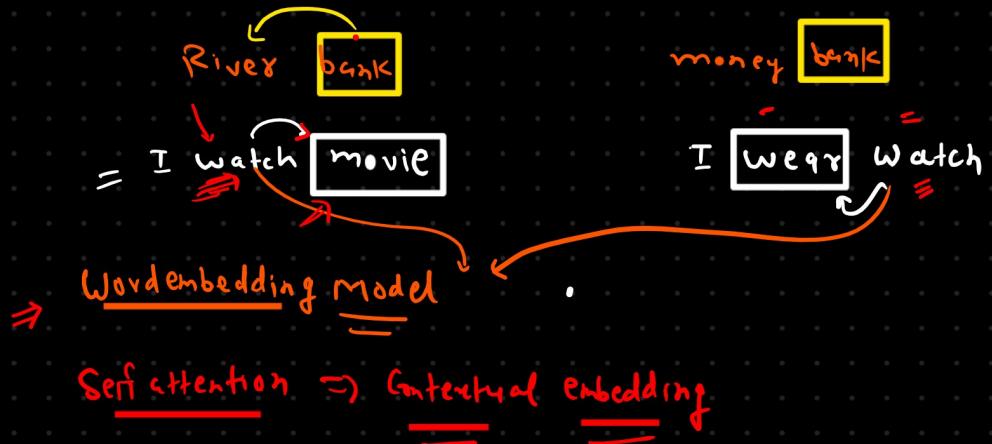


NLP →



- ① An apple a day keeps the doctor away
  - ② Apple is healthy
  - ③ Apple better than the orange
  - ④ Apple make great phone
- Word2vec → [0.2 0.8] ← Apple

new sent → Apple launch new phone while I was eating apple  
→ [0.2, 0.8] dynamic embedding → [0.2, 0.8] X

money bank

River bank

many bank grow

River bank flow

$$\boxed{\text{bank}} = w_1 \text{money} + w_2 \text{bank} + w_3 \text{grow}$$

$$\boxed{\text{bank}} = w_1 \text{River} + w_2 \text{bank} + w_3 \text{flow}$$

Money, bank, value



$S_1$

$(\sum w_i)_{\text{inv}} = 1$  (Random weight)

$$\begin{aligned} \boxed{\text{Money}} &= 0.7 \text{money} + 0.2 \text{bank} + 0.1 \text{grow} \\ \boxed{\text{bank}} &= 0.25 \text{money} + 0.7 \text{bank} + 0.05 \text{grow} \\ \boxed{\text{grow}} &= 0.1 \text{money} + 0.2 \text{bank} + 0.7 \text{grow} \end{aligned}$$

$S_2$

$$\left. \begin{aligned} \text{River} &= 0.8 \text{River} + 0.1 \text{bank} + 0.05 \text{flow} \\ \text{bank} &= 0.25 \text{River} + 0.78 \text{bank} + 0.02 \text{flow} \\ \text{flow} &= 0.4 \text{River} + 0.01 \text{bank} + 0.02 \text{flow} \end{aligned} \right\}$$



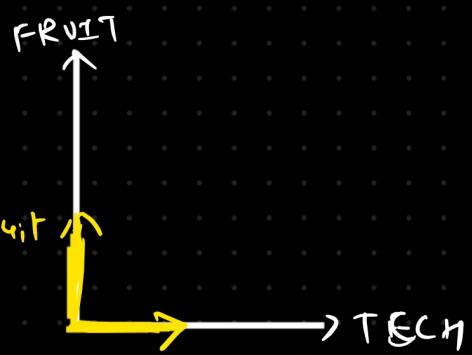
$$\boxed{\text{Tea}} = \frac{0.5}{\text{milk}} \quad \frac{0.2}{\text{Tea leaves}} \quad \frac{0.2}{\text{Sugar}} \quad \frac{0.1}{\text{Ginger}}$$
$$= \frac{\text{half L}}{1 \text{ spoon}} \quad \Rightarrow \quad \frac{1 \text{ spoon}}{1 \text{ spoon}} \quad \frac{1 \text{ spoon}}{20 \text{ gm}}$$

