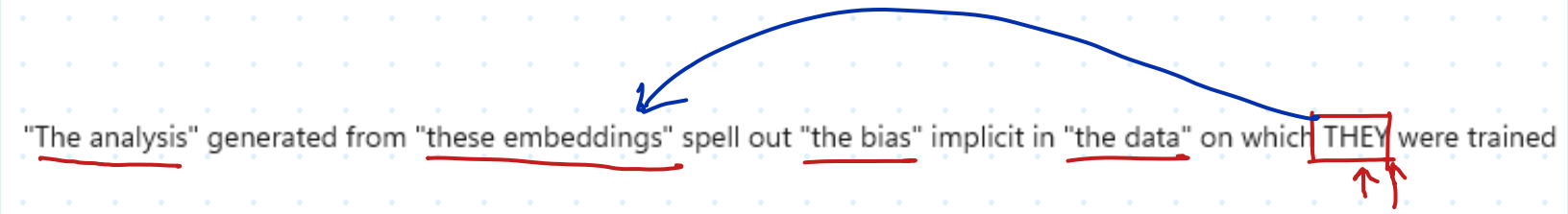
RNN / LSTM



PROBLEM :

→ No context of whats going on in sentence.

⇒ Long Context Dependencies.

2017

"Attention is all you need"

Self Attention

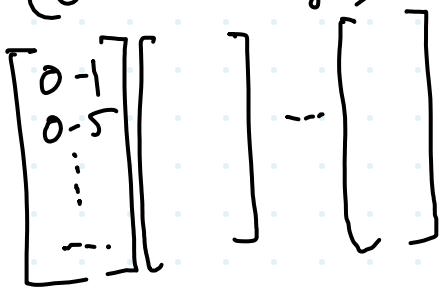
↑

Solution.

Process:

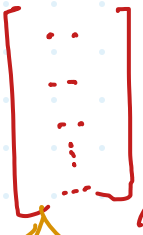
Sequence Input → Tokenize → Vectors (Embeddings)

NORMAL

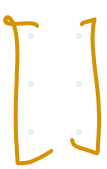


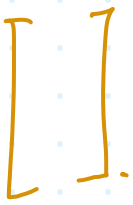
- Sentence
- word
- sub word ★
- character

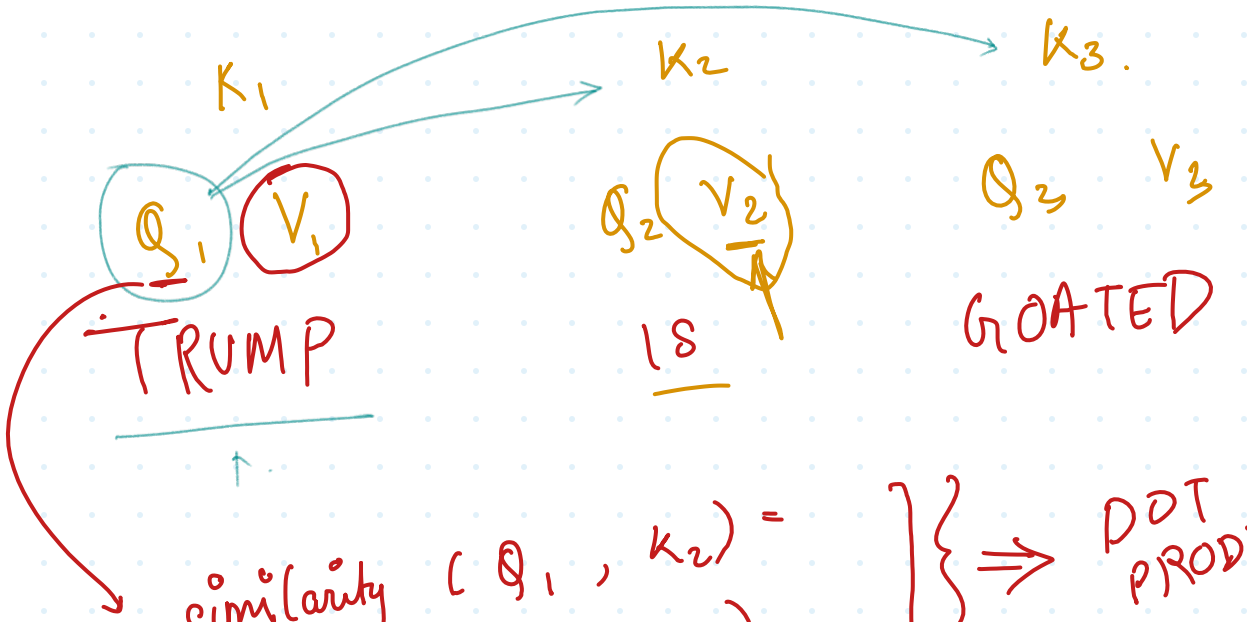
Implementations:

[TRUMP] ⇐ 

small NN. ⇐ Q (Query) K (Key) V (Value).



