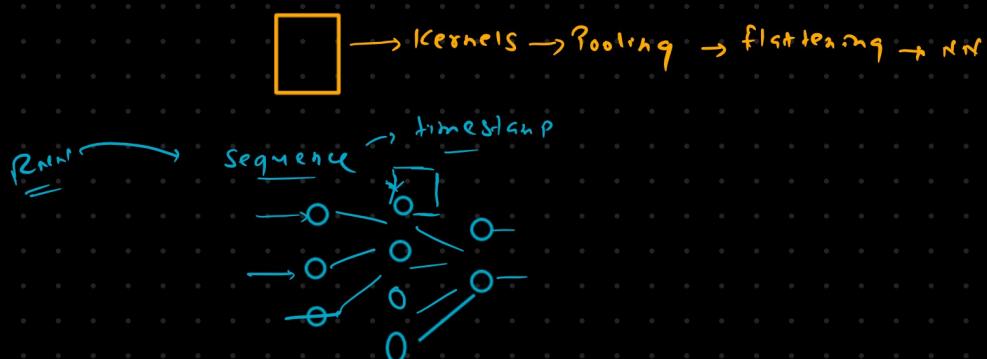
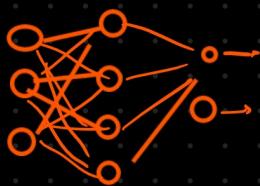


# Transformer

- 1 What is transformer
- 2 Problem with previous
- 3 Attention & Self Attention
- 4 mathematical intuition behind self attention

What is transformer?

- 1 ANN
- 2 CNN
- 3 RNN



(encoder & decoder)

seq-to-seq -

=  
= (translation)

=  
= INP  $\rightarrow$  (1) ML  
= (2) question ans  
= (3) summarization

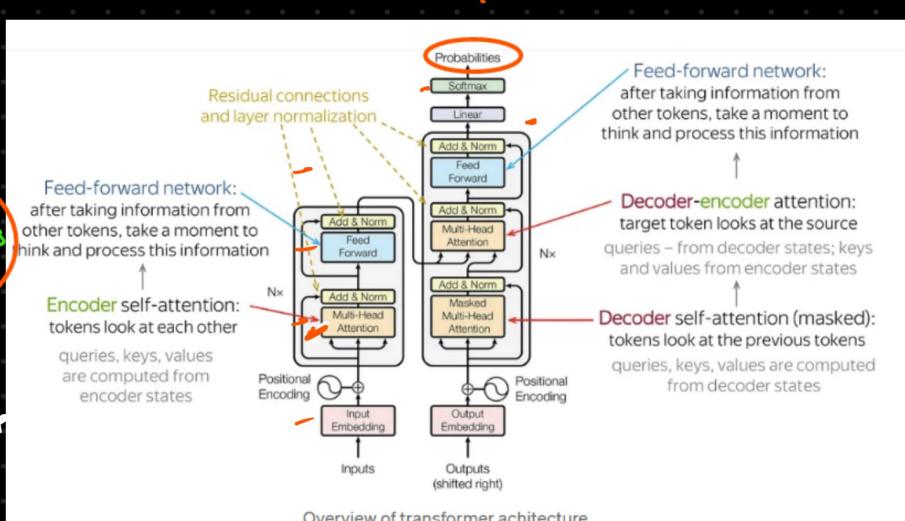
KSTM, RNN, GRU  
(seq) timestamp

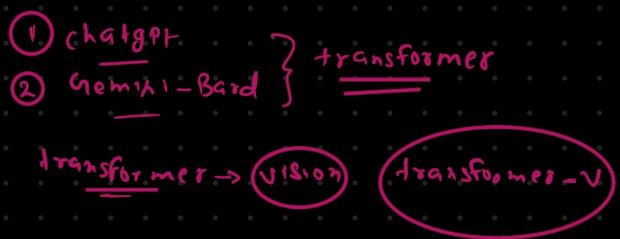
NN (Self attention)

Parallel Processing

## Self Attention

transformer video, image, text





Addition is all you need ⇒ 2017, Google Brain

### Impact of transformer

- ① Revolution in NLP
- ② transfer learning & fine-tuning (huge amt)

