0\\2\end{pmatrix}\begin{pmatrix}3\\0\end{pmatrix}\begin{pmatrix}0\\4\end{pmatrix}\begin{pmatrix}0.1\\0.2\\0.3\\0.4\end{pmatrix}$$

Attention

$$\boldsymbol{y^i = \sum_j \alpha'_{i,j} * v^j}$$



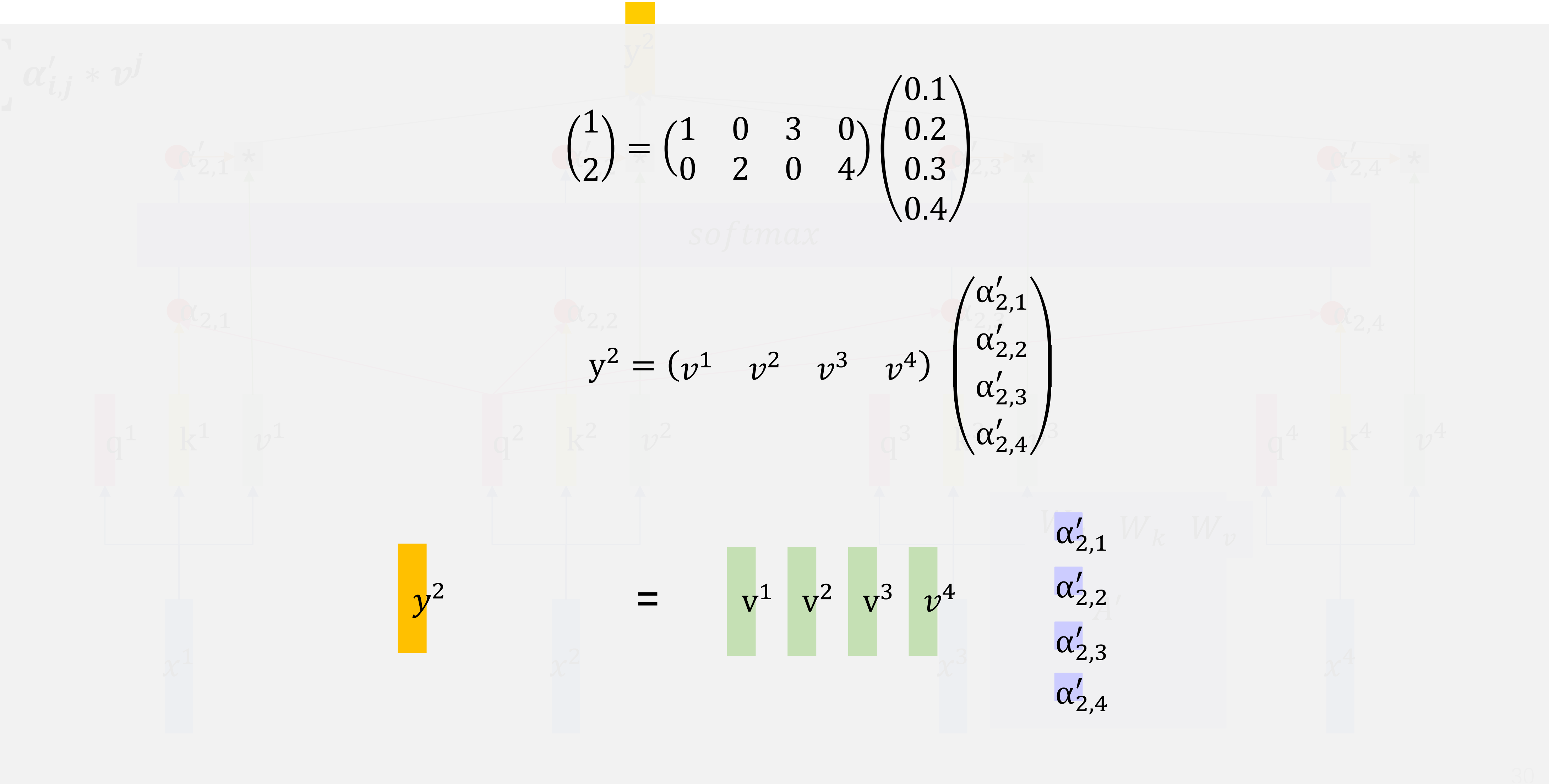
$$y^2 = \alpha'_{2,1}v^1 + \alpha'_{2,2}v^2 + \alpha'_{2,3}v^3 + \alpha'_{2,4}v^4$$

$$\begin{pmatrix}1\\2\end{pmatrix} = 0.1\begin{pmatrix}1\\0\end{pmatrix} + 0.2\begin{pmatrix}0\\2\end{pmatrix} + 0.3\begin{pmatrix}3\\0\end{pmatrix} + 0.4\begin{pmatrix}0\\4\end{pmatrix}$$

$$\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}1 & 0 & 3 & 0\\0 & 2 & 0 & 4\end{pmatrix}\begin{pmatrix}0.1\\0.2\\0.3\\0.4\end{pmatrix} \qquad y^2 = \begin{pmatrix}v^1 & v^2 & v^3 & v^4\end{pmatrix}\begin{pmatrix}\alpha'_{2,1}\\\alpha'_{2,2}\\\alpha'_{2,3}\\\alpha'_{2,4}\end{pmatrix}$$

Attention

$$y^i = \sum_j \alpha'_{i,j} * v^j$$

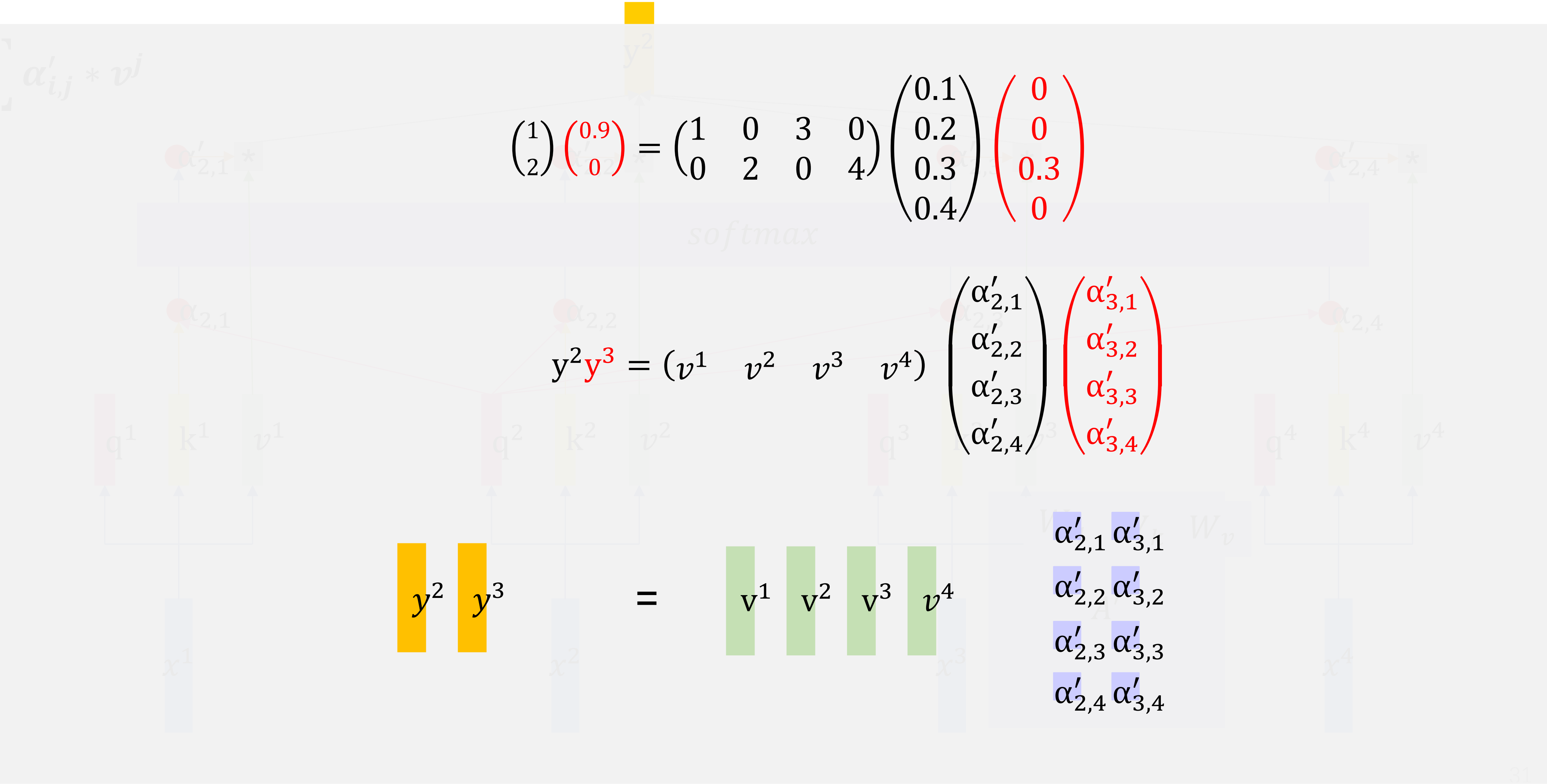


$$\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}1 & 0 & 3 & 0\\0 & 2 & 0 & 4\end{pmatrix}\begin{pmatrix}0.1\\0.2\\0.3\\0.4\end{pmatrix}$$

$$y^2 = \begin{pmatrix}v^1 & v^2 & v^3 & v^4\end{pmatrix}\begin{pmatrix}\alpha'_{2,1}\\\alpha'_{2,2}\\\alpha'_{2,3}\\\alpha'_{2,4}\end{pmatrix}$$

