

Self-Atten

Piya Amara Palamure

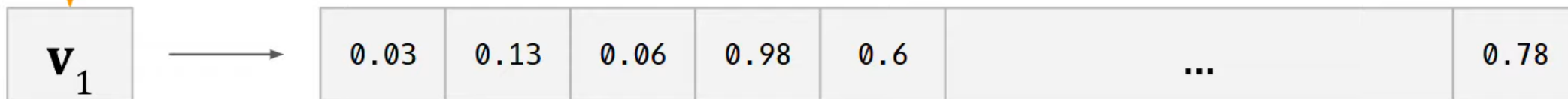
Feb 29, 2024

I swam across the river to get to the other bank

tokenize



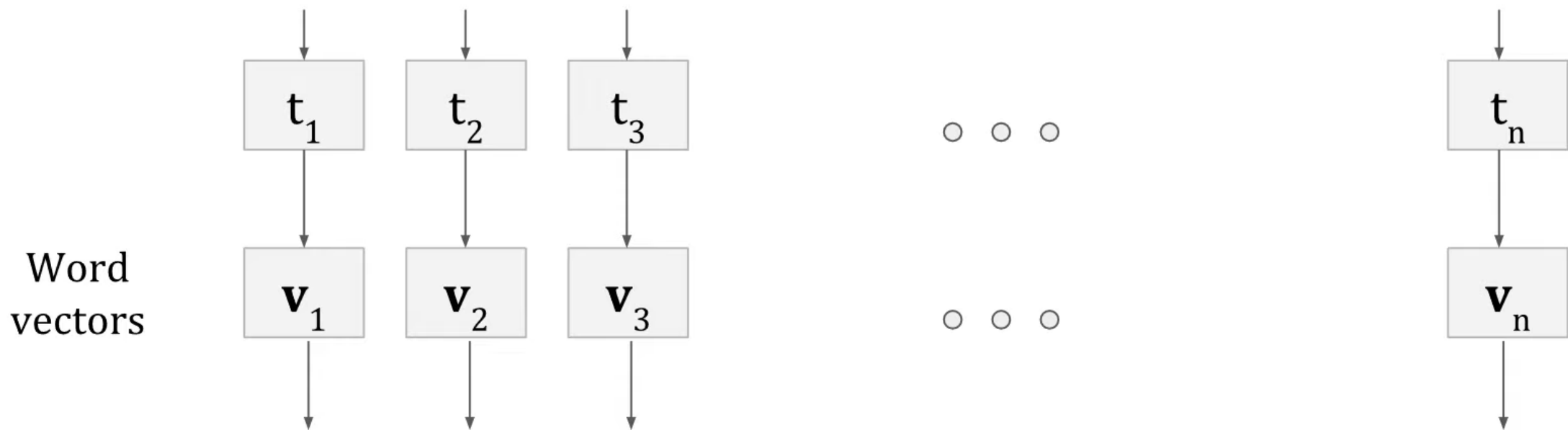
vectorize



N dim vector

50 dim, 100 dim, 200 dim, 300 dim

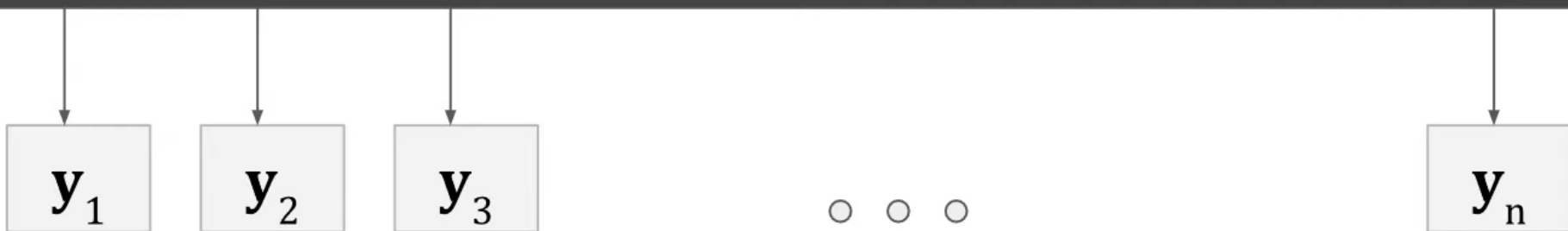
I swam across the river to get to the other bank