timestamp (scratch)  
seq  
LSTM, RNN, GRU

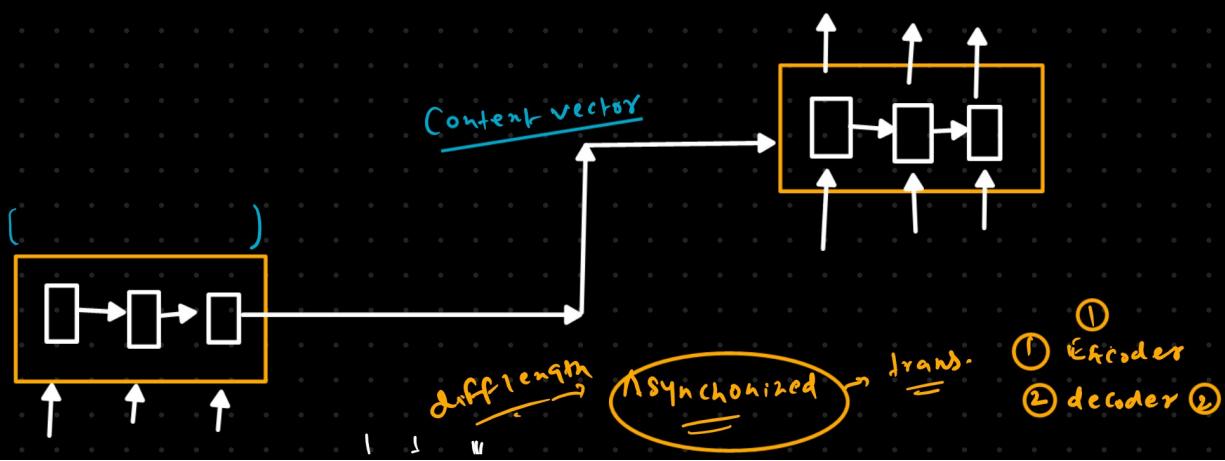
- ③ Multimodal Capability
  - Image → CNN
  - text, audio → RNN
  - Image, text & audio → heterogeneous
  - (video, audio, text)

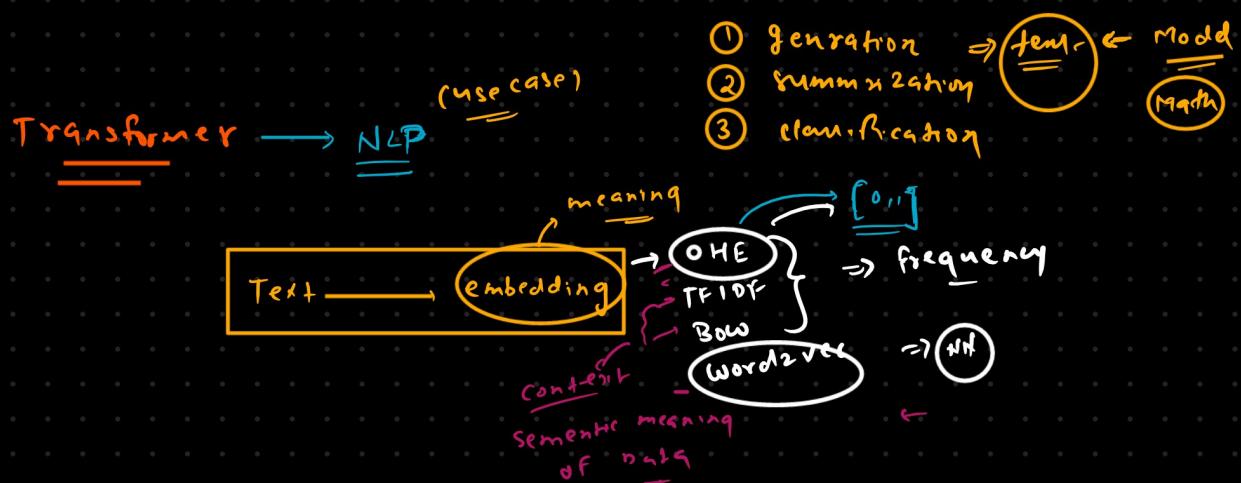
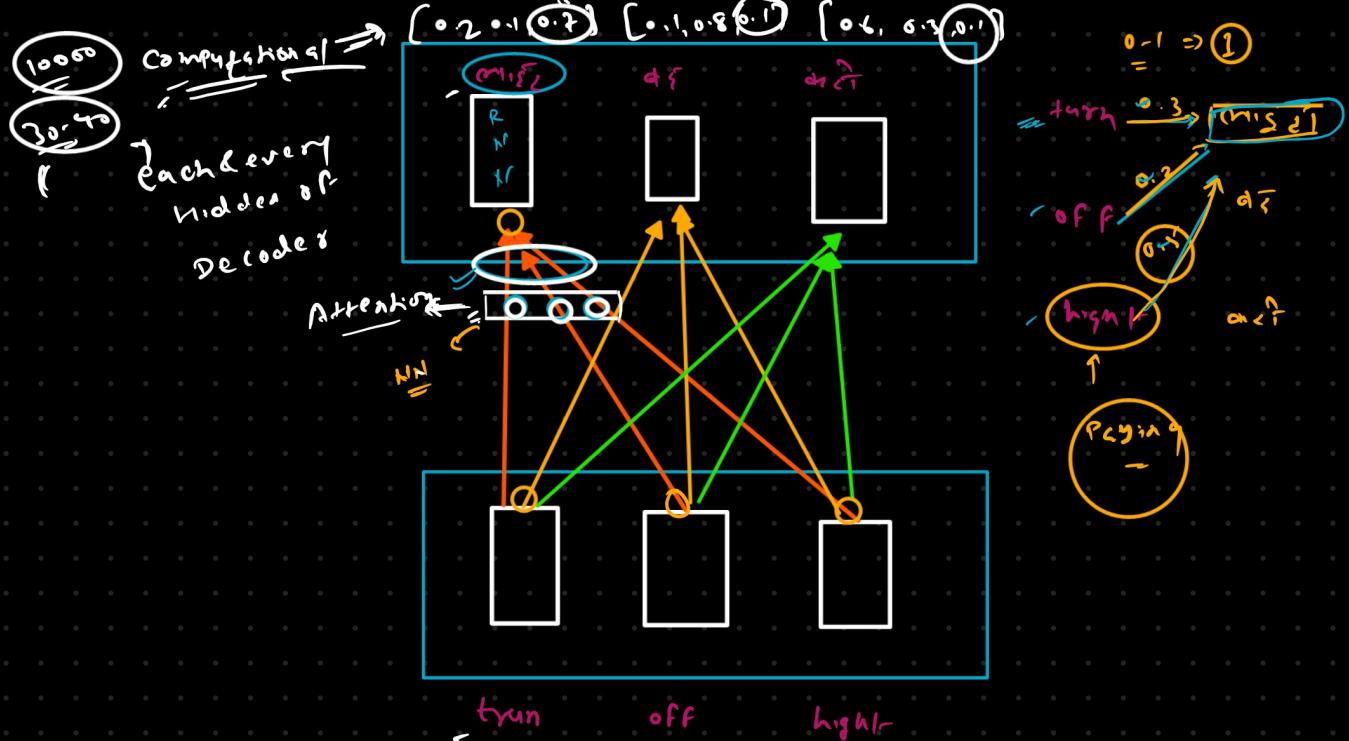
- ④ GenAI (Generate) Image, text, audio