Attention

$$\boldsymbol{y^i} = \sum_{\boldsymbol{j}} \alpha'_{i,j} * v^j$$



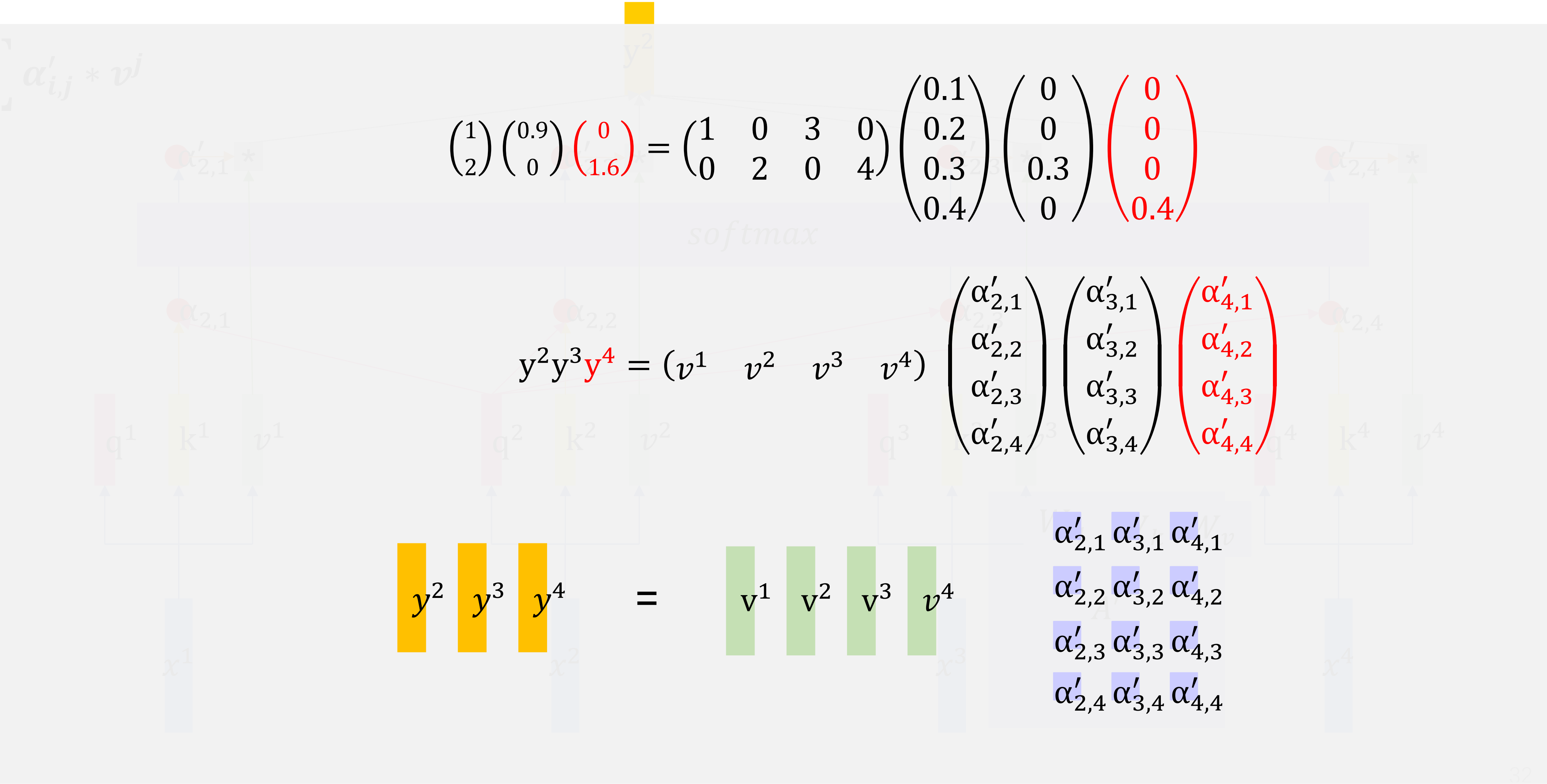
$$\begin{pmatrix}1\\2\end{pmatrix}\begin{pmatrix}0.9\\0\end{pmatrix} = \begin{pmatrix}1 & 0 & 3 & 0\\0 & 2 & 0 & 4\end{pmatrix}\begin{pmatrix}0.1\\0.2\\0.3\\0.4\end{pmatrix}\begin{pmatrix}0\\0\\0.3\\0\end{pmatrix}$$

$$\mathrm{y}^2\mathrm{y}^3 = \begin{pmatrix}v^1 & v^2 & v^3 & v^4\end{pmatrix}\begin{pmatrix}\alpha'_{2,1}\\\alpha'_{2,2}\\\alpha'_{2,3}\\\alpha'_{2,4}\end{pmatrix}\begin{pmatrix}\alpha'_{3,1}\\\alpha'_{3,2}\\\alpha'_{3,3}\\\alpha'_{3,4}\end{pmatrix}$$

$$y^2 \quad y^3 \quad = \quad \mathrm{v}^1 \quad \mathrm{v}^2 \quad \mathrm{v}^3 \quad v^4 \quad \begin{matrix}\alpha'_{2,1} & \alpha'_{3,1}\\\alpha'_{2,2} & \alpha'_{3,2}\\\alpha'_{2,3} & \alpha'_{3,3}\\\alpha'_{2,4} & \alpha'_{3,4}\end{matrix}$$

Attention

$$y^i = \sum_j \alpha'_{i,j} * v^j$$



$$\begin{pmatrix}1\\2\end{pmatrix}\begin{pmatrix}0.9\\0\end{pmatrix}{\color{red}\begin{pmatrix}0\\1.6\end{pmatrix}} = \begin{pmatrix}1&0&3&0\\0&2&0&4\end{pmatrix}\begin{pmatrix}0.1\\0.2\\0.3\\0.4\end{pmatrix}\begin{pmatrix}0\\0\\0.3\\0\end{pmatrix}{\color{red}\begin{pmatrix}0\\0\\0\\0.4\end{pmatrix}}$$

$$\mathrm{y}^2\mathrm{y}^3{\color{red}\mathrm{y}^4} = \begin{pmatrix}v^1 & v^2 & v^3 & v^4\end{pmatrix}\begin{pmatrix}\alpha'_{2,1}\\\alpha'_{2,2}\\\alpha'_{2,3}\\\alpha'_{2,4}\end{pmatrix}\begin{pmatrix}\alpha'_{3,1}\\\alpha'_{3,2}\\\alpha'_{3,3}\\\alpha'_{3,4}\end{pmatrix}{\color{red}\begin{pmatrix}\alpha'_{4,1}\\\alpha'_{4,2}\\\alpha'_{4,3}\\\alpha'_{4,4}\end{pmatrix}}$$

$$\begin{matrix}y^2 & y^3 & y^4\end{matrix} = \begin{matrix}\mathrm{v}^1 & \mathrm{v}^2 & \mathrm{v}^3 & v^4\end{matrix} \quad \begin{matrix}\alpha'_{2,1} & \alpha'_{3,1} & \alpha'_{4,1}\\ \alpha'_{2,2} & \alpha'_{3,2} & \alpha'_{4,2}\\ \alpha'_{2,3} & \alpha'_{3,3} & \alpha'_{4,3}\\ \alpha'_{2,4} & \alpha'_{3,4} & \alpha'_{4,4}\end{matrix}$$

Attention

$$y^i = \sum_j$$

$$\begin{pmatrix}0.1\\0\end{pmatrix}\begin{pmatrix}1\\2\end{pmatrix}\begin{pmatrix}0.9\\0\end{pmatrix}\begin{pmatrix}0\\1.6\end{pmatrix} = \begin{pmatrix}1 & 0 & 3 & 0\\0 & 2 & 0 & 4\end{pmatrix}\begin{pmatrix}0.1\\0\\0\\0\end{pmatrix}\begin{pmatrix}0.1\\0.2\\0.3\\0.4\end{pmatrix}\begin{pmatrix}0\\0\\0.3\\0\end{pmatrix}\begin{pmatrix}0\\0\\0\\0.4\end{pmatrix}$$

$$\mathrm{y}^1\mathrm{y}^2\mathrm{y}^3\mathrm{y}^4 = \begin{pmatrix}v^1 & v^2 & v^3 & v^4\end{pmatrix}\begin{pmatrix}\alpha'_{1,1}\\\alpha'_{1,2}\\\alpha'_{1,3}\\\alpha'_{1,4}\end{pmatrix}\begin{pmatrix}\alpha'_{2,1}\\\alpha'_{2,2}\\\alpha'_{2,3}\\\alpha'_{2,4}\end{pmatrix}\begin{pmatrix}\alpha'_{3,1}\\\alpha'_{3,2}\\\alpha'_{3,3}\\\alpha'_{3,4}\end{pmatrix}\begin{pmatrix}\alpha'_{4,1}\\\alpha'_{4,2}\\\alpha'_{4,3}\\\alpha'_{4,4}\end{pmatrix}$$

$$y^1 \quad y^2 \quad y^3 \quad y^4 \quad = \quad \mathrm{v}^1 \quad \mathrm{v}^2 \quad \mathrm{v}^3 \quad v^4 \quad \begin{matrix}\alpha'_{1,1} & \alpha'_{2,1} & \alpha'_{3,1} & \alpha'_{4,1}\\\alpha'_{1,2} & \alpha'_{2,2} & \alpha'_{3,2} & \alpha'_{4,2}\\\alpha'_{1,3} & \alpha'_{2,3} & \alpha'_{3,3} & \alpha'_{4,3}\\\alpha'_{1,4} & \alpha'_{2,4} & \alpha'_{3,4} & \alpha'_{4,4}\end{matrix}$$