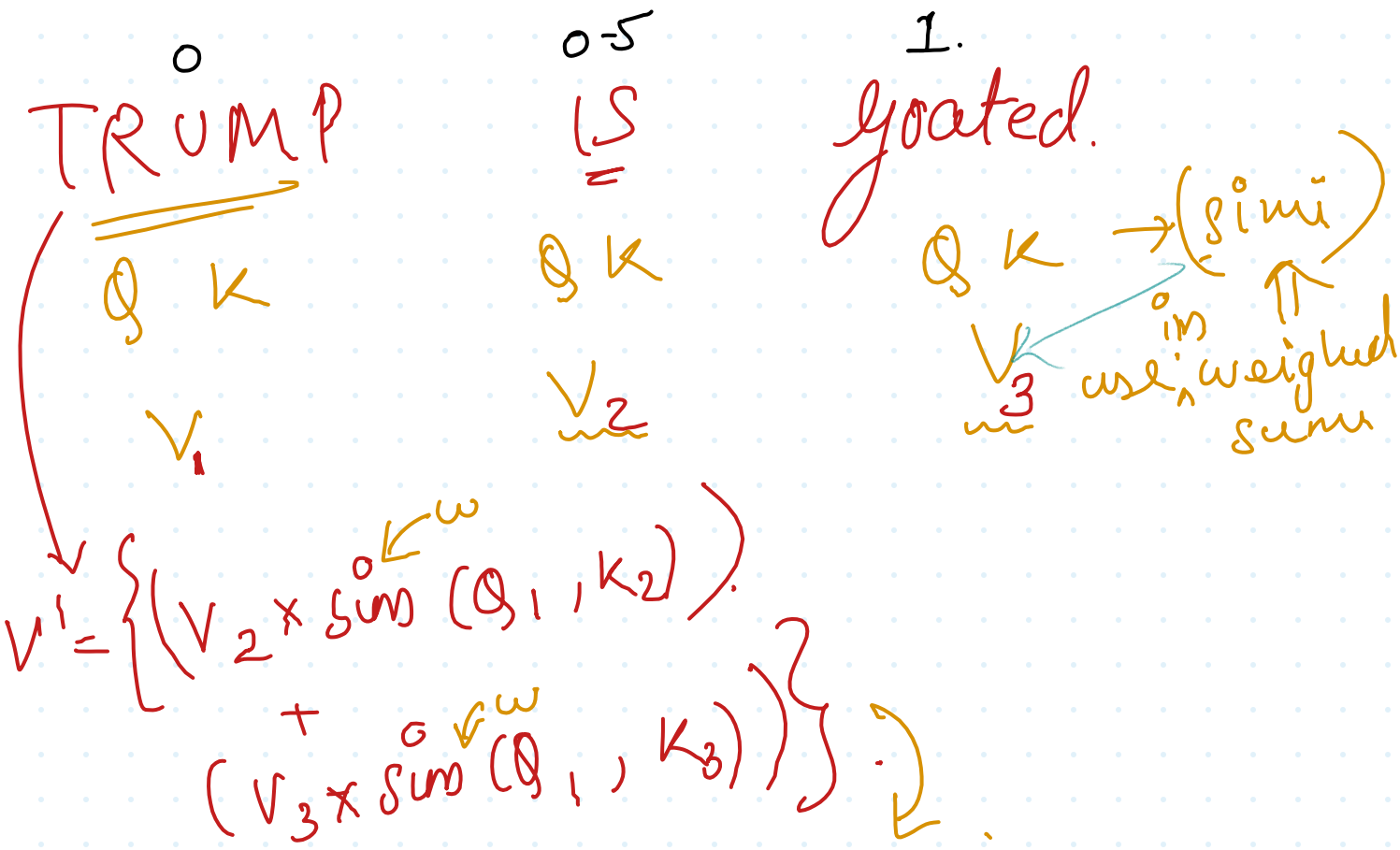




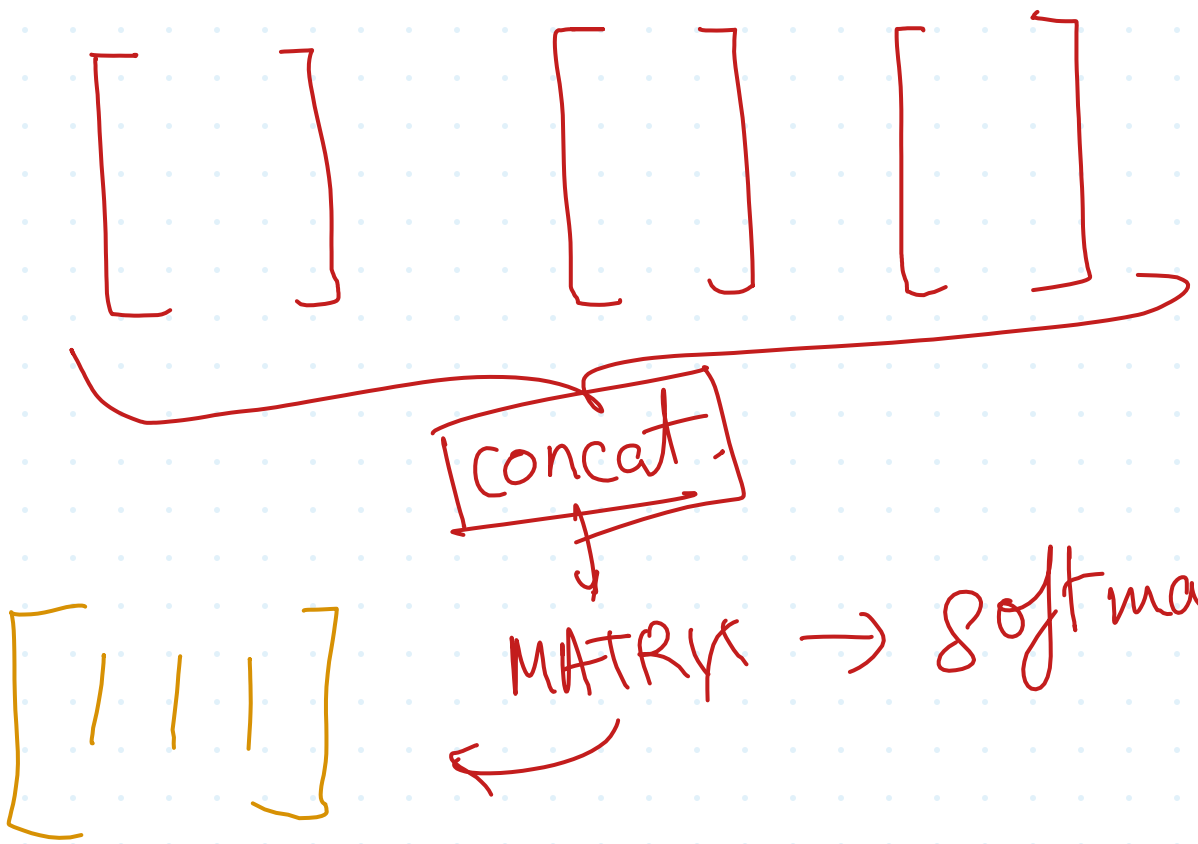
similarity (Q_1, K_2) =
sim (Q_1, K_3) =
⋮
" (Q_2, K_1) =
" (Q_2, K_3) =
" (Q_3, K_1) =
" (Q_3, K_2) =

$\Rightarrow$ DOT PRODUCT

weight




```
newValue += value * similarity(query, key)
```



PROBLEM:

→ Order of the sentence.

↓

Positional Encoding

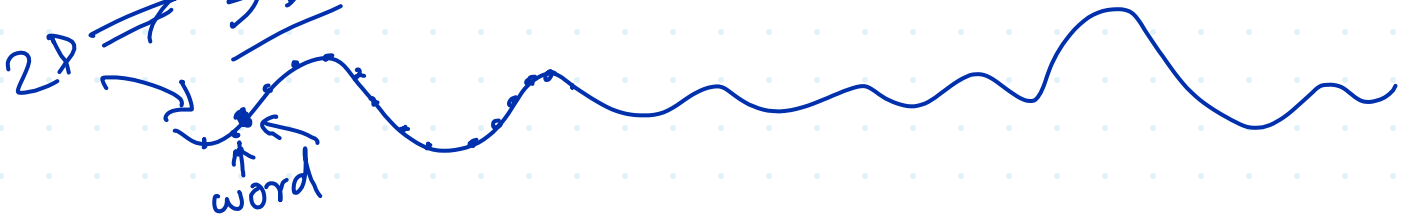
Sin waves.