NLP

- 2000-2014 → RNN / LSTM  
 2014 → **encoder-decoder**  
 2017 → **transformer**  
 2018 → BERT / GPT (transfer learning)  
 2021 → GenAI

2022/2023 → ChatGPT / Bard / Stable Diffusion





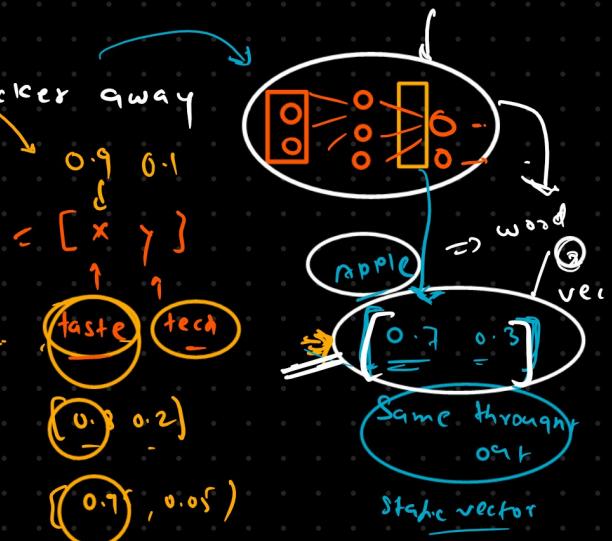
Word2vec  $\rightarrow$  Word to vector



## Problem of word2vec

$\Rightarrow$  AugVec

- 1 An **apple** a day keeps the doctor away
- 2 **Apple** is healthy -
- 3 **Apple** is better than orange
- 4 **Apple** makes great phones



