

bilinear : a function takes two vectors and return a scalar

---

$$[x_{i1}, x_{i2}, x_{i3}] \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} x_{j1} \\ x_{j2} \\ x_{j3} \end{bmatrix}$$

$$= (x_{i1}w_{11} + x_{i2}w_{21} + x_{i3}w_{31}) \cdot x_{j1} + (x_{i1}w_{12} + x_{i2}w_{22} + x_{i3}w_{32}) \cdot x_{j2} + (x_{i1}w_{13} + x_{i2}w_{23} + x_{i3}w_{33}) \cdot x_{j3}$$

---

$$[a_1, \dots, a_n] \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \dots & w_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \sum_{i,j} a_i w_{ij} b_j \quad \text{— bilinear product.}$$

$$[a_1, \dots, a_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \sum_{i,j} a_i \cdot b_j \quad \text{— dot product}$$

# Self Attention:

$$\alpha_{k,i,j} = \text{softmax} (X_i^T W_k X_j) \quad \left| \quad X'_{k,i} = \sum_{j=1}^n \alpha_{k,i,j} V_k X_j$$

↑     ↑  
1-n   1-n

---

$$\text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V$$

n x n

$d_k = 64 - 3 \text{ blocks}$   
 $d_x = 512 - 4 \text{ blocks}$

~~$$\text{softmax} \left( \frac{X W^Q (X W^K)^T}{\sqrt{d_k}} \right) X W^V$$~~

$$\text{softmax} \left( \frac{X W^Q (X W^K)^T}{\sqrt{d_k}} \right) X W^V$$

$d_x \times d_v$

---

$$Q = X W^Q$$

/     /  
 $n \times d_k$     $n \times d_x$     $d_x \times d_k$

$O(n^2 d)$

# of words tokens