- Multiple waves
- Diff freq.

→ Merged together.

Ripples in water

2D → 3D → ND.



SUMMARIZE

→ Self Attention:

→ Q K V

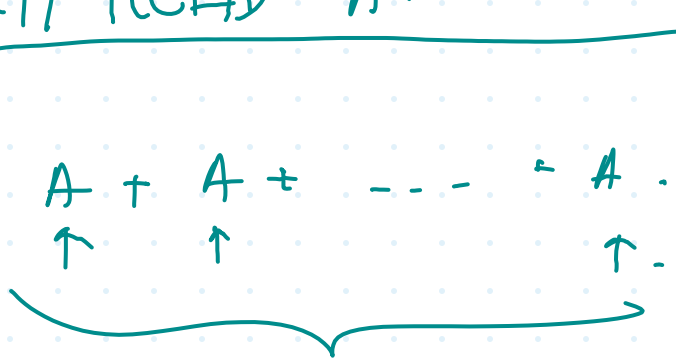
→ Positional Enc.

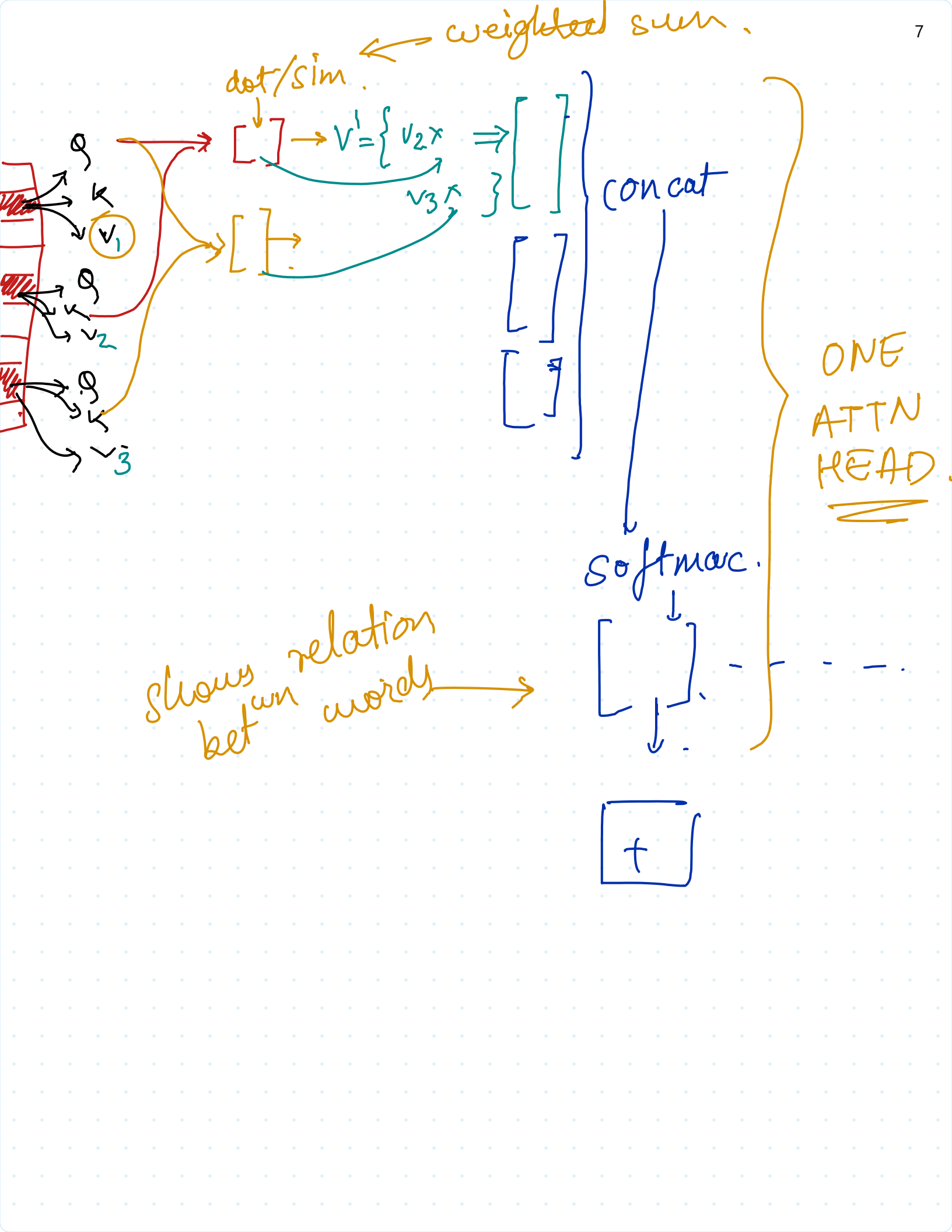
PROBLEM:

→ MULTIPLE PRESPECTIVES.