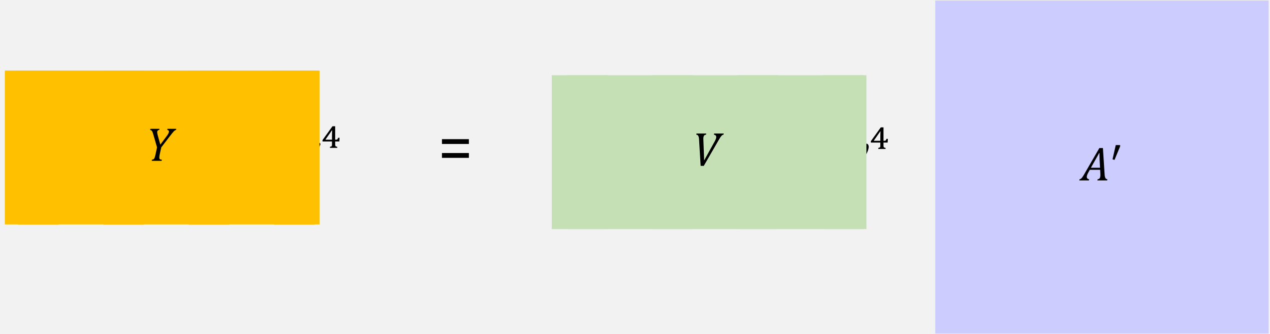
Attention

$$y^i = \sum_j$$

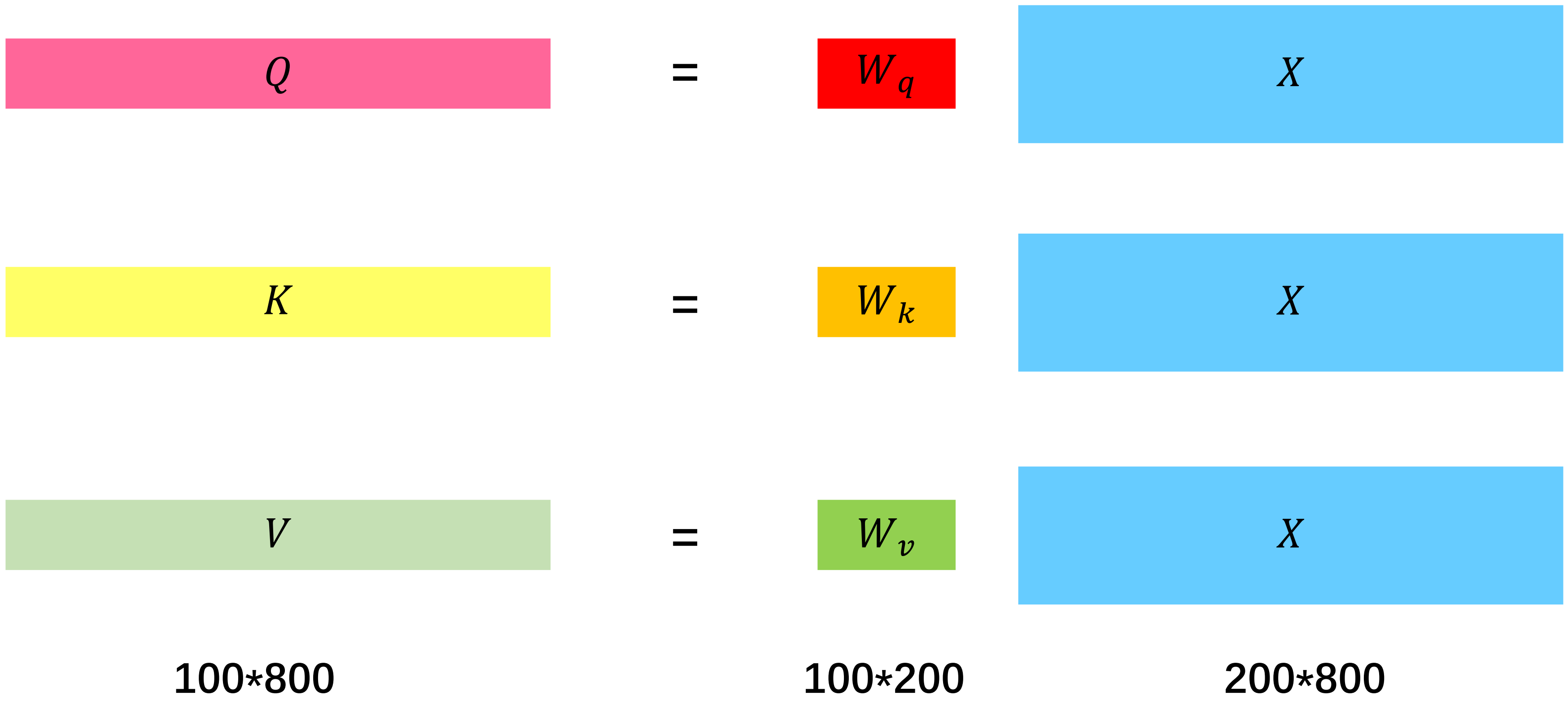
$$\begin{pmatrix} 0.1 & 1 & 0.9 & 0 \\ 0 & 2 & 0 & 1.6 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 4 \end{pmatrix} \begin{pmatrix} 0.1 & 0.1 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0.3 & 0.3 & 0 \\ 0 & 0.4 & 0 & 0.4 \end{pmatrix}$$

$$\mathrm{y}^1\mathrm{y}^2\mathrm{y}^3\mathrm{y}^4 = \begin{pmatrix} v^1 & v^2 & v^3 & v^4 \end{pmatrix} \begin{pmatrix} \alpha'_{1,1} & \alpha'_{2,1} & \alpha'_{3,1} & \alpha'_{4,1} \\ \alpha'_{1,2} & \alpha'_{2,2} & \alpha'_{3,2} & \alpha'_{4,2} \\ \alpha'_{1,3} & \alpha'_{2,3} & \alpha'_{3,3} & \alpha'_{4,3} \\ \alpha'_{1,4} & \alpha'_{2,4} & \alpha'_{3,4} & \alpha'_{4,4} \end{pmatrix}$$

$$\boldsymbol{Y = VA'}$$

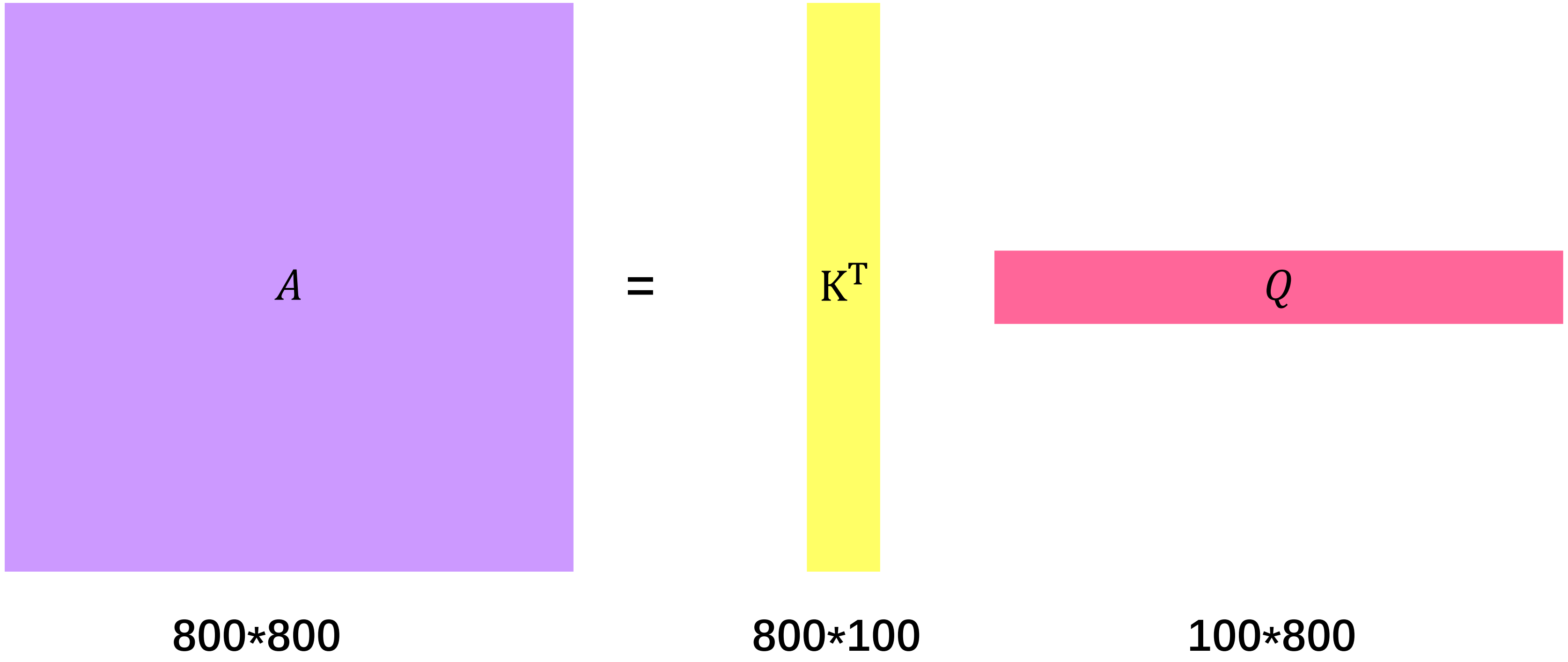


Attention



$$\boldsymbol{Q} = \boldsymbol{W}_q \boldsymbol{X}$$
$$\boldsymbol{K} = \boldsymbol{W}_k \boldsymbol{X}$$
$$\boldsymbol{V} = \boldsymbol{W}_v \boldsymbol{X}$$

