Transformers

Plan

Review

Self-Attention

Cross-Attention

Positional Encoding

Logistics

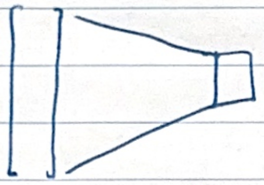
check in

scribe

zoom

Review

Motivation: words as vectors



$f: \mathbb{R}^{|V|} \rightarrow \mathbb{R}^d$

(x, x^+) close $\Rightarrow f(x) \cdot f(x^+)$ large

(x, x^-) far $\Rightarrow |f(x) \cdot f(x^-)|$ small

Contrastive learning $\subseteq$ unsupervised learning

Principal Component Analysis

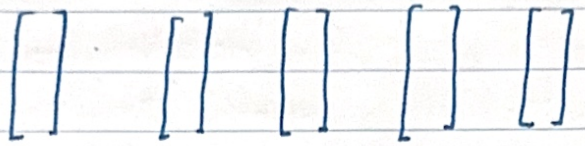
capture variation of ~~eigenvectors~~ data via eigenvectors

$$X^T X = \sum_{i=1}^{r} \lambda_i v^{(i)} v^{(i)T} \quad \text{for} \quad v^{(i)} \cdot v^{(j)} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases}$$

$$\max_{v: \|v\|_2 = 1} \quad \|Xv\|_2^2 \leftarrow v^T X^T X v \leftarrow \sum_{i=1}^{n} \lambda_i [v^T v^{(i)}]^2$$

Motivation: sentences as vectors

"Vermont is chilly and beautiful"



How can we understand sequences of vectors?