New

Apple Launched a new phone which was good than previous Apple phone

company eating an apple

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone

white was

new

apple phone

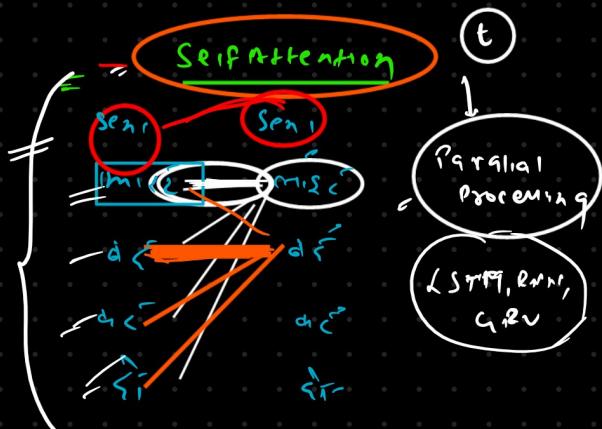
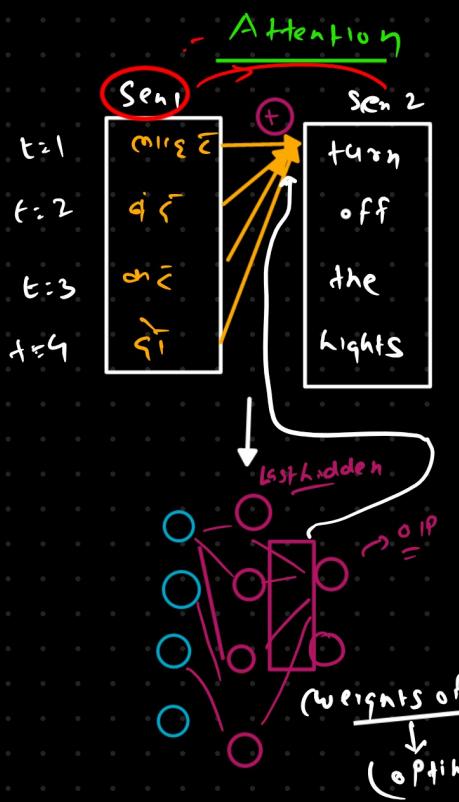
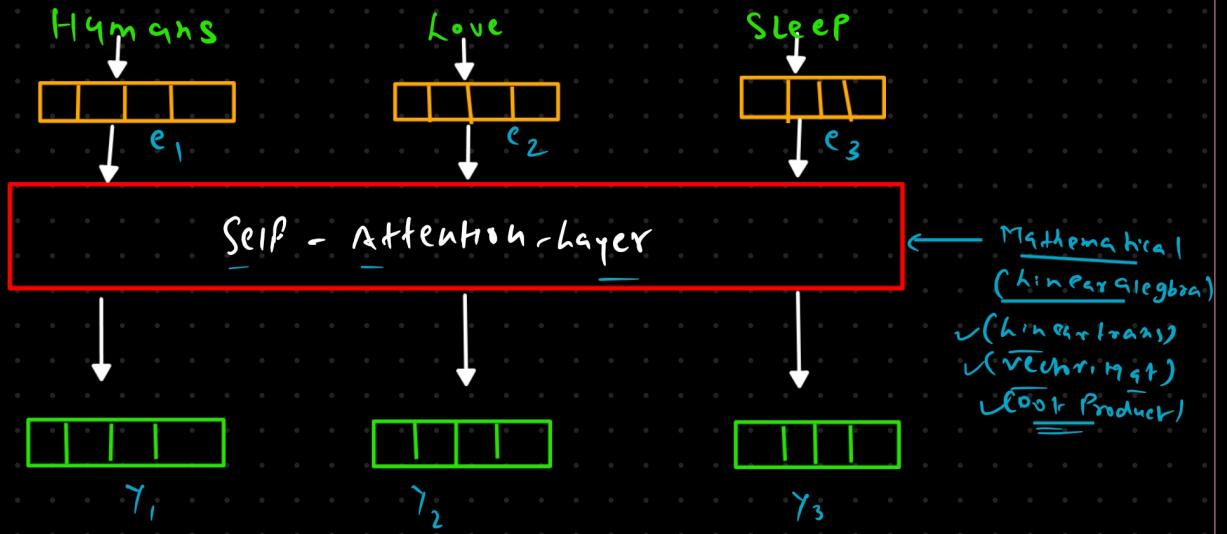
which was good than previous Apple phone

white was

new

apple phone

which was good than previous Apple phone



## Example

Text: "Money bank grows" is split into tokens: Money, bank, grows.

Text: "bank" is mapped to index 1, weight 2.

Text: "Sen. 1" is highlighted.

Equation:  $\text{bank} \rightarrow 0.3 \times \text{Money} + 0.7 \times \text{bank} + 0.1 \times \text{grows}$

Text: "River bank flows" is split into tokens: River, bank, flows.

Text: "Bank" is mapped to index 1, weight 2.

Equation:  $\text{Bank} \rightarrow 0.5 \times \text{River} + 0.4 \times \text{Bank} + 0.1 \times \text{flows}$

$$\begin{aligned}
 & \text{Money} \rightarrow 0.7 \text{money} + 0.2 \text{bank} + 0.1 \text{grows} \\
 & \text{bank} \rightarrow 0.25 \text{money} + 0.2 \text{bank} + 0.05 \text{grows} \\
 & \text{grows} \rightarrow 0.1 \text{money} + 0.2 \text{bank} + 0.7 \text{grows}
 \end{aligned}$$

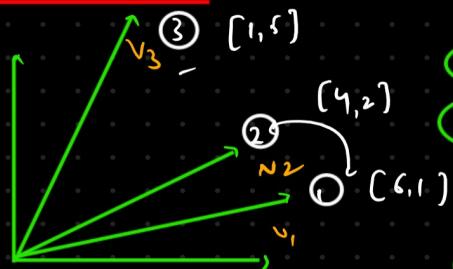
$$\begin{cases}
 \text{River} \rightarrow 0.8 \text{River} + 0.15 \text{bank} + 0.05 \text{flows} \\
 \text{bank} \rightarrow 0.2 \text{River} + 0.25 \text{Bank} + 0.02 \text{flows} \\
 \text{flows} \rightarrow 0.9 \text{River} + 0.01 \text{Bank} + 0.55 \text{flows}
 \end{cases}$$