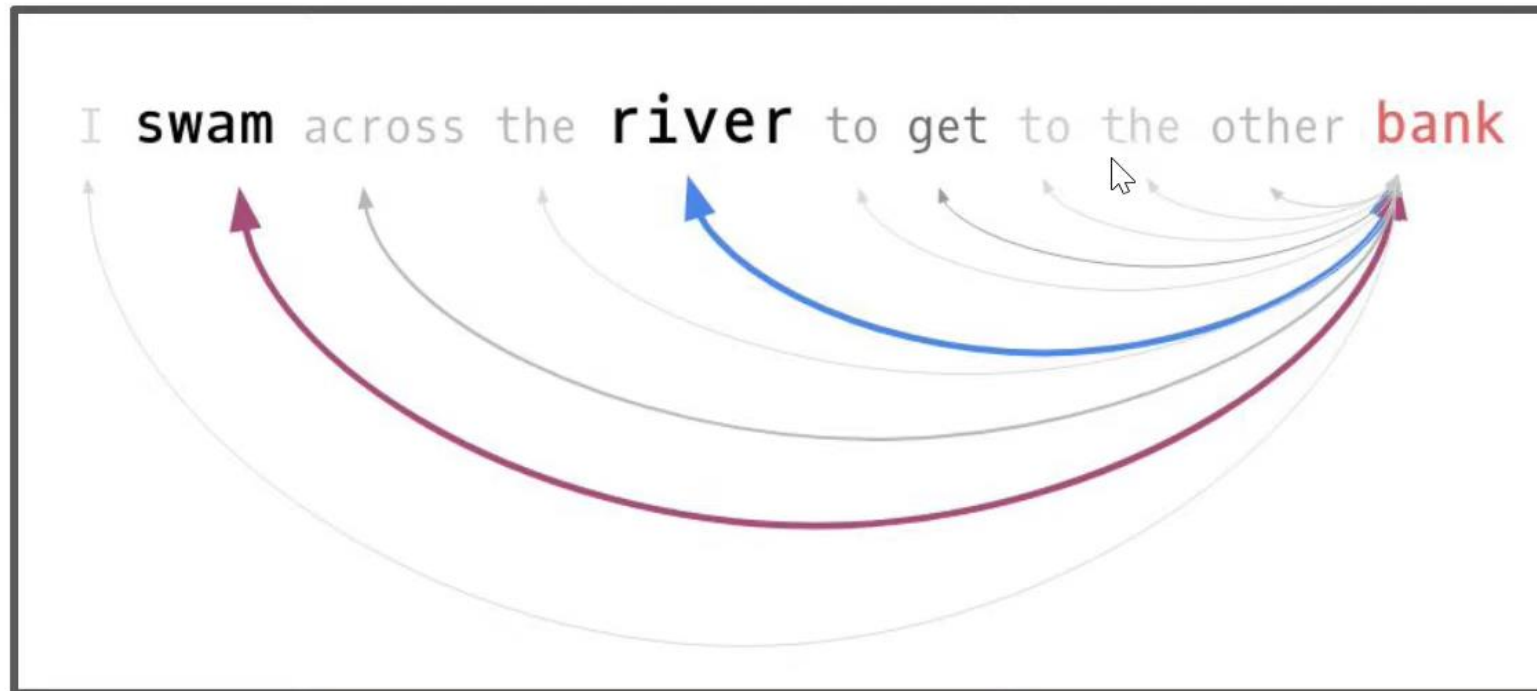
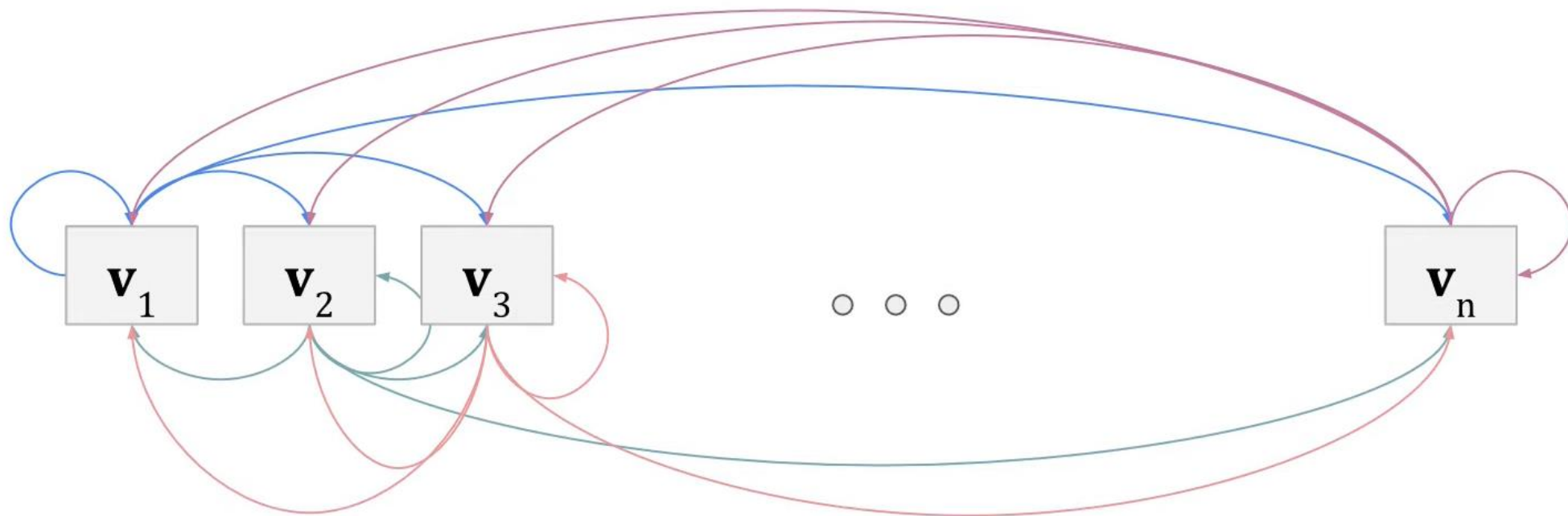