↳ Recurrent networks

↳ LSTM

Attention! (self first) $X \in \mathbb{R}^{n \times d}$

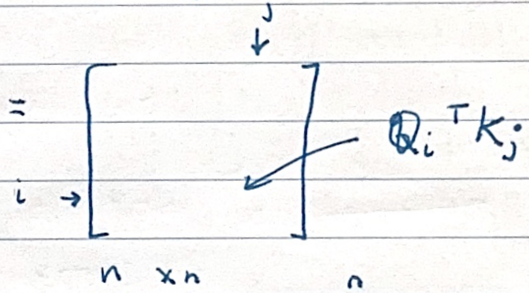
Goal: Combine similar words/tokens

Queries: $W^{(Q)} X = Q$ $W^{(Q)} \in \mathbb{R}^{r \times n}$

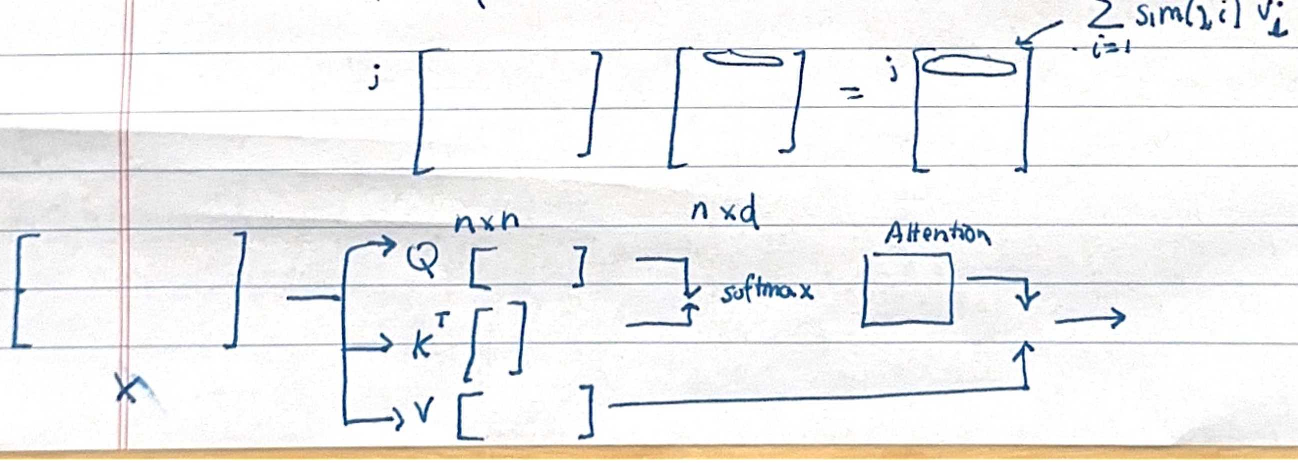
Keys: $W^{(K)} X = K$

Values: $W^{(V)} X = V$ $W^{(V)} \in \mathbb{R}^{d \times n}$

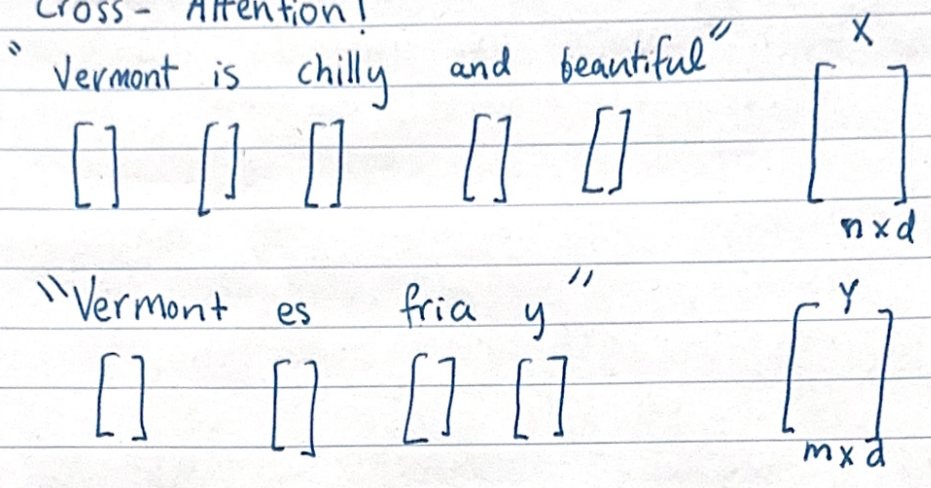
attention : $\underset{\uparrow \text{ to rows}}{\text{softmax}}(Q K^T) =$



result: $\text{softmax}(Q K^T) V$



Cross-Attention!