Self Attention

- Key, Query, Value
- $W_1$  embed

Similarity  $\Rightarrow$  Dot Product

Dot



$$\begin{aligned}
 v_1 \cdot v_2 &= [6, 1] \cdot [4, 2] = 6 \times 4 + 1 \times 2 = 26 \\
 v_1 \cdot v_3 &= [6, 1] \cdot [1, 5] = 6 \times 1 + 1 \times 5 = 11
 \end{aligned}$$

$$v_1 \text{ dot } v_2 > v_1 \text{ dot } v_3$$

$$\begin{aligned}
 ? &= [x, 0], [0, y] \\
 x \times 0 + y \times 0 &= 0
 \end{aligned}$$

$$w * \text{money} + w * \text{bank} + w * \text{grows}$$

$$e_{\text{bank}} = [e_{\text{bank}} \cdot e_{\text{money}}] * \text{money} + [e_{\text{bank}} \cdot e_{\text{bank}}] * \text{bank} + [e_{\text{bank}} \cdot e_{\text{grows}}] * \text{grows}$$

Similarity score

$$\begin{array}{c}
 \boxed{\quad} \quad \boxed{\quad} \\
 + \quad \boxed{\quad} \quad + \quad \boxed{\quad} \quad \boxed{\quad} \quad \times \quad \boxed{\quad} \\
 \end{array}$$

$e_{\text{money}}$

$s_{21}$

$e_{\text{bank}}$

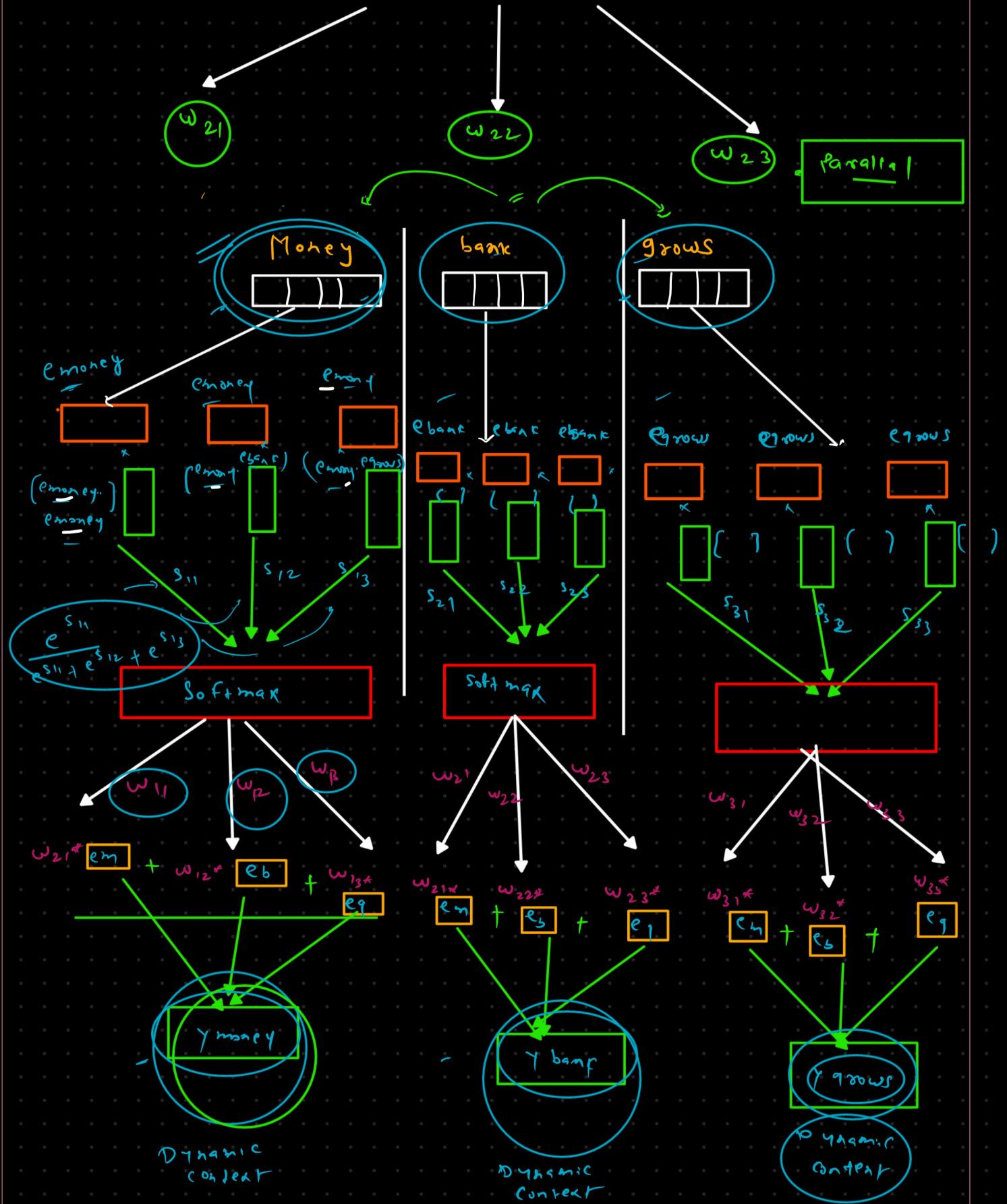
$s_{22}$

$e_{\text{grows}}$

$s_{23}$

$$w_{22} = \frac{e^{s_{22}}}{e^{s_{21}} + e^{s_{22}} + e^{s_{23}}}$$

SoftMax Layer



key, value, query)