Attention

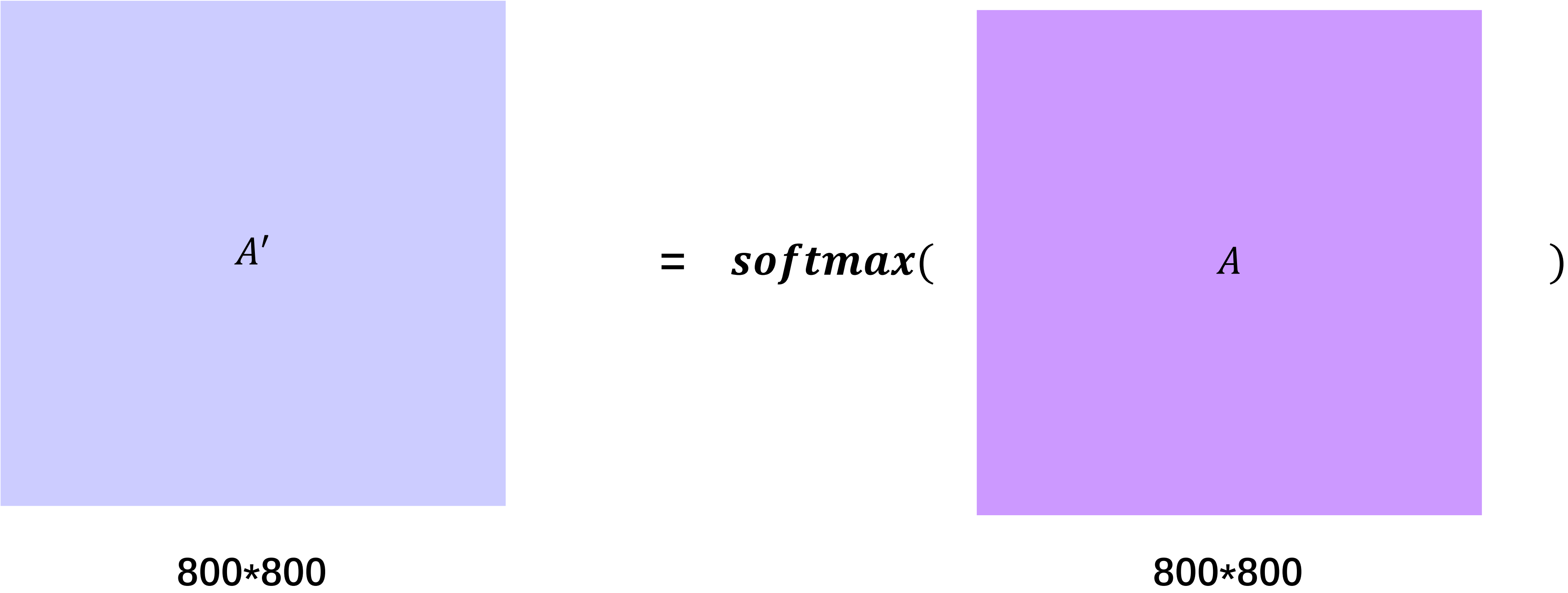


A = K^{T} Q

800*800 800*100 100*800

$$\boldsymbol{A = K^T Q}$$
$$\boldsymbol{A' = softmax(A)}$$

Attention



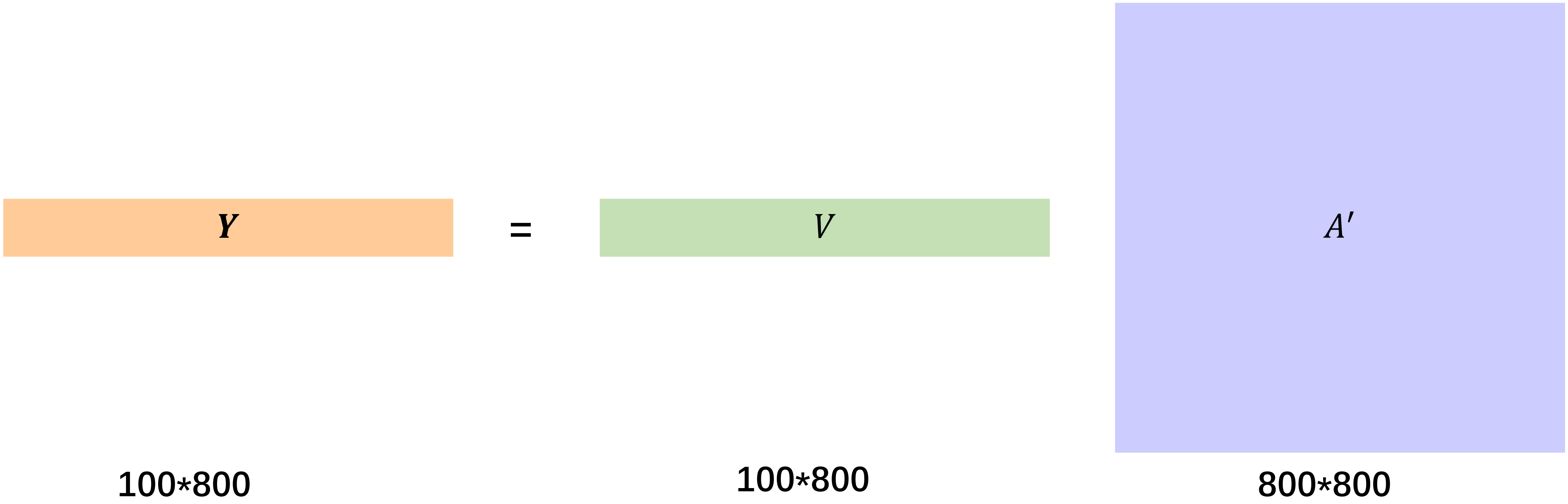
A' = $\boldsymbol{softmax}($ A $)$

800*800

800*800

$\boldsymbol{A' = softmax(A)}$

Self-Attention



$\boldsymbol{Y}$ = V A'