Contextual embedding

$$\boxed{\text{bank}} = w_1 e_{\text{money}} + w_2 e_{\text{bank}} + w_3 e_{\text{flow}} = 1$$
$$\begin{aligned} w_1 &= 0.25 \\ w_2 &= 0.7 \\ w_3 &= 0.05 \end{aligned}$$
$$\begin{aligned} (e_{\text{money}}) &= (\text{Similarity score}) \\ (e_{\text{bank}}) &= (\text{Similarity score}) \\ (e_{\text{flow}}) &= (\text{Similarity score}) \end{aligned}$$

$$\underline{e_{\text{bank}}} = [e_{\text{bank}} \cdot e_{\text{money}}] e_{\text{money}} + [e_{\text{bank}} \cdot e_{\text{bank}}] e_{\text{bank}} + [e_{\text{bank}} \cdot e_{\text{flow}}] e_{\text{flow}}$$

$$\begin{array}{c}
 \text{[TECH]} \quad \text{[FRUIT]} \\
 \text{PHONEAPPLE} \quad | \quad \bullet \leftarrow [1,0] \\
 \text{FRUITAPPLE} \quad 0 \quad | \quad 1 \leftarrow [0,1]
 \end{array}$$



## Similarity

$$\begin{array}{l} \text{Phone} \rightarrow [1, 0] \\ \text{Fruit} \rightarrow [0, 1] \end{array} \quad ] \quad \underline{\text{dat.-Produkt}}$$

$$\begin{array}{l}
 \text{Sc} \quad \left( [1, 0] \cdot [0, 1] \right) = \quad 1 \times 0 + 0 \times 1 \\
 \Rightarrow \quad \quad \quad \quad \quad = \quad \quad \quad 0 + 0 \\
 \text{(Phone)} \quad \quad \quad \quad \quad \boxed{\text{Sc} = 0} \\
 \text{(fruit)}^T
 \end{array}$$

$$\begin{array}{c} x_1 \ y_1 \\ [1, 0] \\ \hline x_2 \ y_2 \\ [0, 1] \end{array} \rightarrow \text{dot-Product} = \underline{\hspace{10em}}$$

word2vec → 3dim

$$\begin{aligned}
 CE_{bank} &= \underbrace{\left[ e_{bank}^T e_{money} \right]}_{\downarrow SC} \underbrace{e_{money}^T}_{=} \left[ e_{bank}^T e_{bank} \right] \\
 &\quad \downarrow SC \\
 &\quad \left( \begin{array}{|c|c|} \hline \text{ebank} \\ \hline \end{array} \right) \cdot \left( \begin{array}{|c|} \hline \text{emoney} \\ \hline \end{array} \right) \quad \left( \begin{array}{|c|c|c|} \hline \text{ebank} \\ \hline \end{array} \right) + \left( \begin{array}{|c|c|} \hline \text{ebank} \\ \hline \end{array} \right) \cdot \left( \begin{array}{|c|} \hline \text{ebank} \\ \hline \end{array} \right) \\
 &\quad e_{bank} + \left( \begin{array}{|c|c|} \hline \text{ebank} \\ \hline \end{array} \right) \underbrace{e_{grow}}_{\downarrow SC} \\
 &\quad \left( \begin{array}{|c|c|} \hline \text{ebank} \\ \hline \end{array} \right) + \left( \begin{array}{|c|c|} \hline \text{ebank} \\ \hline \end{array} \right) \cdot \left( \begin{array}{|c|} \hline \text{ebank} \\ \hline \end{array} \right) \underbrace{e_{grow}}_{\downarrow SC}
 \end{aligned}$$

Diagram illustrating softmax normalization:

The diagram shows three input nodes labeled  $SC_1$ ,  $SC_2$ , and  $SC_3$  connected to a central box labeled "softmax". The connections are labeled  $w_1$ ,  $w_2$ , and  $w_3$  respectively. Red arrows point from the labels  $w_1$ ,  $w_2$ , and  $w_3$  to their corresponding terms in the softmax formula below.