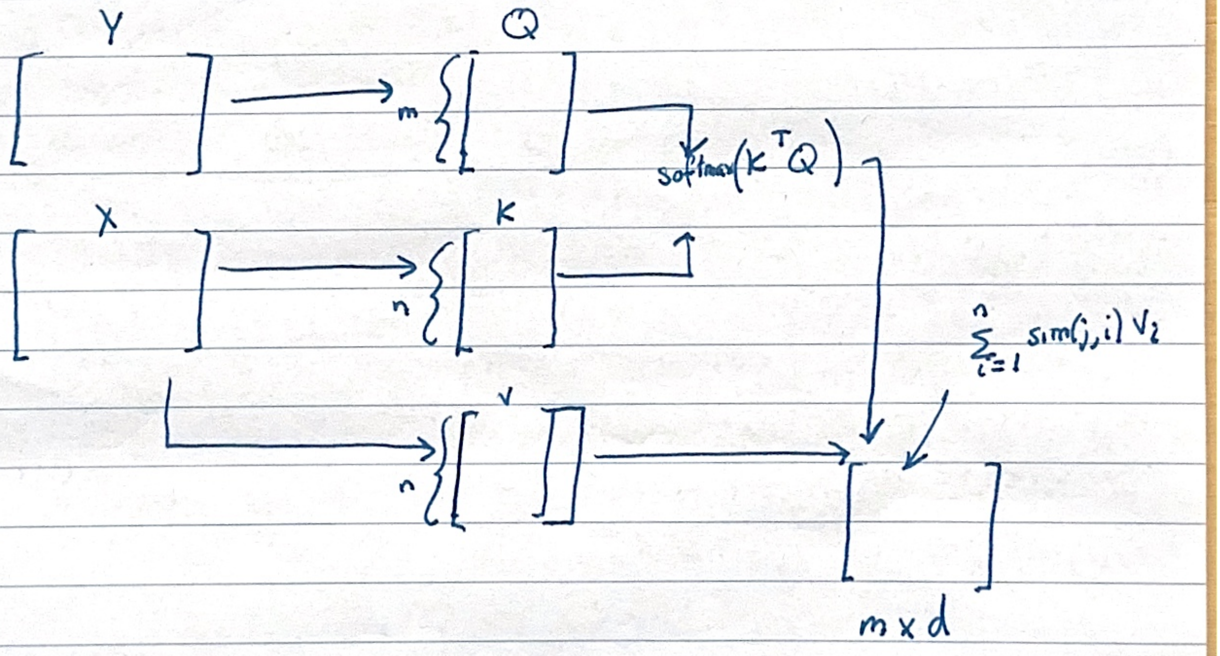


Goal: Represent sequence as linear combo of another

Queries: $W^{(Q)} \cancel{X} = Q$ $\qquad W^{(Q)} \in \mathbb{R}^{r \times m}$

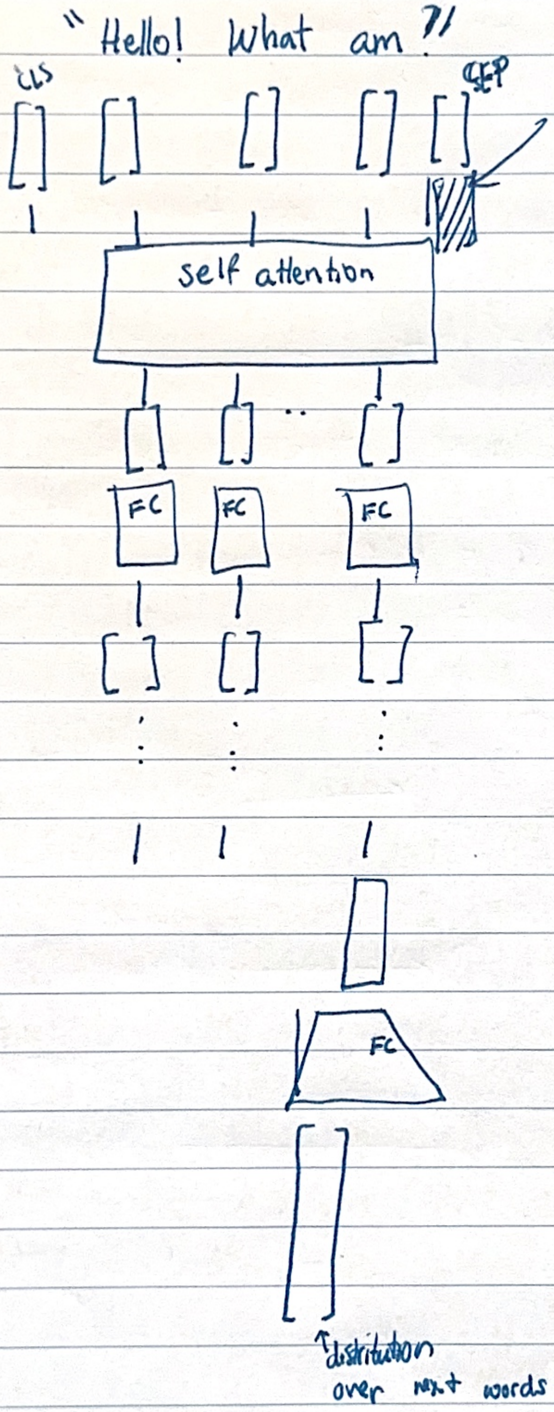
$W^{(K)} X = K$ $\qquad W^{(K)} \in \mathbb{R}^{r \times n}$

$W^{(V)} X = V$ $\qquad W^{(V)} \in \mathbb{R}^{d \times n}$



Make upwards direction?

Large Language Models

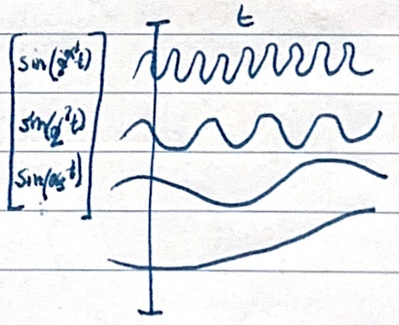


Positional Encoding?

We represent time as
11:24am Tuesday, Jan 14, 2025
rather than
1,065,066,880 min since 0 BC

↳ min captures schedule
↳ hour captures time of day
↳ day captures schedule
↳ date captures schedule
↳ month captures time of year
↳ year captures years passed

"Hello! What am I?"
0 1 2 = t


t
sin(8^{1}t)
sin(2^{?}t)
sin(0.5^{?}t)