100*800

100*800

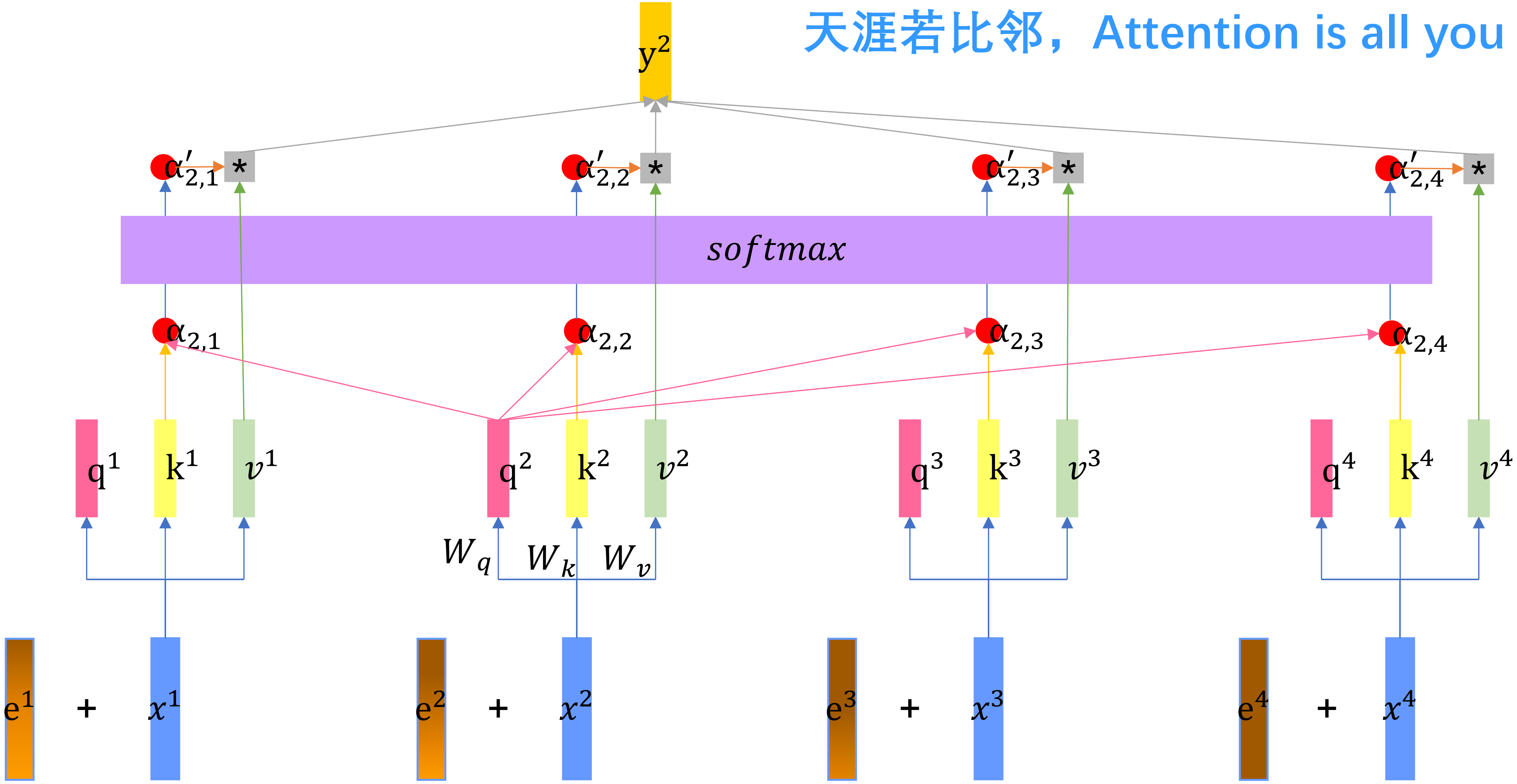
800*800

$$\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{A}'$$

Attention

位置编码

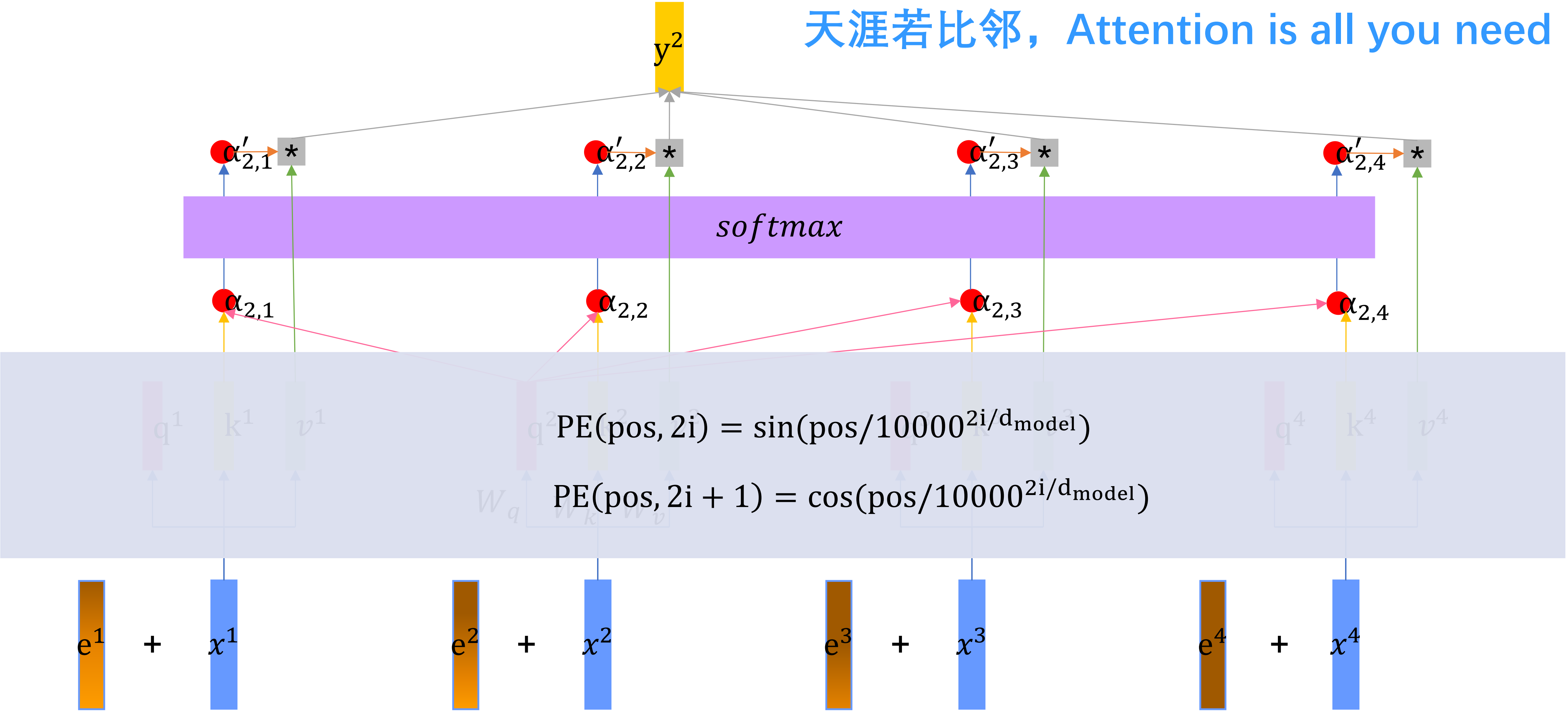
天涯若比邻，Attention is all you need



Attention

位置编码

天涯若比邻，Attention is all you need



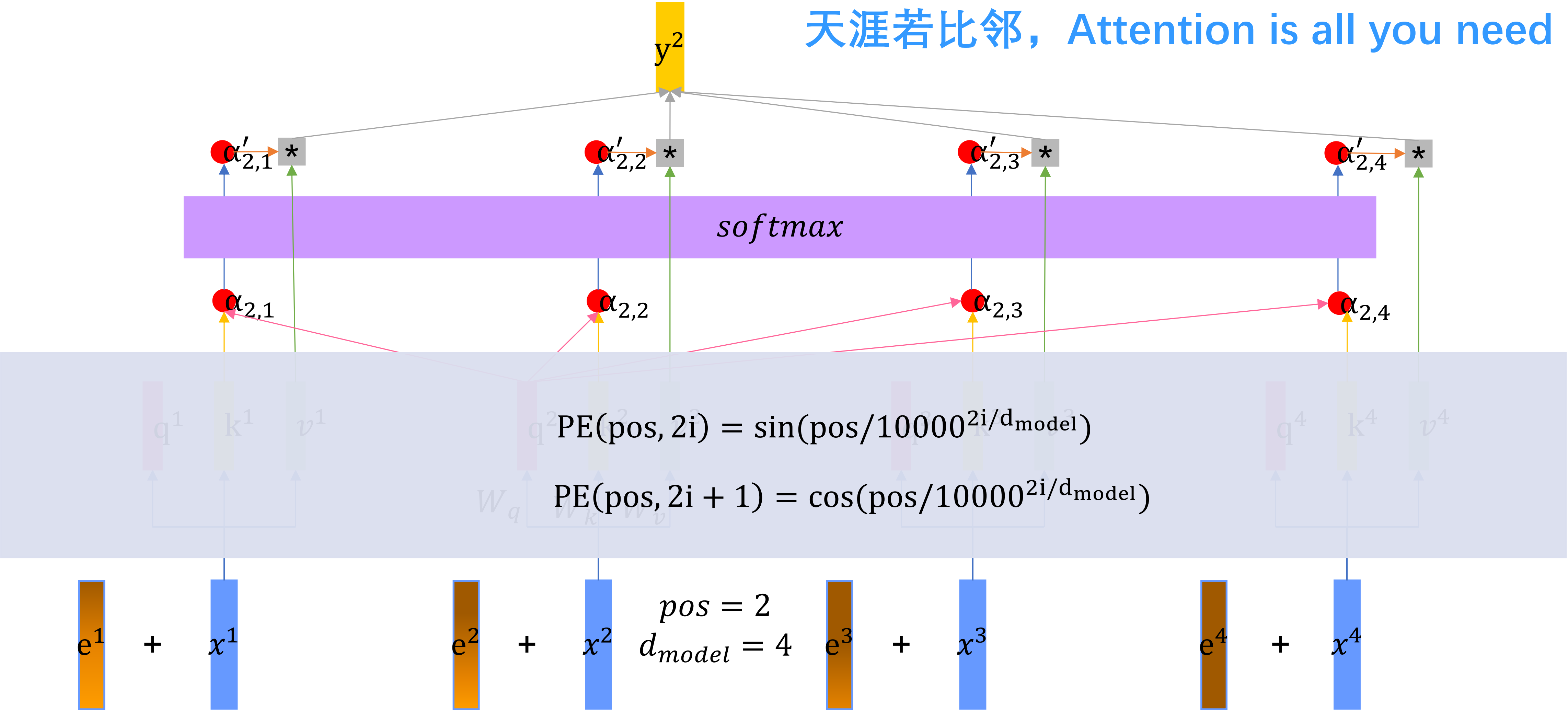
$$PE(pos, 2i) = \sin(pos/10000^{2i/d_{model}})$$

$$PE(pos, 2i + 1) = \cos(pos/10000^{2i/d_{model}})$$

Attention

位置编码

天涯若比邻，Attention is all you need



$$\mathrm{PE}(\mathrm{pos}, 2\mathrm{i}) = \sin(\mathrm{pos}/10000^{2\mathrm{i}/\mathrm{d}_{\mathrm{model}}})$$

$$\mathrm{PE}(\mathrm{pos}, 2\mathrm{i}+1) = \cos(\mathrm{pos}/10000^{2\mathrm{i}/\mathrm{d}_{\mathrm{model}}})$$