Below the diagram, the softmax function is defined as:

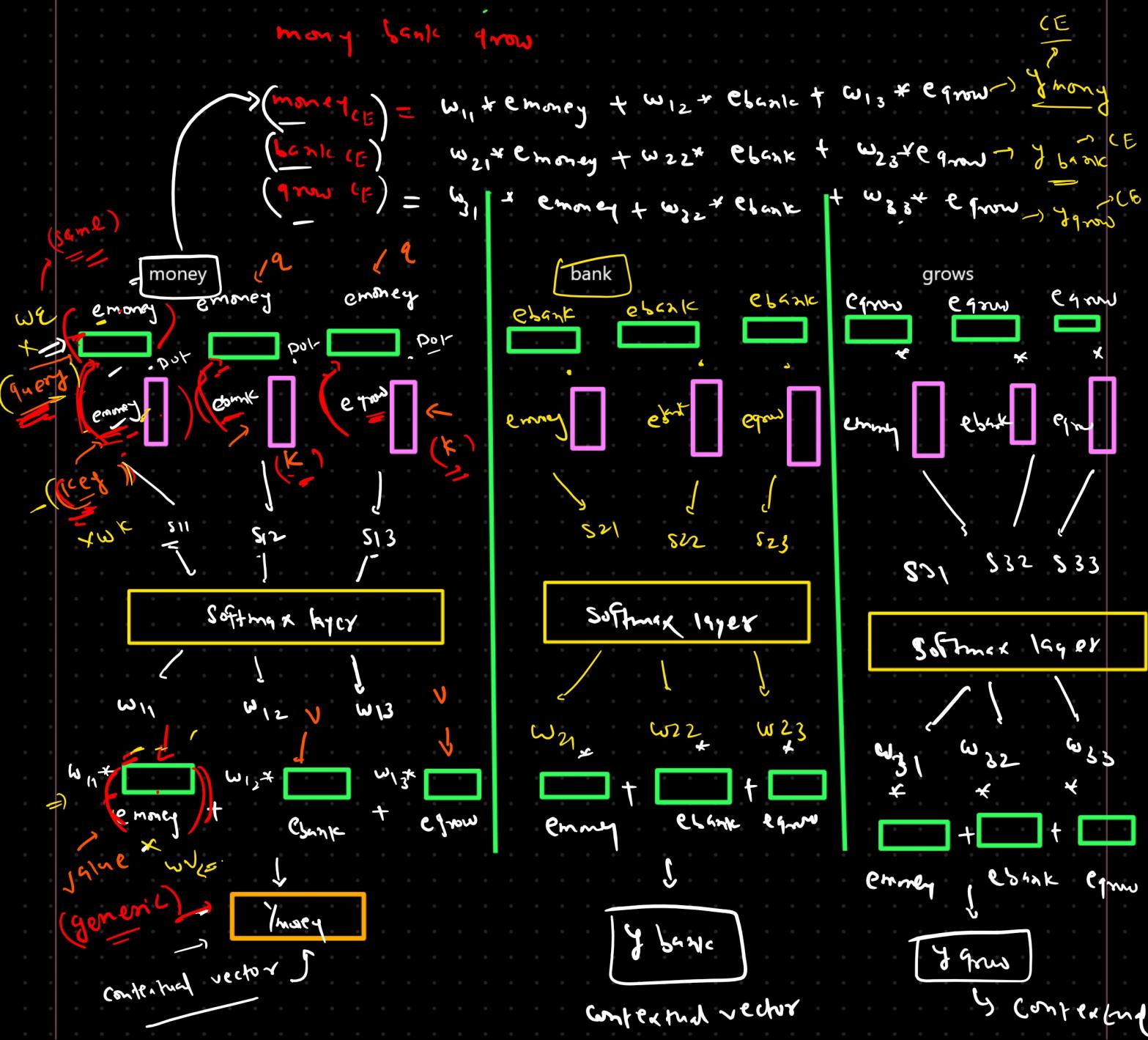
$$w_i = \frac{e^{SC_i}}{e^{SC_1} + e^{SC_2} + e^{SC_3}}$$

$$w_1 * \boxed{11} \text{ emoney} + w_2 * \boxed{11} \text{ ebank} + w_3 * \boxed{11} \text{ eqnow}$$