Bonus  $\Rightarrow$  transform  $\rightarrow$  interviews (genAI)

## Points

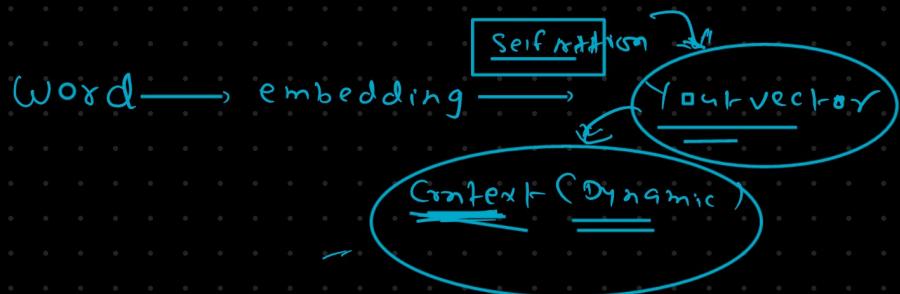
1

parallel operation

best

2

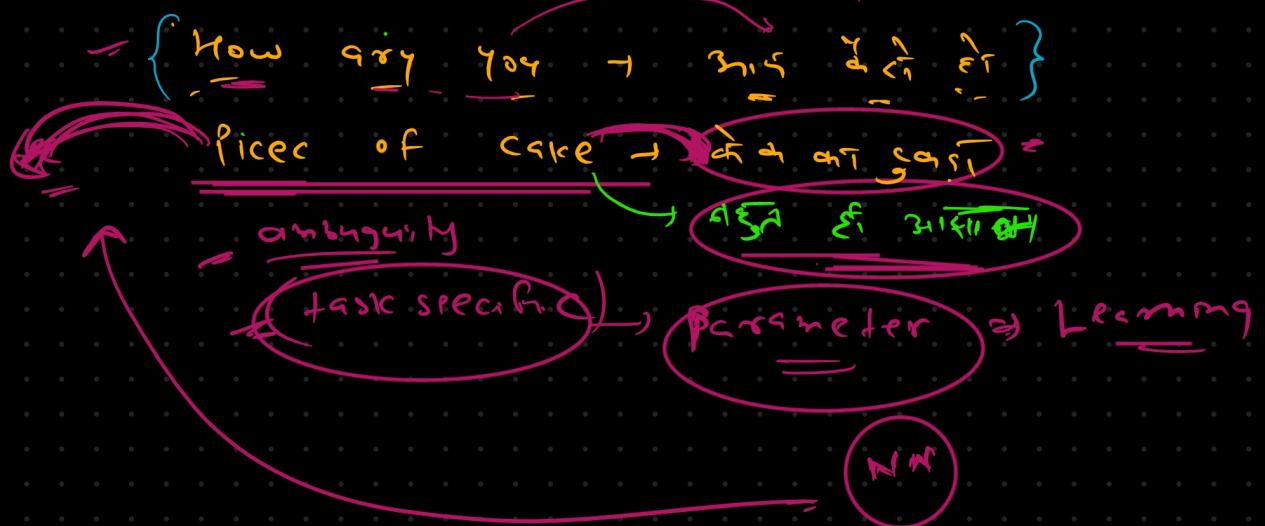
there are no parameters  $\Rightarrow$  learning parameter



how are you  
 $\downarrow$        $\downarrow$        $\downarrow$   
 $e_1$      $e_2$      $e_3$

Self Attention

(general contextual Embedding)



## Compressor

## (SelfAttention)

word |

l'money
e bank
e grows

$3 \times n$



Dot

S <sub>11</sub>	S <sub>12</sub>	S <sub>13</sub>
S <sub>21</sub>	S <sub>22</sub>	S <sub>23</sub>
S <sub>31</sub>	S <sub>32</sub>	S <sub>33</sub>

$3 \times 3$

Softmax

w <sub>11</sub>	w <sub>12</sub>	w <sub>13</sub>
w <sub>21</sub>	w <sub>22</sub>	w <sub>23</sub>
w <sub>31</sub>	w <sub>32</sub>	w <sub>33</sub>