MULTI HEAD ATTENTION

Sol^a.

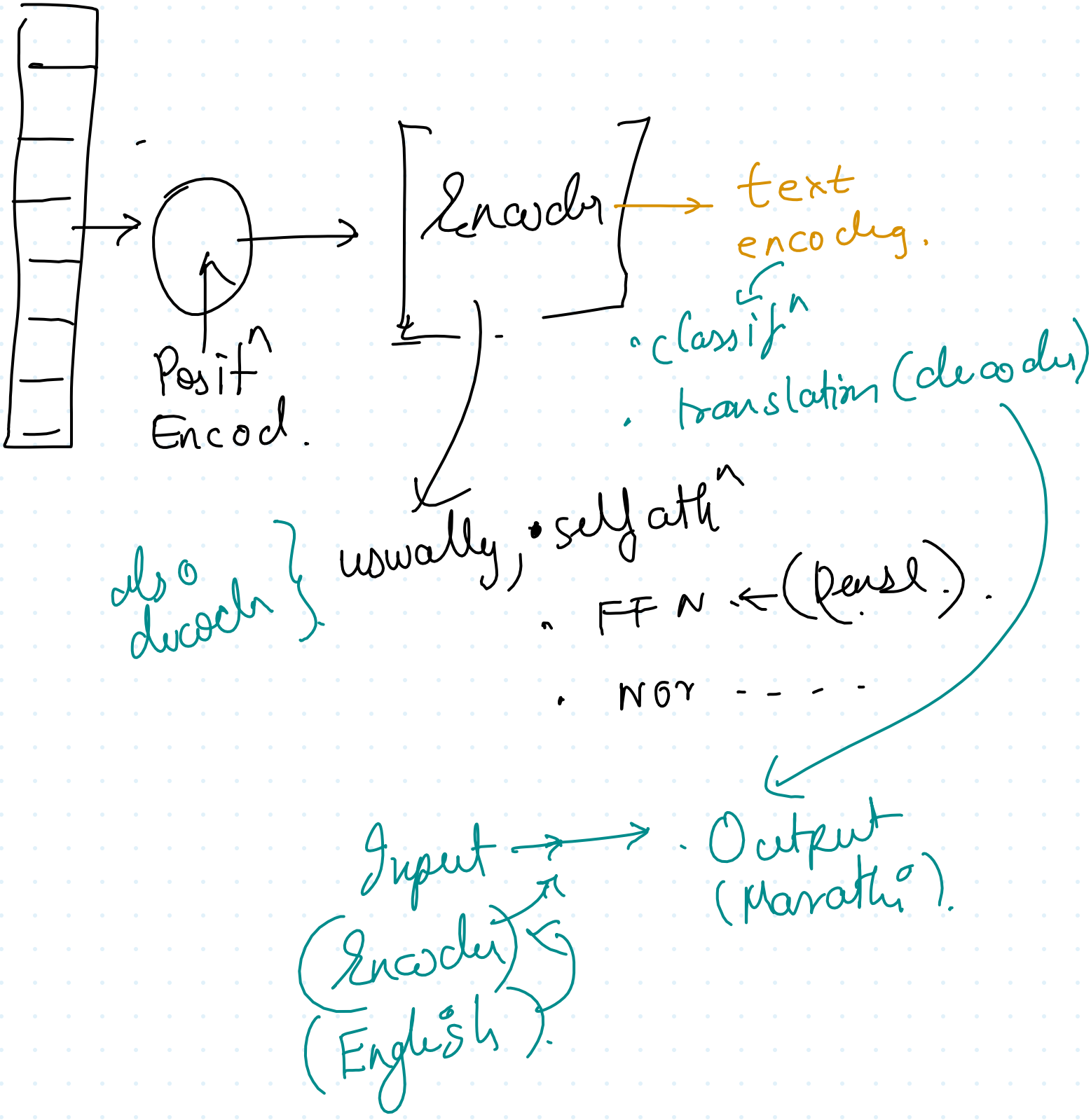



weighted sum.
dot/sim.
v_1
v_2
v_3
concat
softmax.
ONE ATTN HEAD.
shows relation bet un words
t

TRANSFORMER

ENCODER

DECODER



Self Attention

{Cross-} Attention.

"Server, can I have the check?"

"Looks like I just crashed the server"