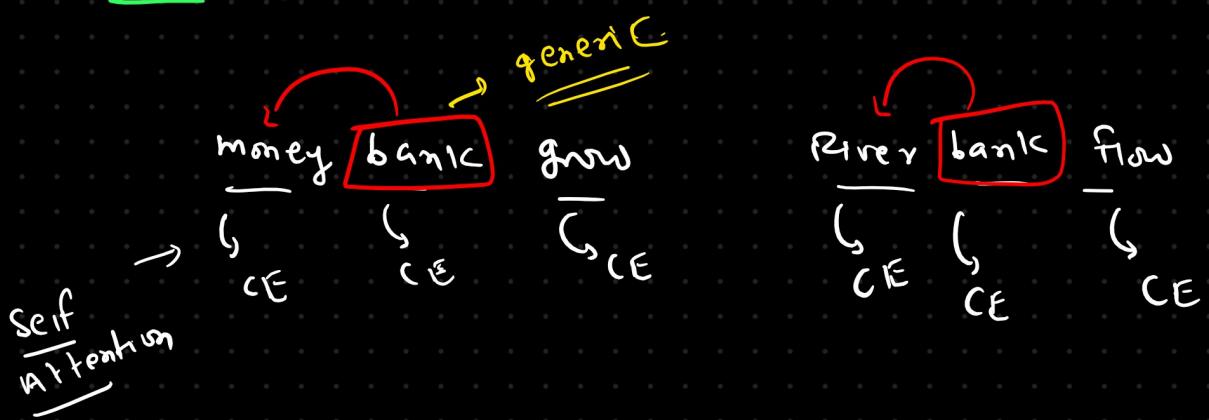
$\boxed{e\text{bank}}$   $\xrightarrow[\Rightarrow]{\substack{\text{(Contextual) vector} \\ \text{query, key, value}}}$

query, key, value

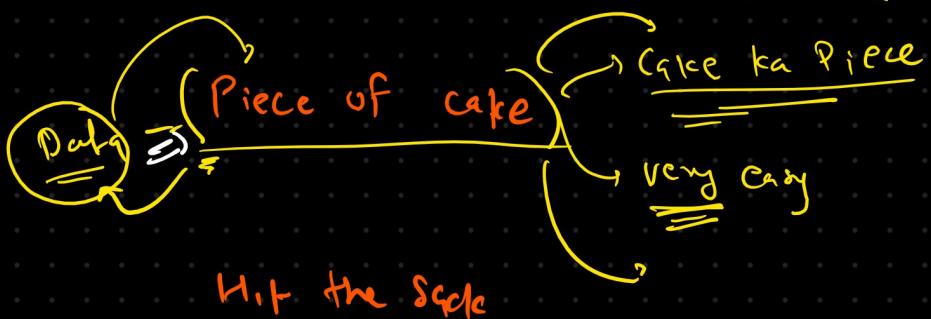
money bank grow



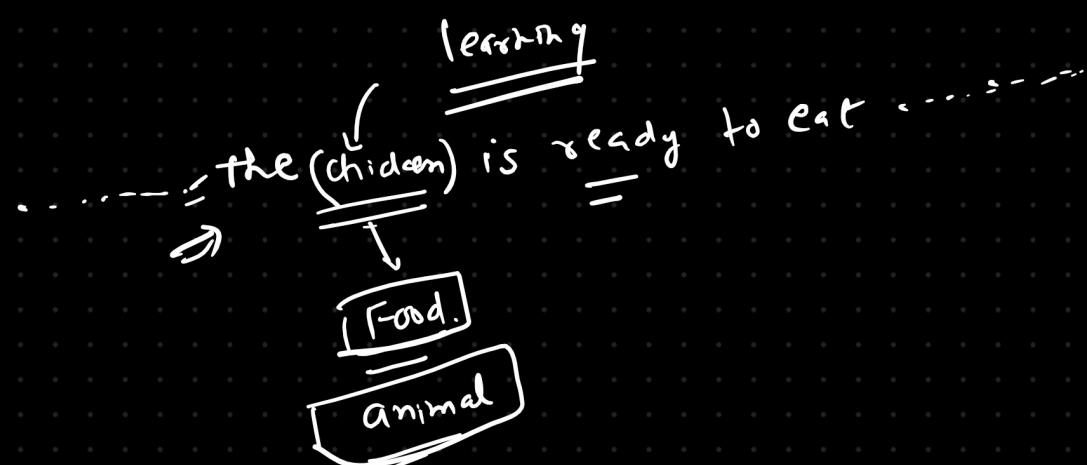
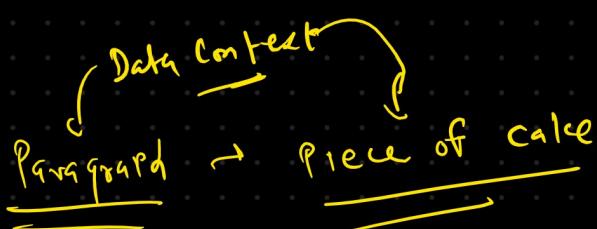
## Query key value