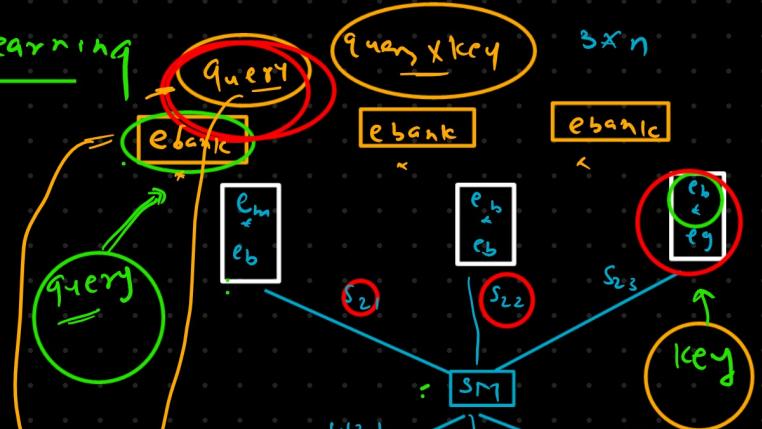
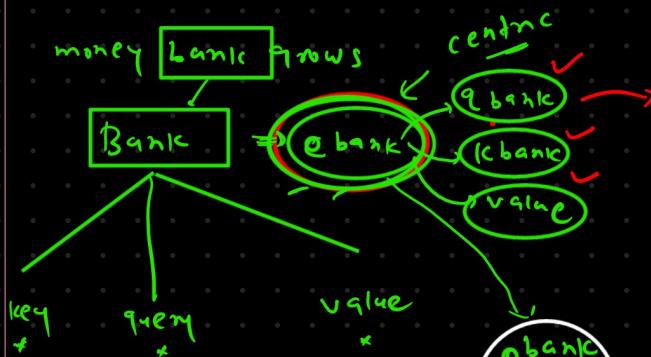
$3 \times 3$