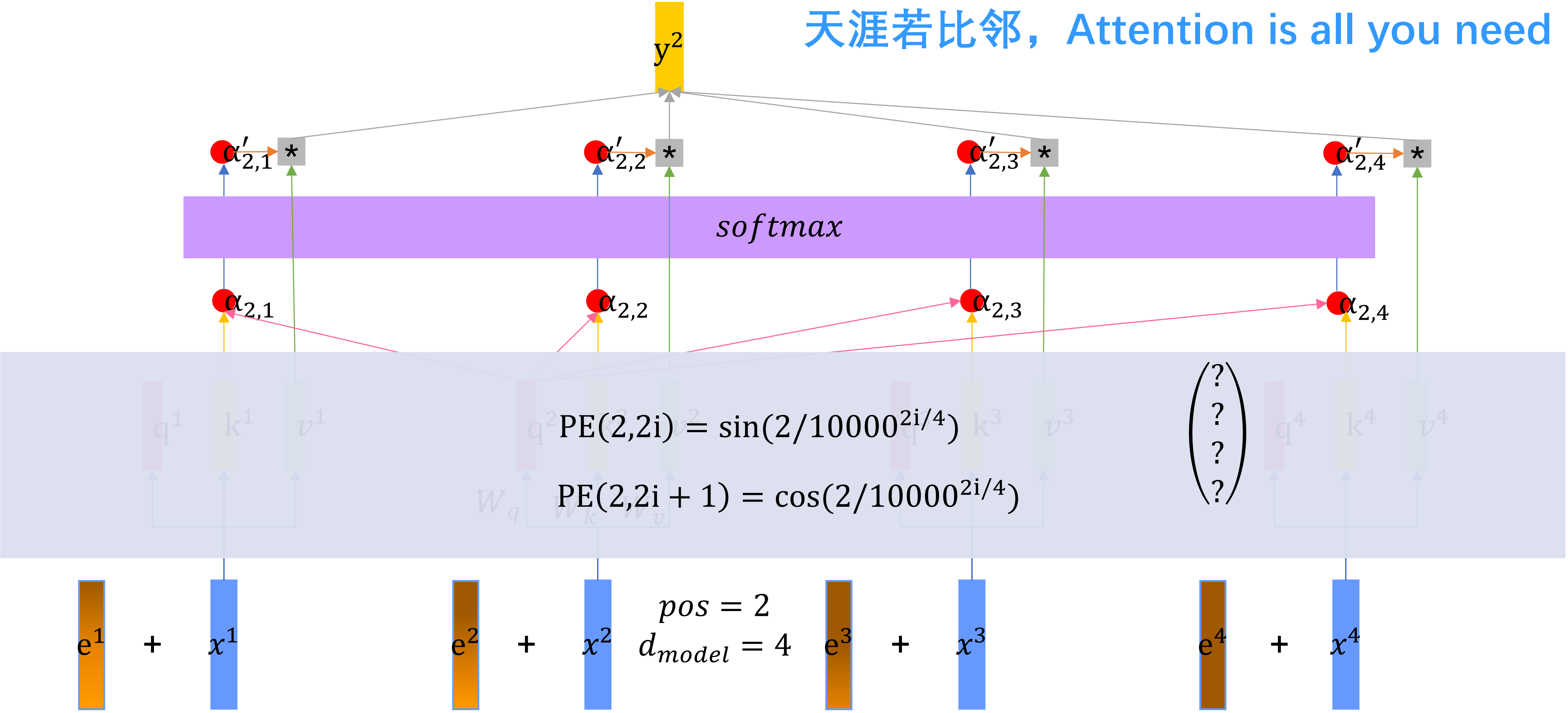
$$pos = 2$$

$$d_{model} = 4$$

Attention

位置编码

天涯若比邻，Attention is all you need



$$\mathrm{PE}(2,2\mathrm{i}) = \sin(2/10000^{2\mathrm{i}/4})$$

$$\mathrm{PE}(2,2\mathrm{i}+1) = \cos(2/10000^{2\mathrm{i}/4})$$

$$\begin{pmatrix} ? \\ ? \\ ? \\ ? \end{pmatrix}$$

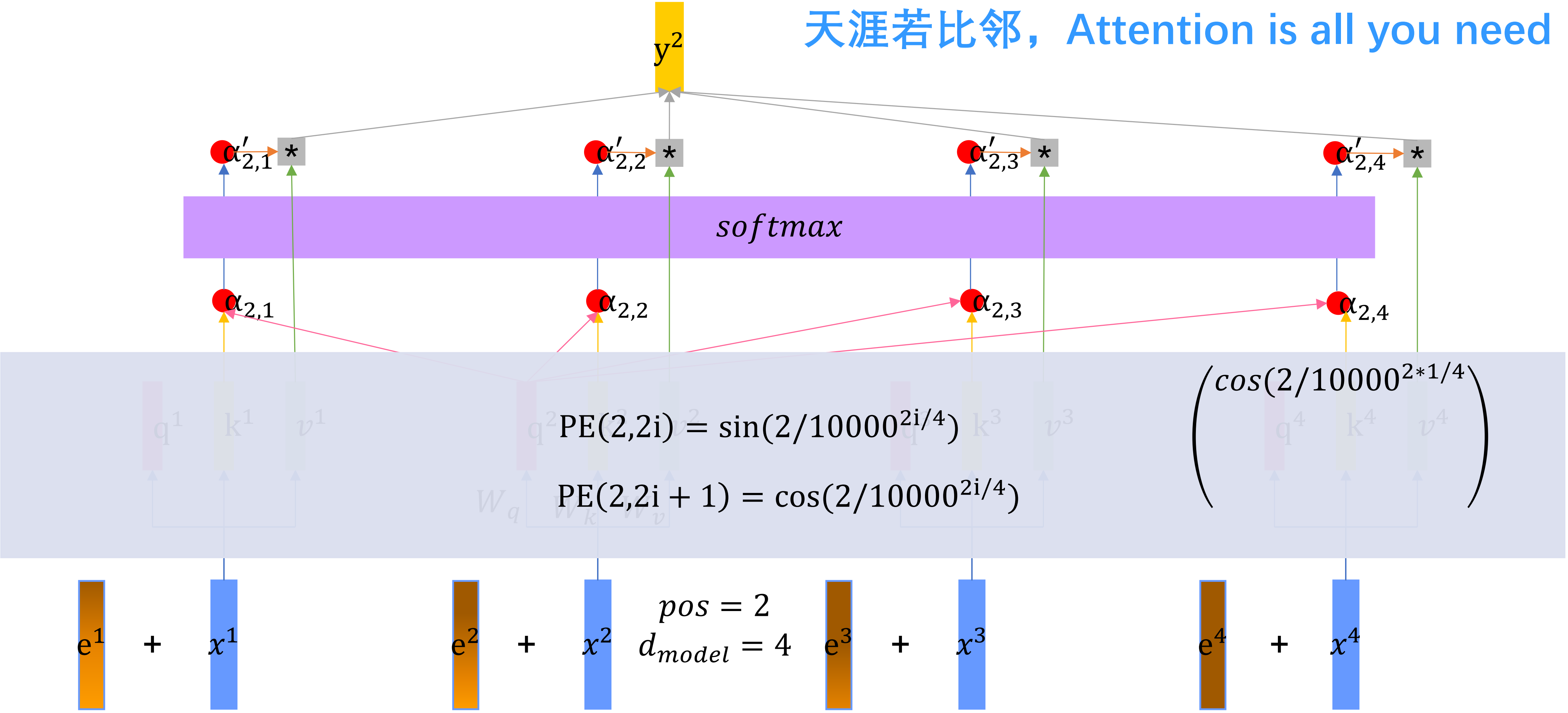
$$pos = 2$$

$$d_{model} = 4$$

Attention

位置编码

天涯若比邻，Attention is all you need



$$\mathrm{PE}(2,2\mathrm{i}) = \sin(2/10000^{2\mathrm{i}/4})$$

$$\mathrm{PE}(2,2\mathrm{i}+1) = \cos(2/10000^{2\mathrm{i}/4})$$

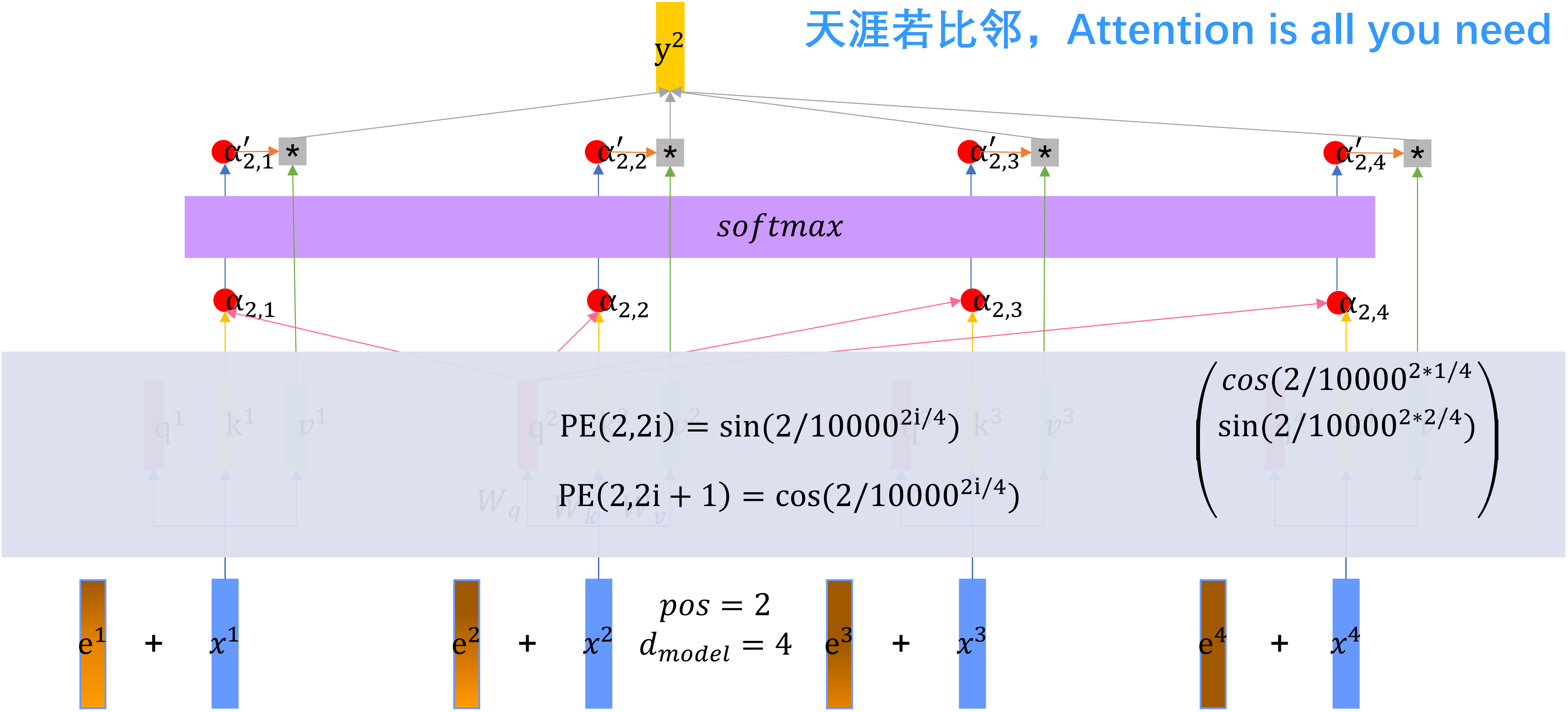
$$\begin{pmatrix} cos(2/10000^{2*1/4}) \\ \\ \\ \end{pmatrix}$$