(Let the cat out of the bag) (Bill to bag se bhar hoga)  
 (Secret reveal learning) (Secret reveal learning)

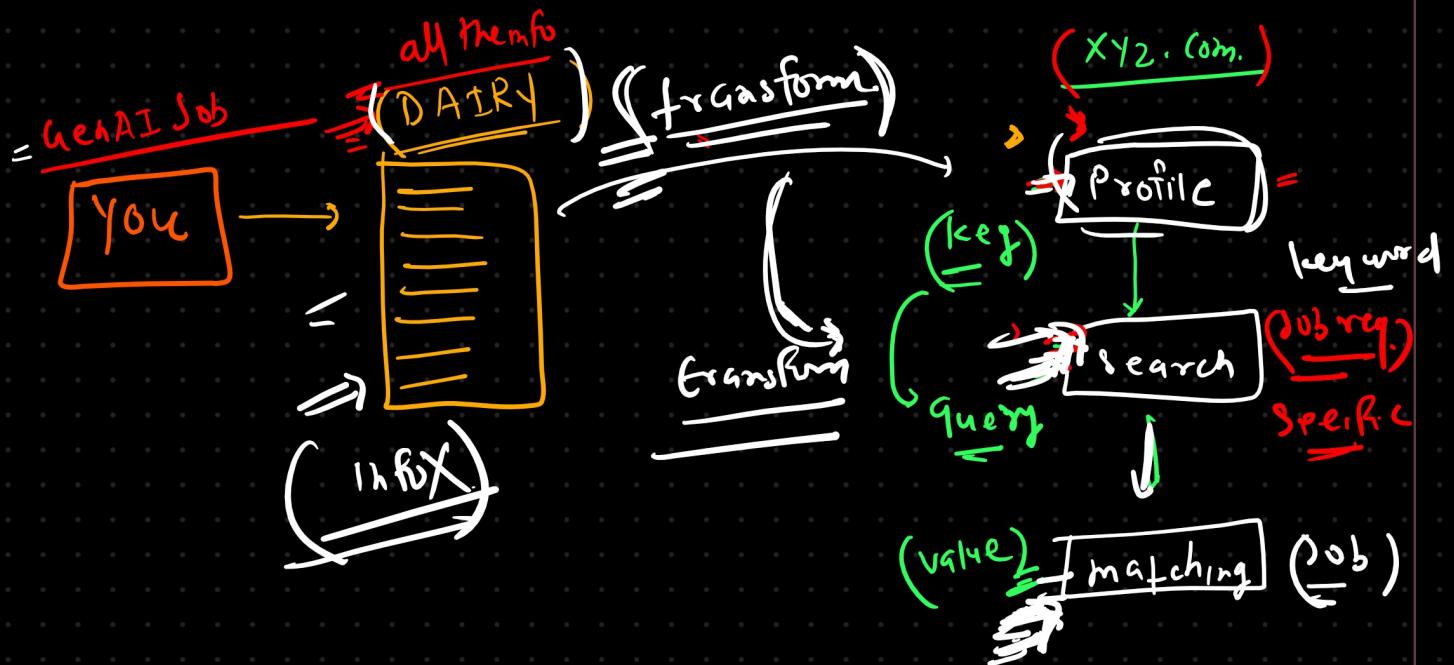
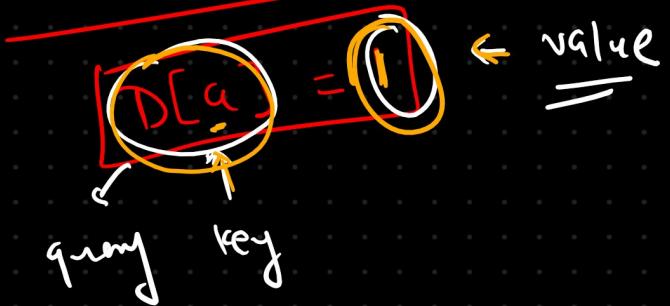