Self-attention mechanism

Weigh word vectors according to context

Contextualized representation of vectors



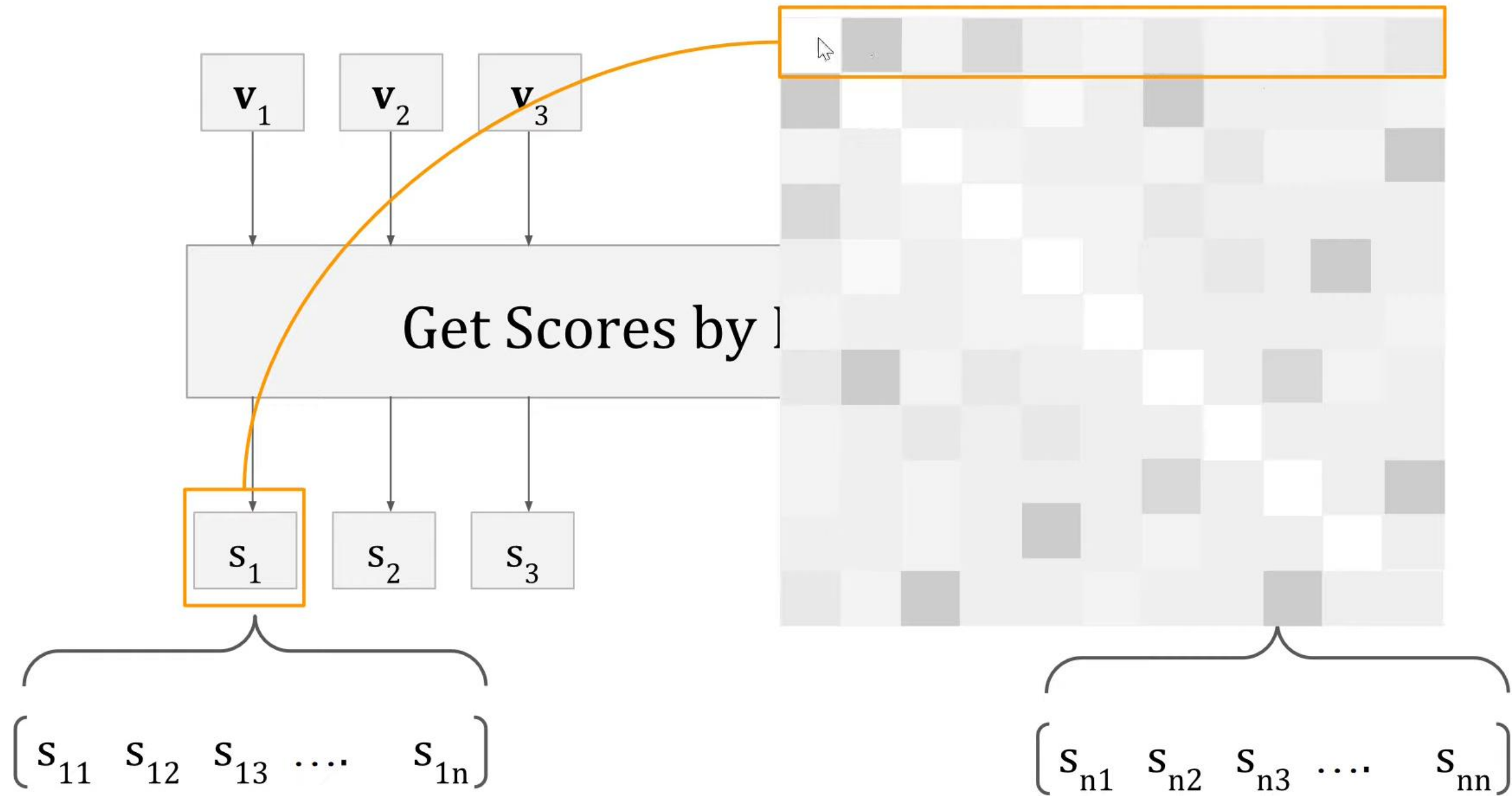