$$pos = 2$$
$$d_{model} = 4$$

Attention

位置编码

天涯若比邻，Attention is all you need



$$\mathrm{PE}(2,2\mathrm{i}) = \sin(2/10000^{2\mathrm{i}/4})$$

$$\mathrm{PE}(2,2\mathrm{i}+1) = \cos(2/10000^{2\mathrm{i}/4})$$

$$\begin{pmatrix} cos(2/10000^{2*1/4} \\ \sin(2/10000^{2*2/4}) \\ \\ \end{pmatrix}$$

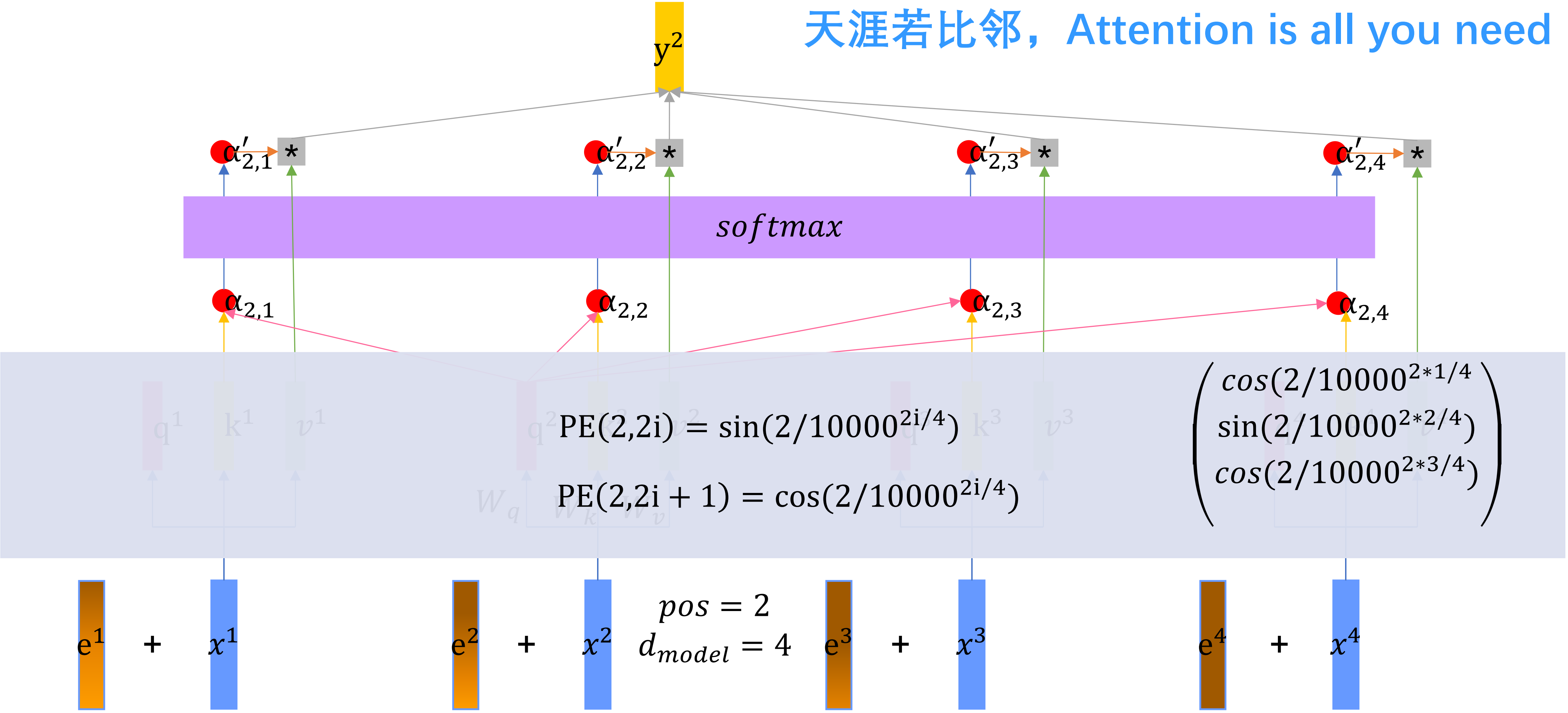
$$pos = 2$$

$$d_{model} = 4$$

Attention

位置编码

天涯若比邻，Attention is all you need



$$\mathrm{PE}(2,2\mathrm{i}) = \sin(2/10000^{2\mathrm{i}/4})$$

$$\mathrm{PE}(2,2\mathrm{i}+1) = \cos(2/10000^{2\mathrm{i}/4})$$

$$\begin{pmatrix} cos(2/10000^{2*1/4} \\ \sin(2/10000^{2*2/4}) \\ cos(2/10000^{2*3/4}) \end{pmatrix}$$