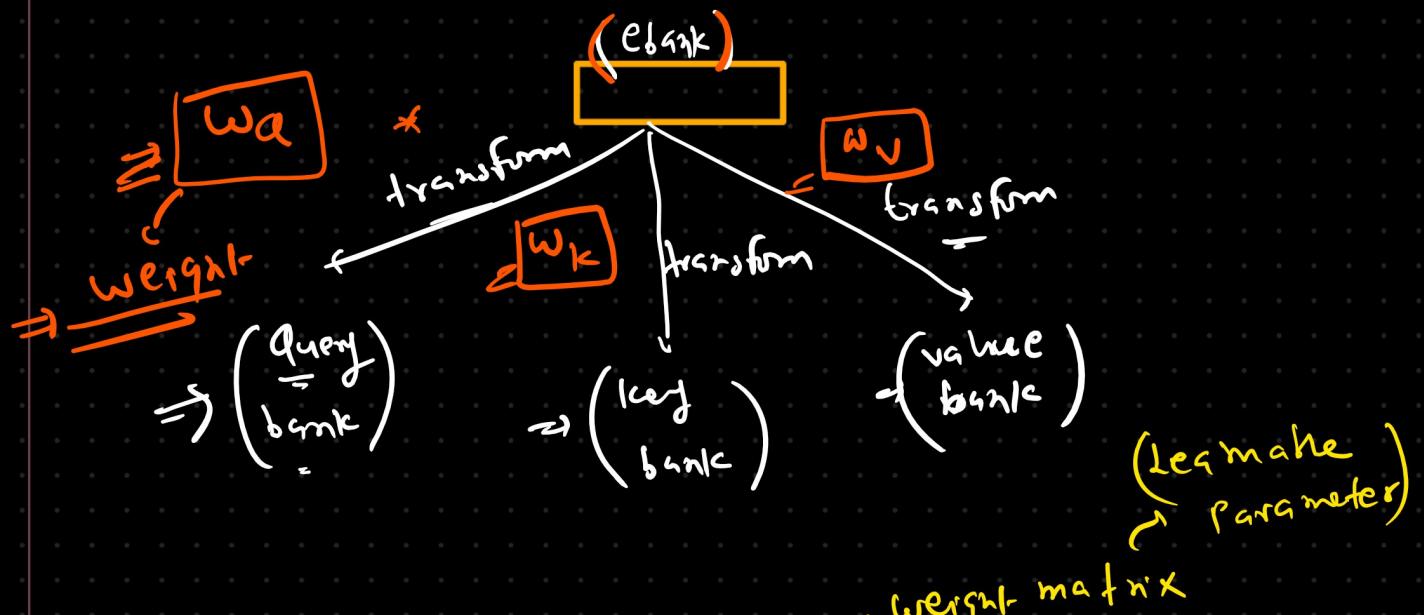
Hit the scale



Q, K, V

$$D = \{a:1, b:2, c:3\}$$





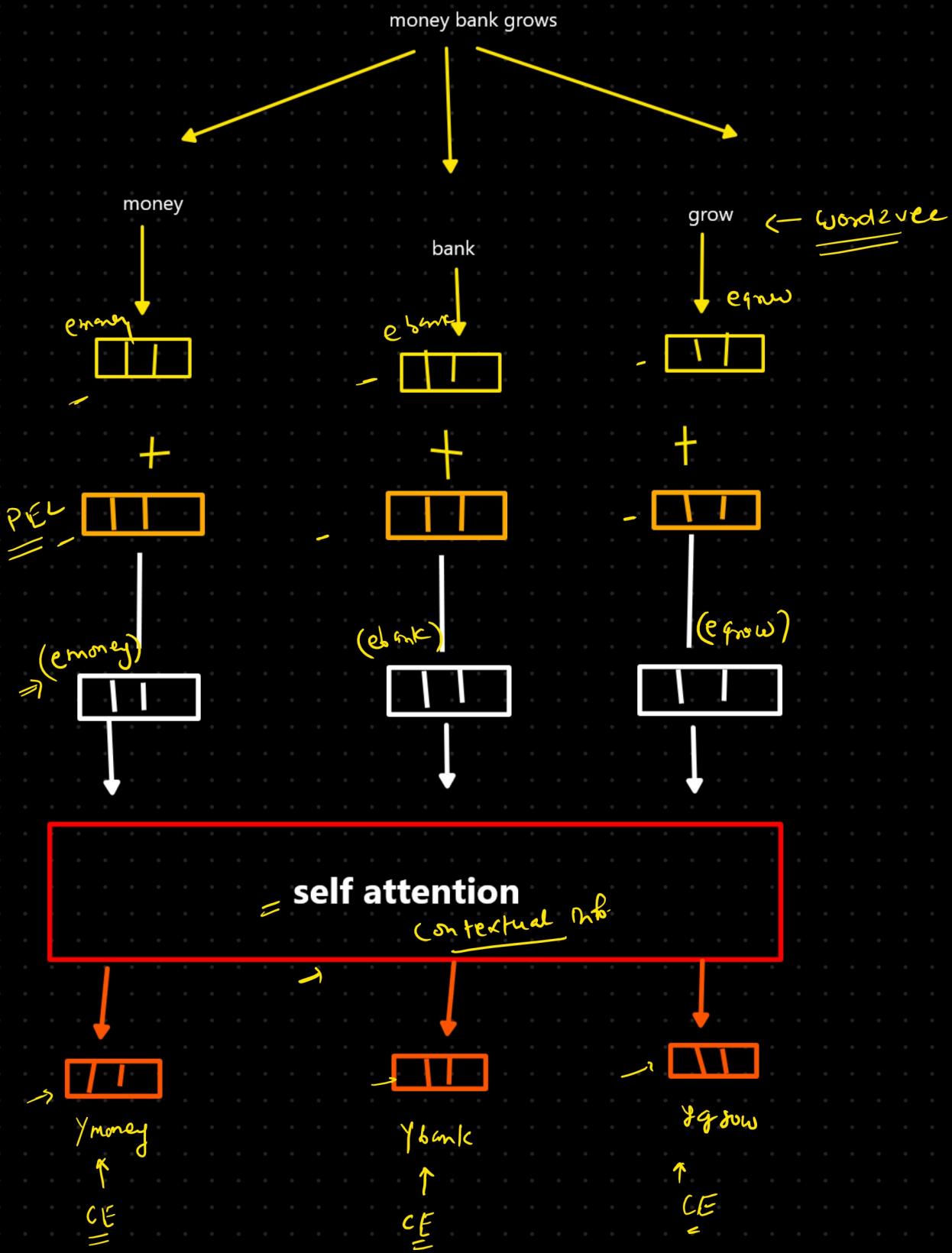
$e_{bank} \rightarrow$

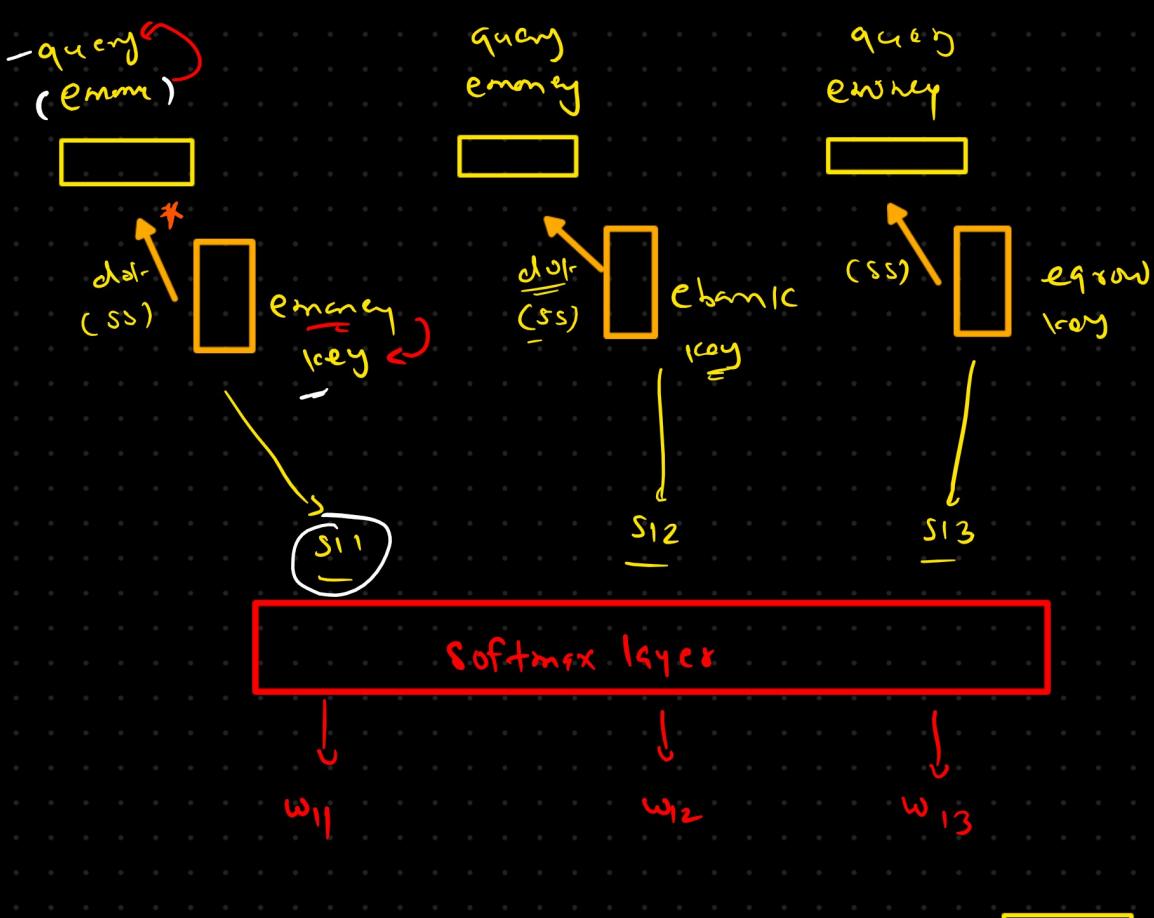
- $q_{bank}$
- $k_{bank}$
- $v_{bank}$

$$\begin{aligned}
 e_{bank} * \underline{w_q} &\Rightarrow q_{bank} \\
 e_{bank} * \underline{w_k} &\Rightarrow k_{bank} \\
 e_{bank} * \underline{w_v} &= v_{bank}