Dot

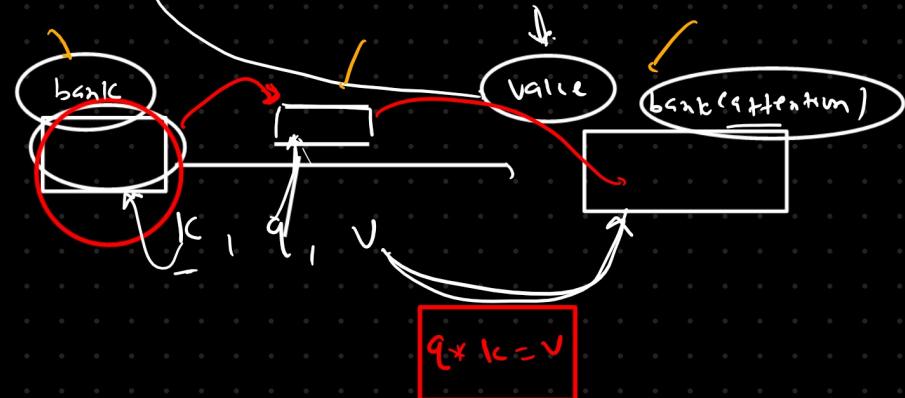
Y'money
e bank
e grows

$3 \times n$

No Learning



$$D = \{a:1, b:2\}$$

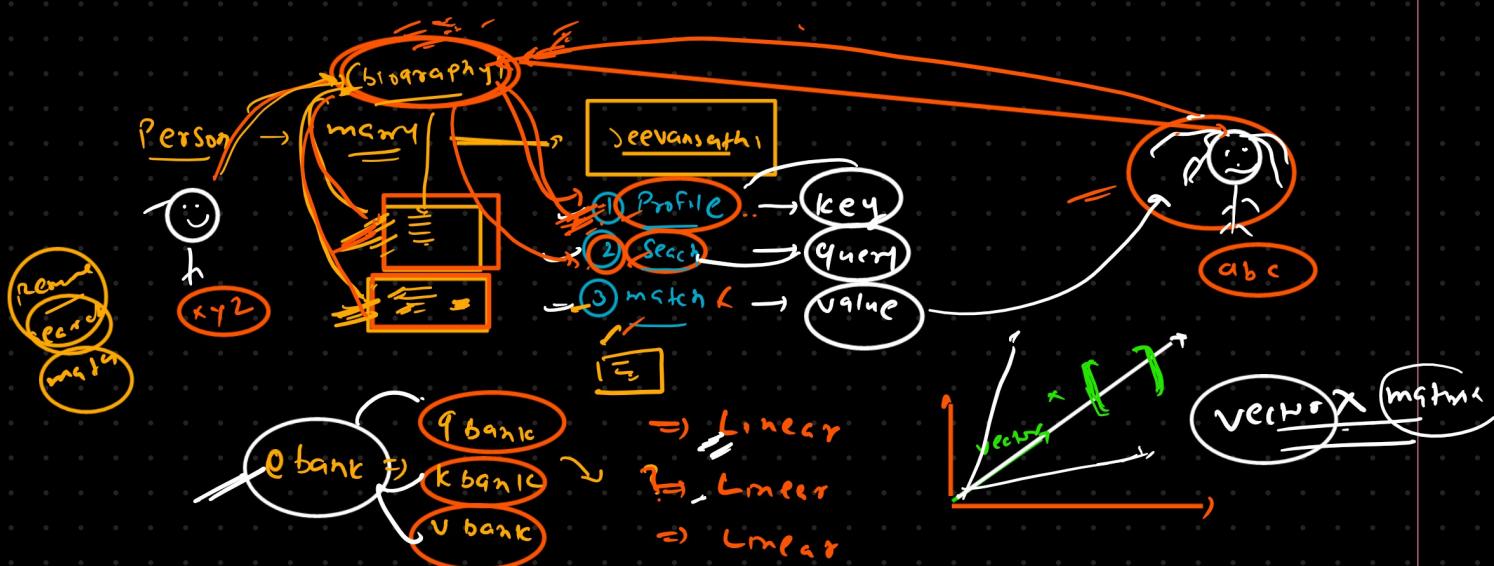
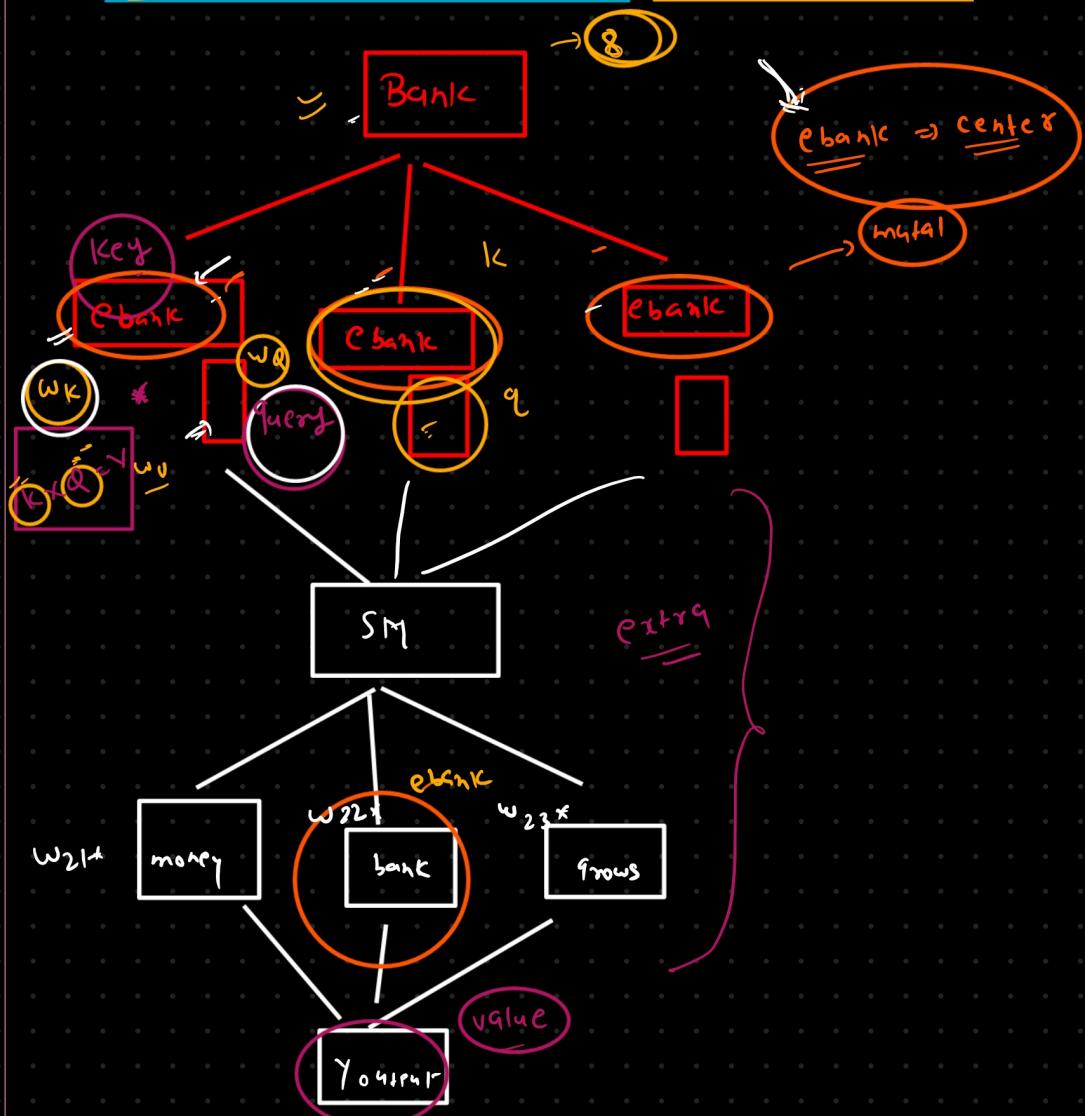


$$Q \times K = V$$

query  $\times$  key  $\approx$  value

$$\begin{aligned} & \text{key} \rightarrow D(q) \\ & \text{query} \rightarrow V \\ & \text{Value} \end{aligned}$$

## Linear transformation ( Multhead attention )



link of the Blog Must read : <https://jalammar.github.io/illustrated-transformer/>

Create a summary of this blog and research paper then post it over the linkedin