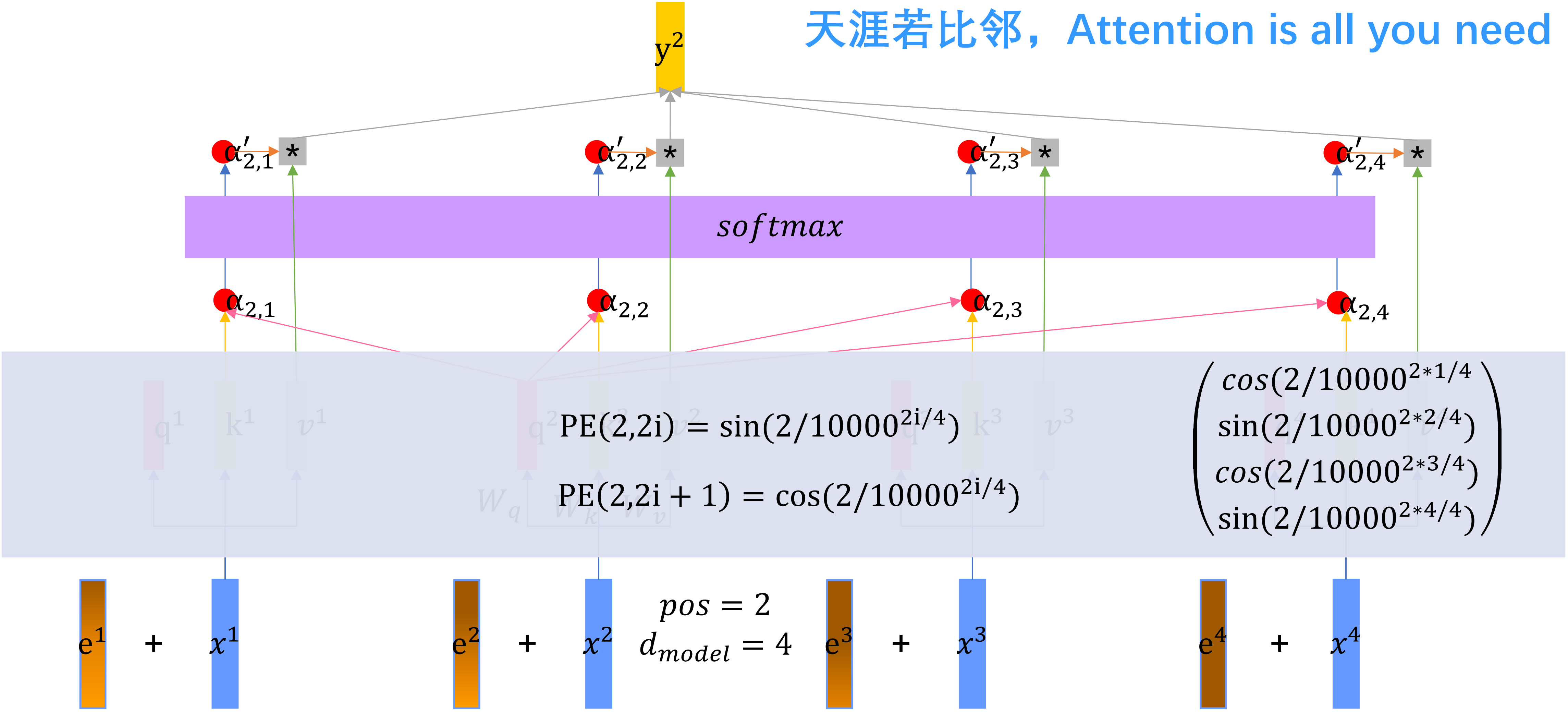
$$pos = 2$$

$$d_{model} = 4$$

Attention

位置编码

天涯若比邻，Attention is all you need



$$PE(2,2i) = \sin(2/10000^{2i/4})$$

$$PE(2,2i+1) = \cos(2/10000^{2i/4})$$

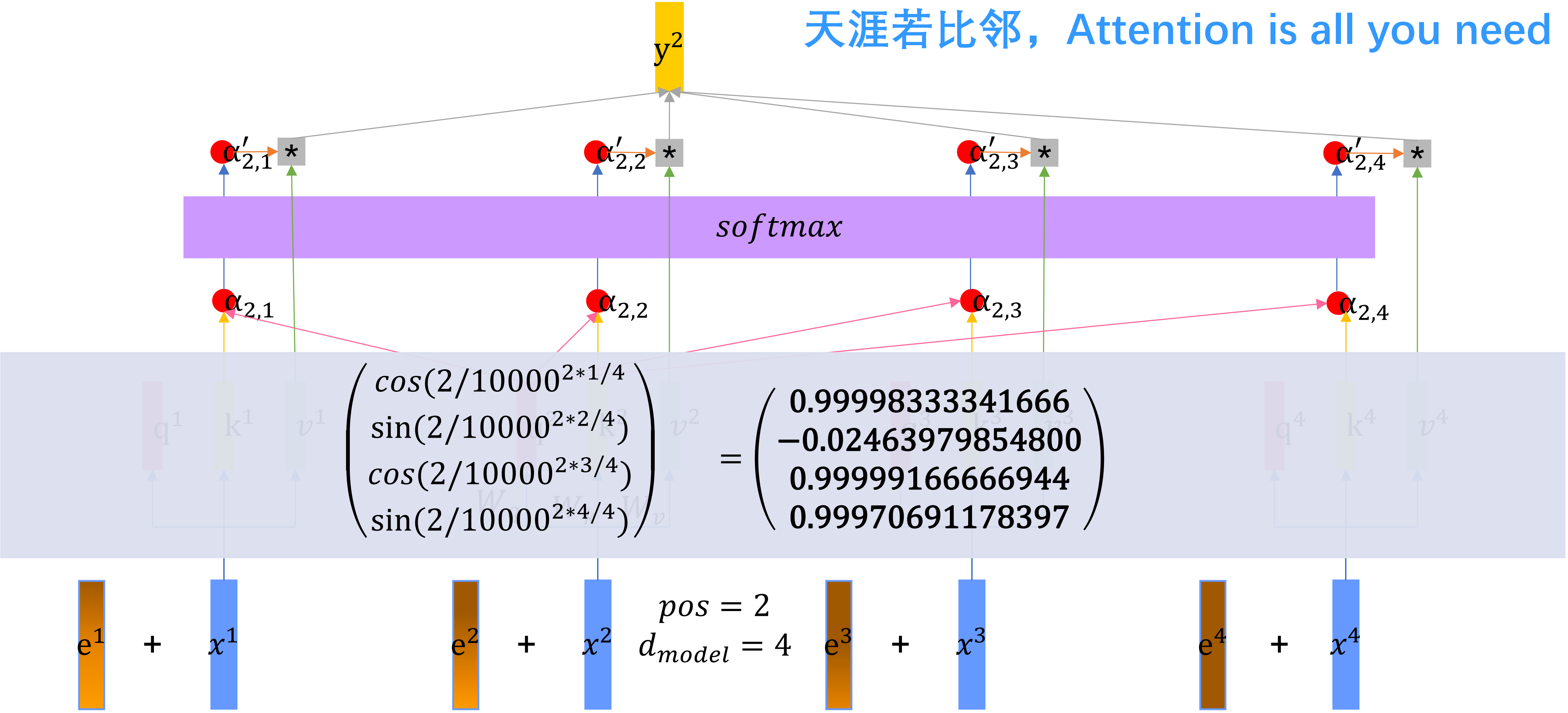
$$\begin{pmatrix} cos(2/10000^{2*1/4}) \\ \sin(2/10000^{2*2/4}) \\ cos(2/10000^{2*3/4}) \\ \sin(2/10000^{2*4/4}) \end{pmatrix}$$

$pos = 2$

$d_{model} = 4$

Attention

位置编码

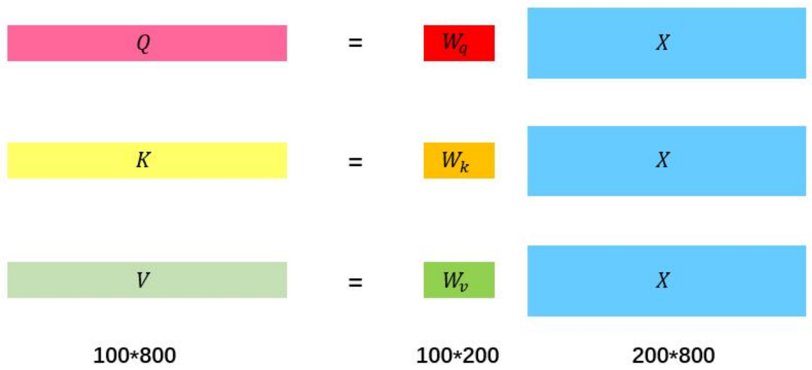


天涯若比邻，Attention is all you need

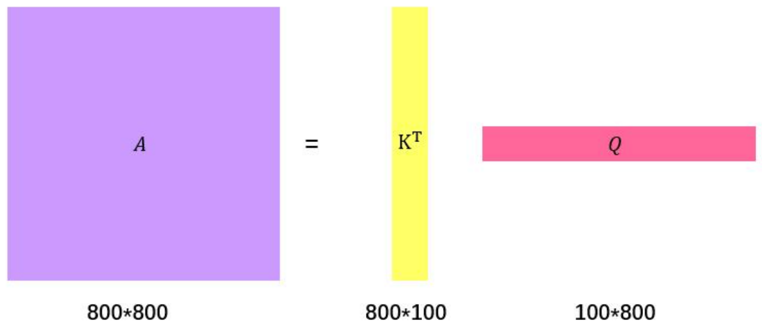
$$\begin{pmatrix} cos(2/10000^{2*1/4}) \\ \sin(2/10000^{2*2/4}) \\ cos(2/10000^{2*3/4}) \\ \sin(2/10000^{2*4/4}) \end{pmatrix} = \begin{pmatrix} 0.99998333341666 \\ -0.02463979854800 \\ 0.99999166666944 \\ 0.99970691178397 \end{pmatrix}$$

$$pos = 2$$
$$d_{model} = 4$$

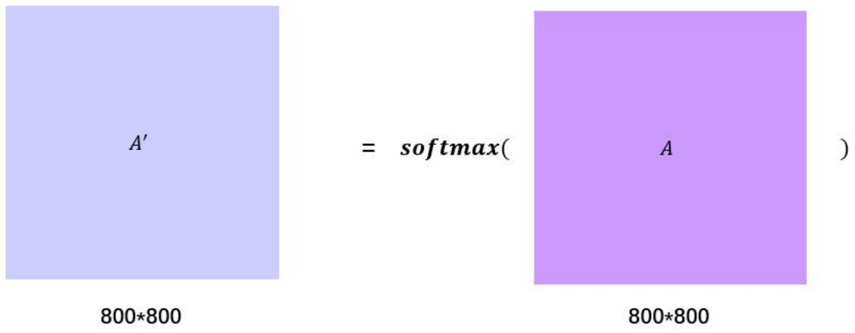
Attention



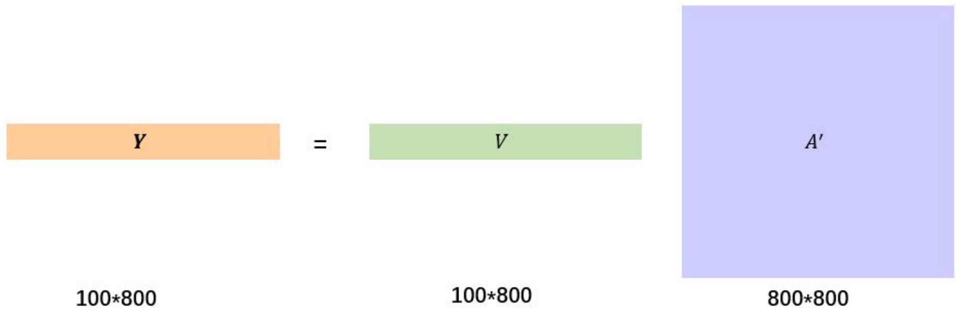
(1)
$$\begin{aligned} Q &= W_q X \\ K &= W_k X \\ V &= W_v X \end{aligned}$$



(2) $$A = K^T Q$$



(3) $$A' = softmax(A)$$



(4) $$Y = VA'$$