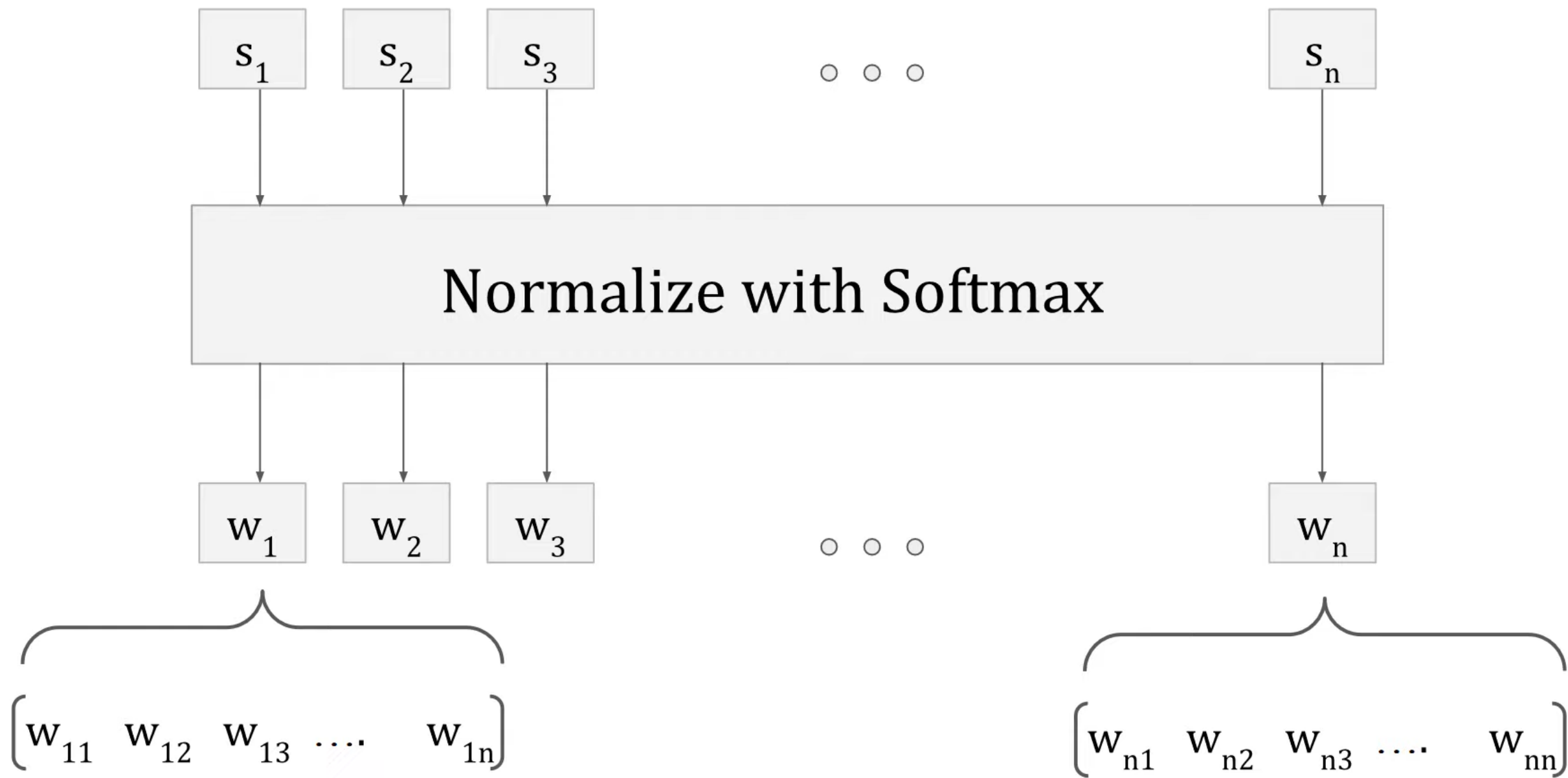
$$S_{n1} = \mathbf{v}_n \mathbf{v}_1^T$$

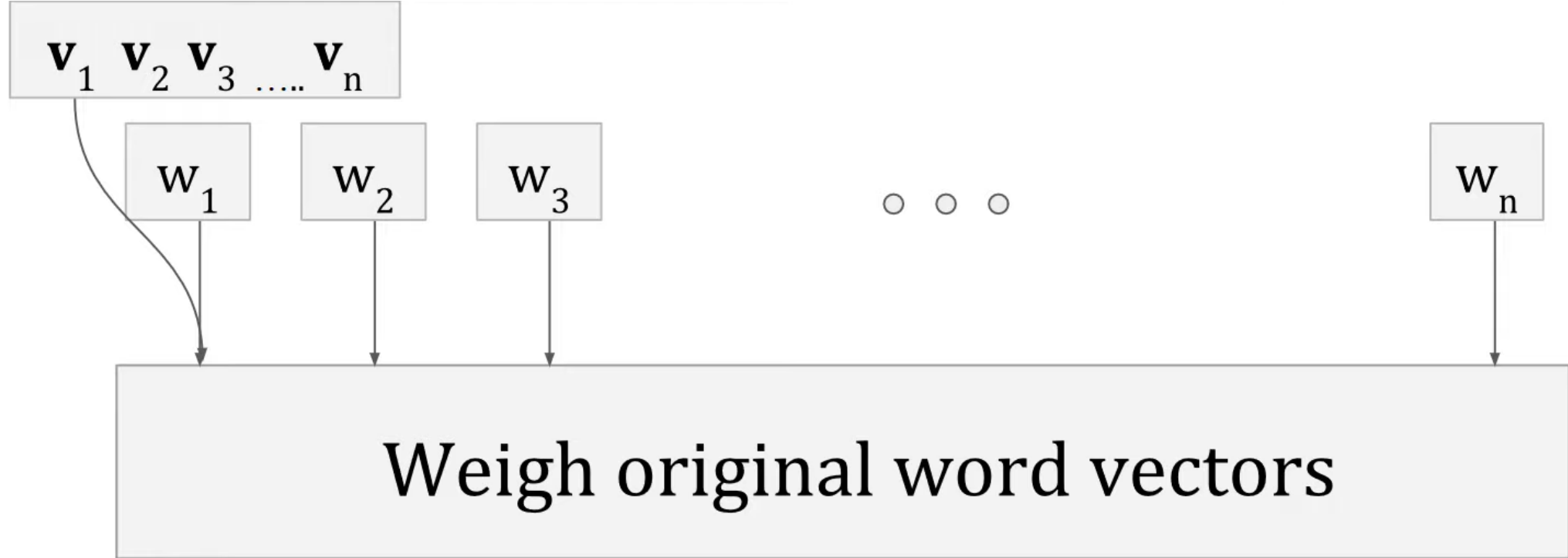
$$S_{n2} = \mathbf{v}_n \mathbf{v}_2^T$$