 \end{aligned}$$

weight matrix

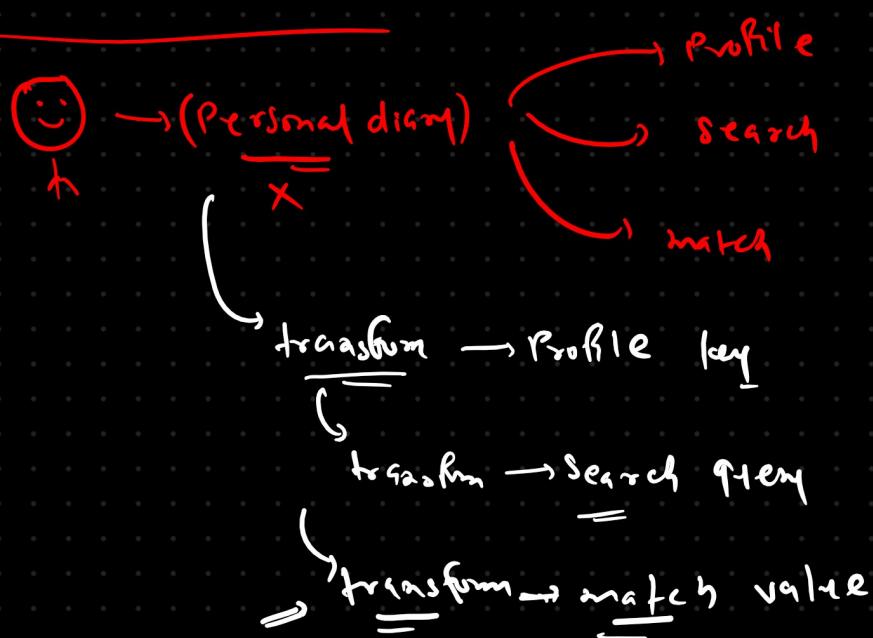
(Learnable  
parameter)





$$\begin{aligned} w_{11}^* \text{emoney} &= \text{values} \\ w_{12}^* \text{ebamik} &= \text{value} \\ w_{13}^* \text{egrow} &= \text{value} \end{aligned}$$

that is not a correct



$$\begin{aligned} \text{emoney} &\rightarrow \text{q money} \rightarrow \text{emoney} * \underline{w \text{ money}} \xrightarrow{\text{transform}} \text{q money} \\ \text{ic money} &\rightarrow \text{ic money} \rightarrow \text{emoney} * \underline{w \text{ ic money}} \xrightarrow{\text{transform}} \text{ic money} \\ \text{v money} &\rightarrow \text{v money} \rightarrow \text{emoney} * \underline{w \text{ v money}} \xrightarrow{\text{transform}} \text{v money} \end{aligned}$$

weight & bias  
 $(N \times D)$   
 ↗ Learnable  
 ↙ Training  
 ↘ Learn