$$S_{n3} = \mathbf{v}_n \mathbf{v}_3^T$$

$$S_{nn} = \mathbf{v}_n \mathbf{v}_n^T$$







$$\mathbf{y}_1 = w_{11} \mathbf{v}_1 + w_{12} \mathbf{v}_2 + w_{13} \mathbf{v}_3 + \dots + w_{1n} \mathbf{v}_n$$

$$\mathbf{y}_2 = w_{21} \mathbf{v}_1 + w_{22} \mathbf{v}_2 + w_{23} \mathbf{v}_3 + \dots + w_{2n} \mathbf{v}_n$$

$$\mathbf{y}_3 = w_{31} \mathbf{v}_1 + w_{32} \mathbf{v}_2 + w_{33} \mathbf{v}_3 + \dots + w_{3n} \mathbf{v}_n$$

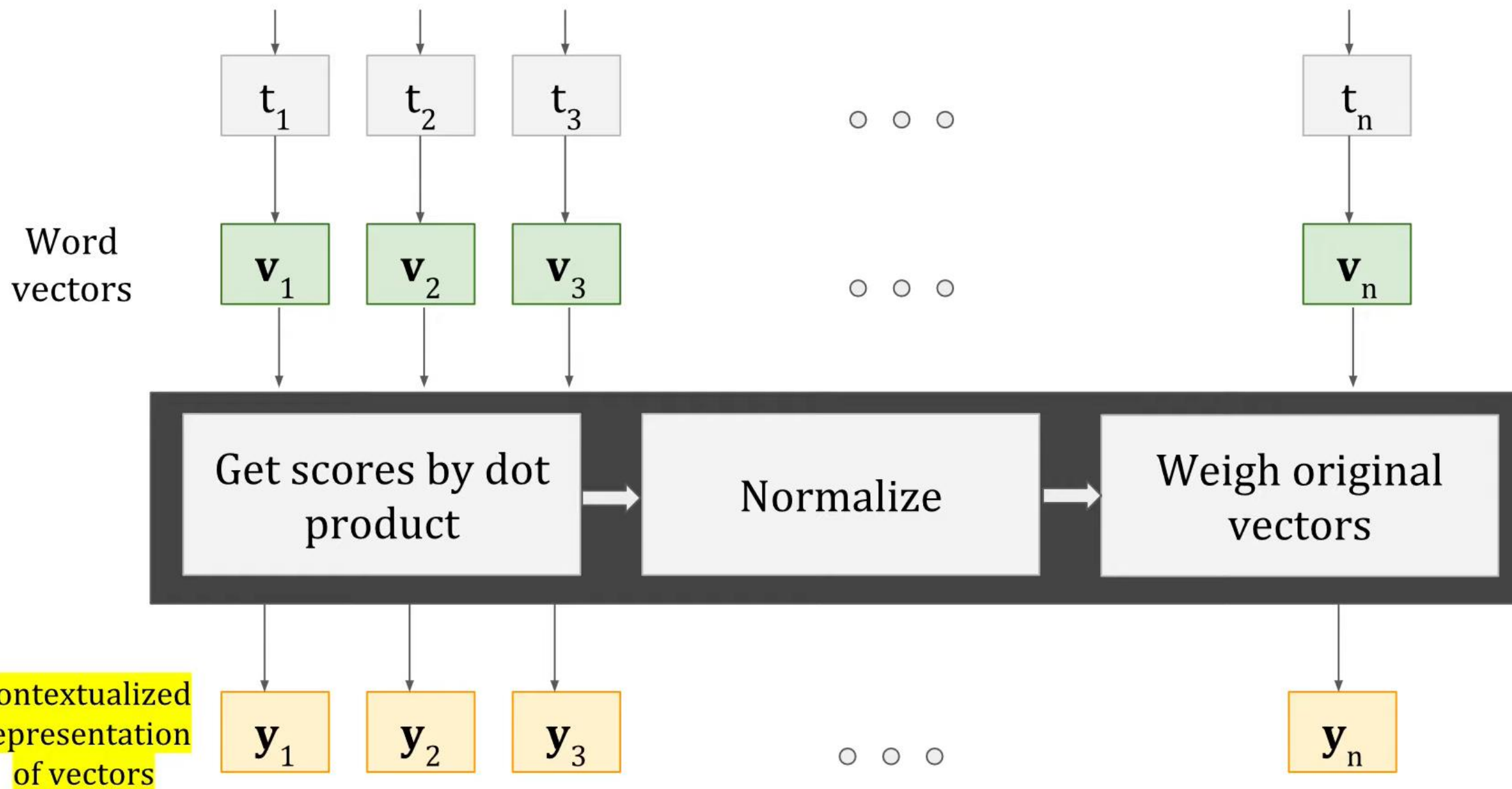
0

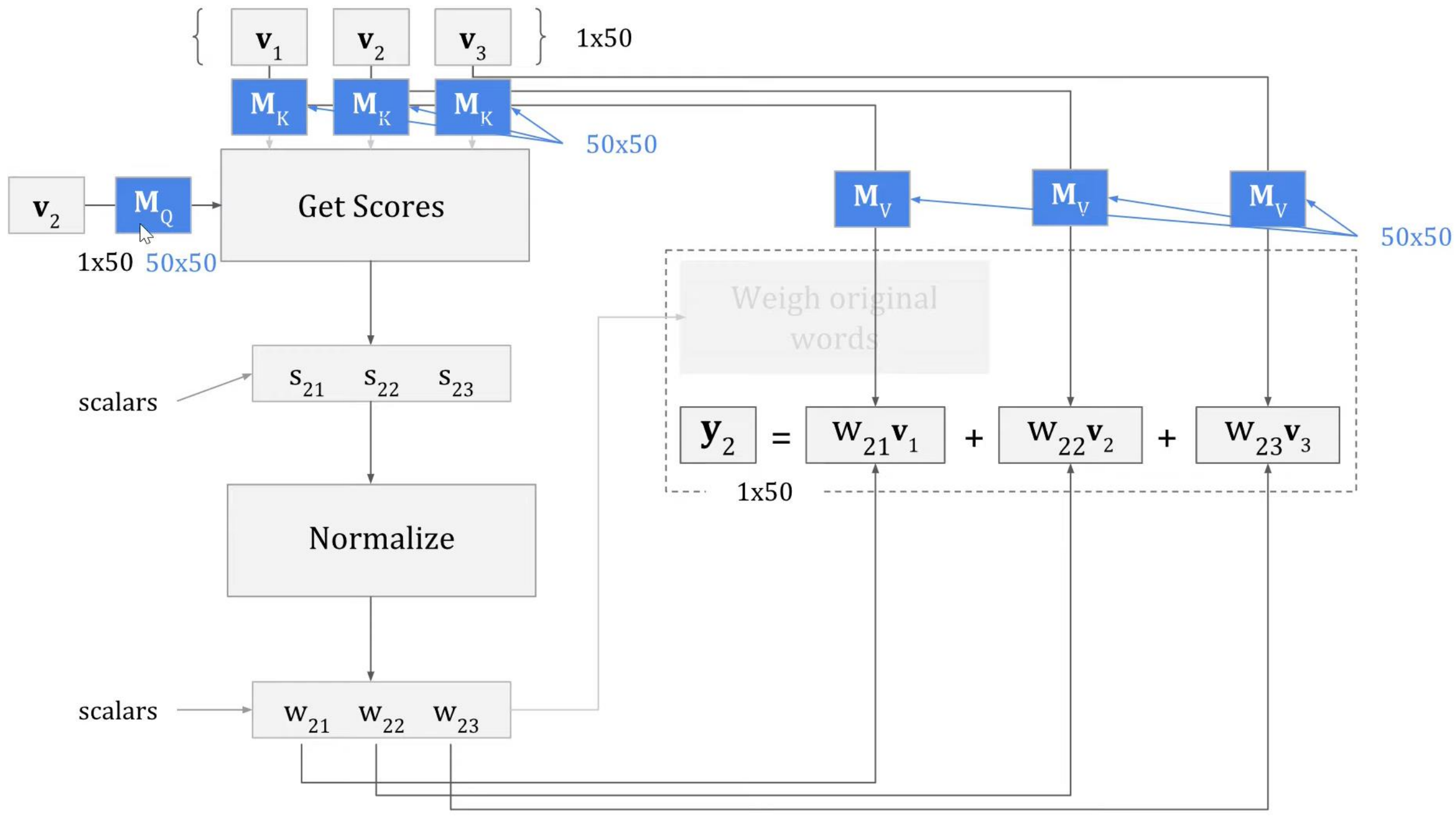
0

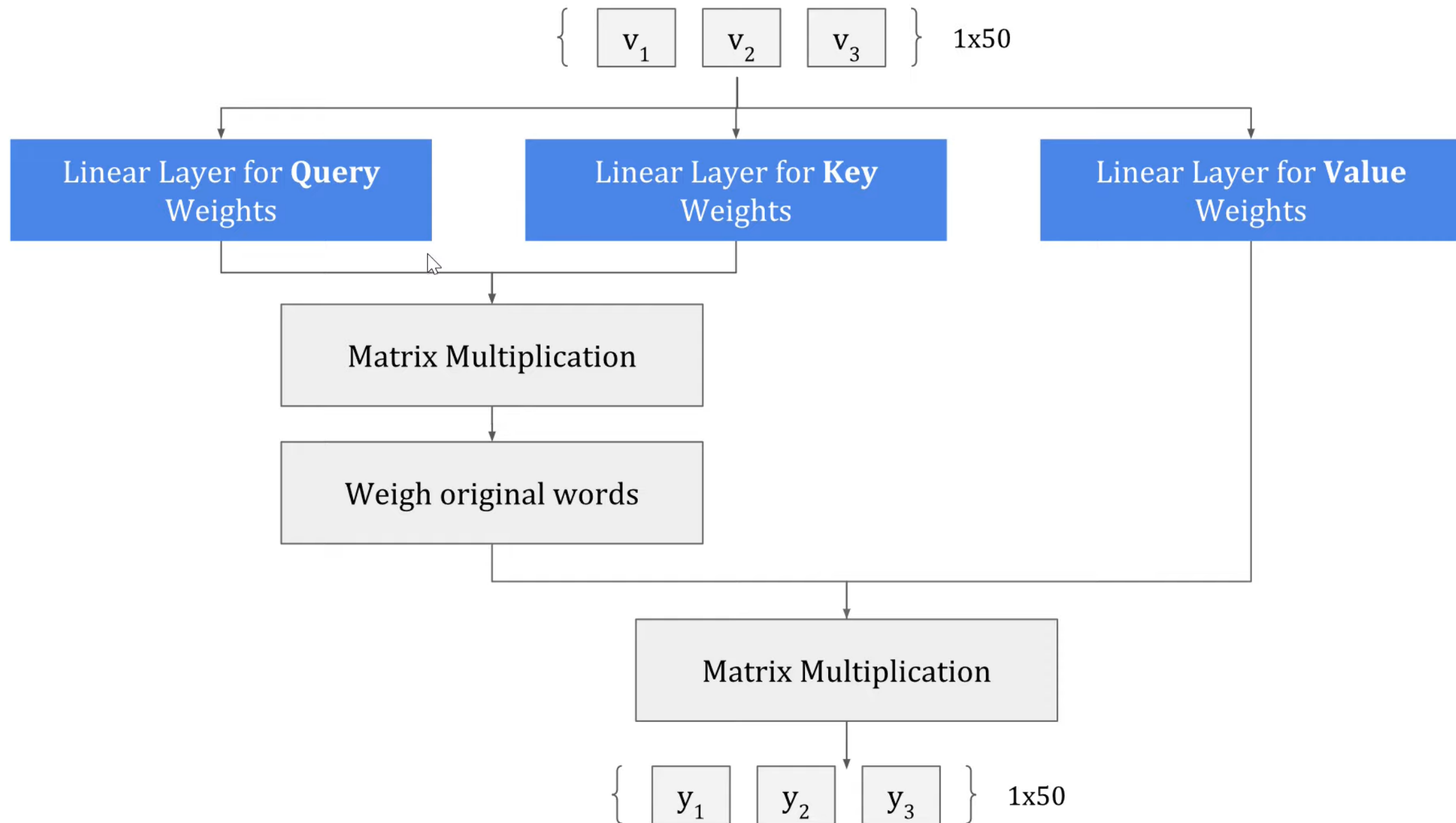
0

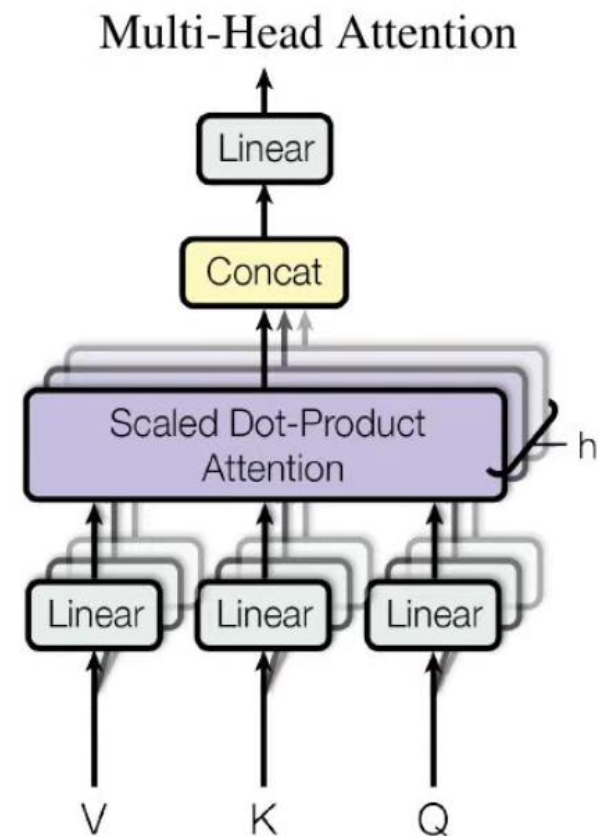
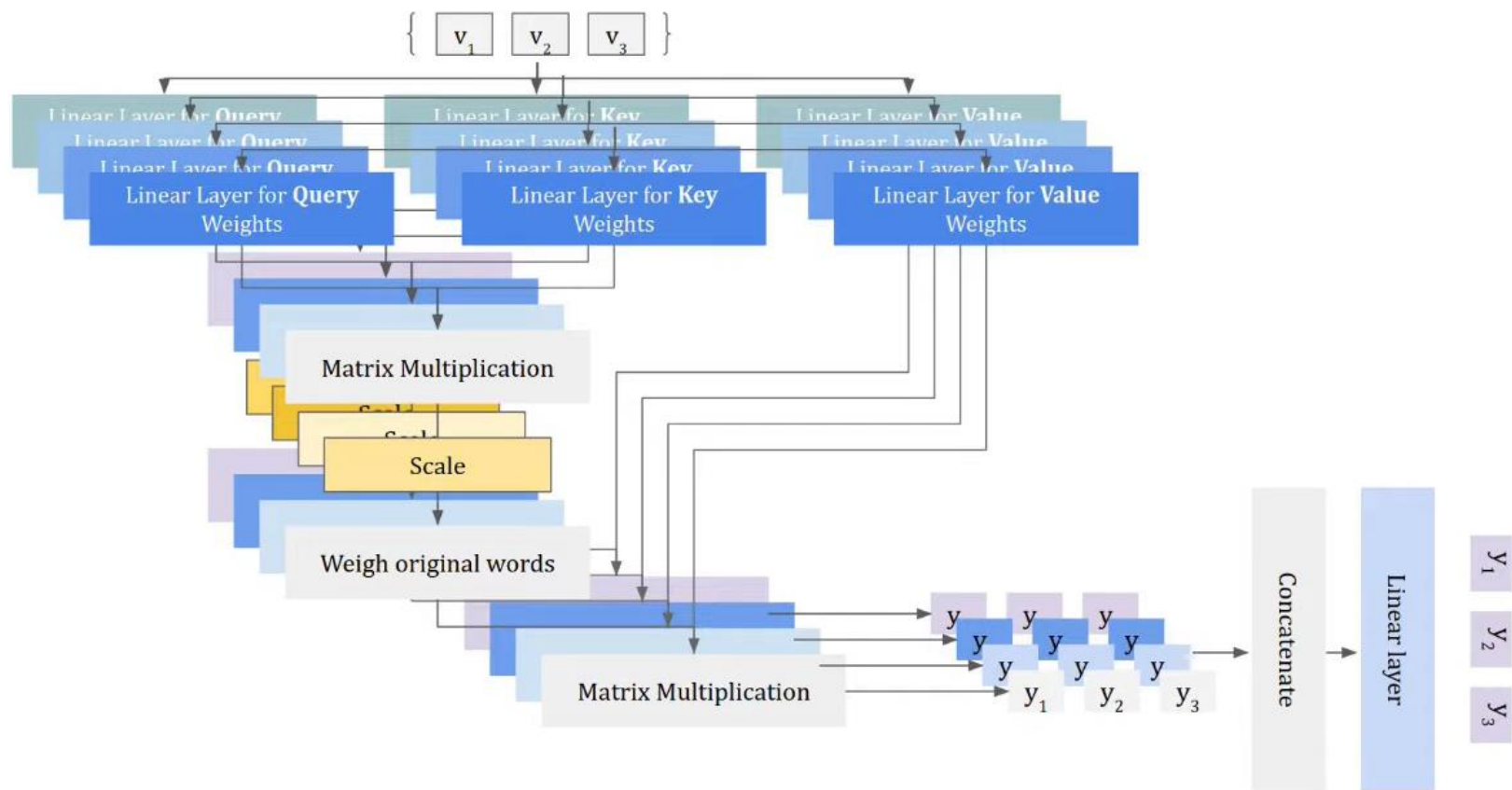
$$\mathbf{y}_n = w_{n1} \mathbf{v}_1 + w_{n2} \mathbf{v}_2 + w_{n3} \mathbf{v}_3 + \dots + w_{nn} \mathbf{v}_n$$

I swam across the river to get to the other bank









Attention is All You Need [Vaswani et al.]