

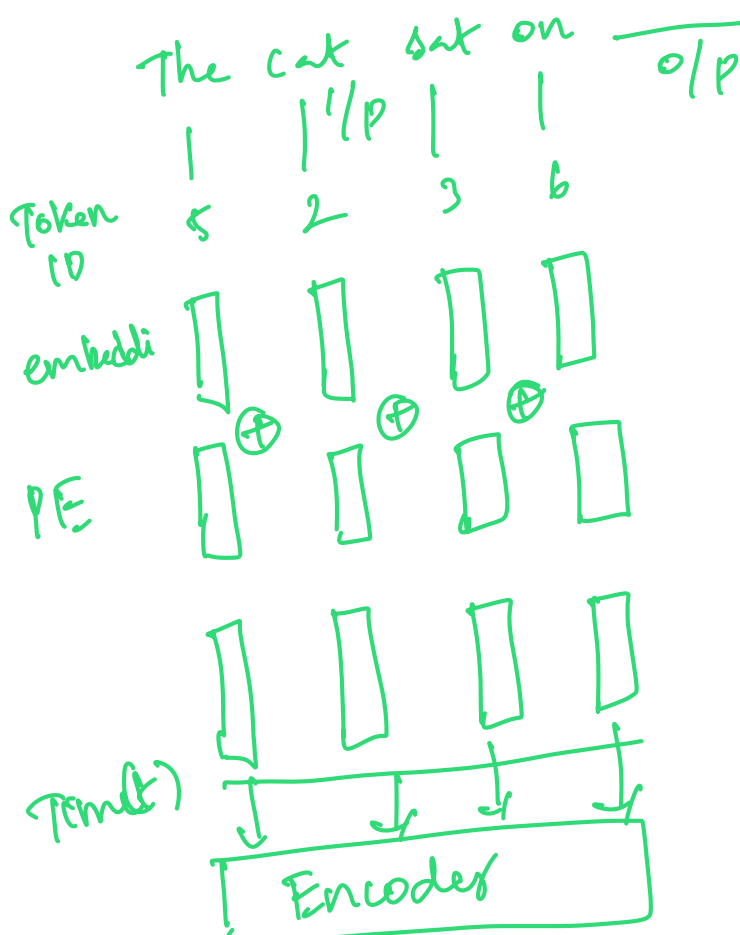
Transformer Architecture

Encoder

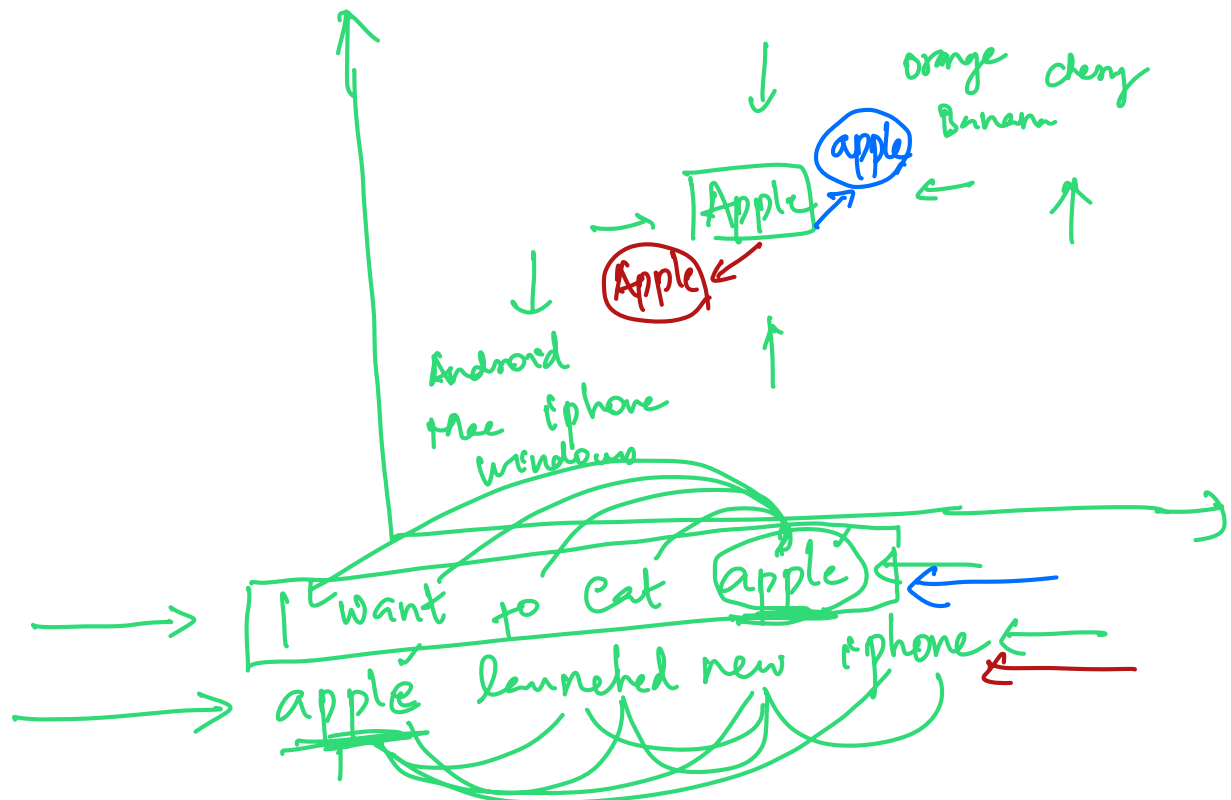
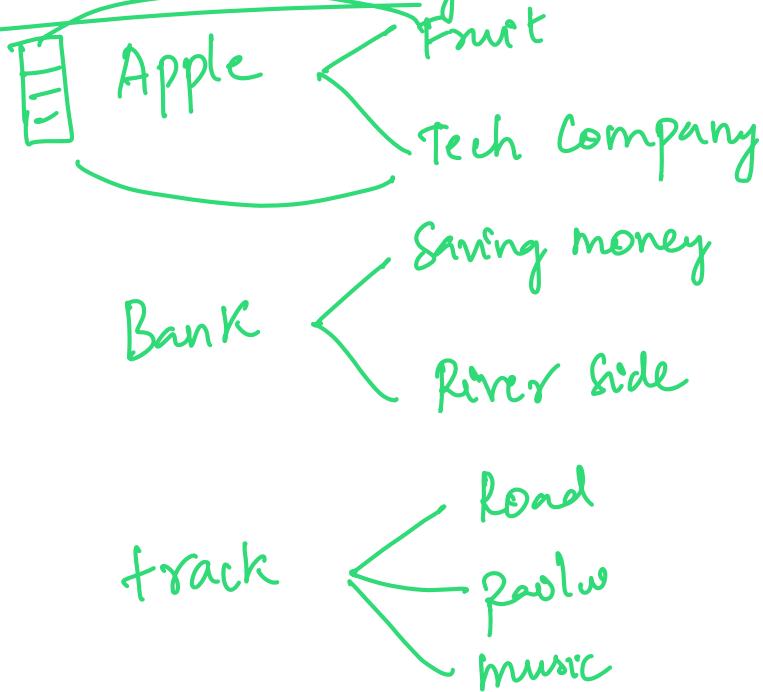
Decoder

Accept the input in tokens parallel

Outputs the token one at a time

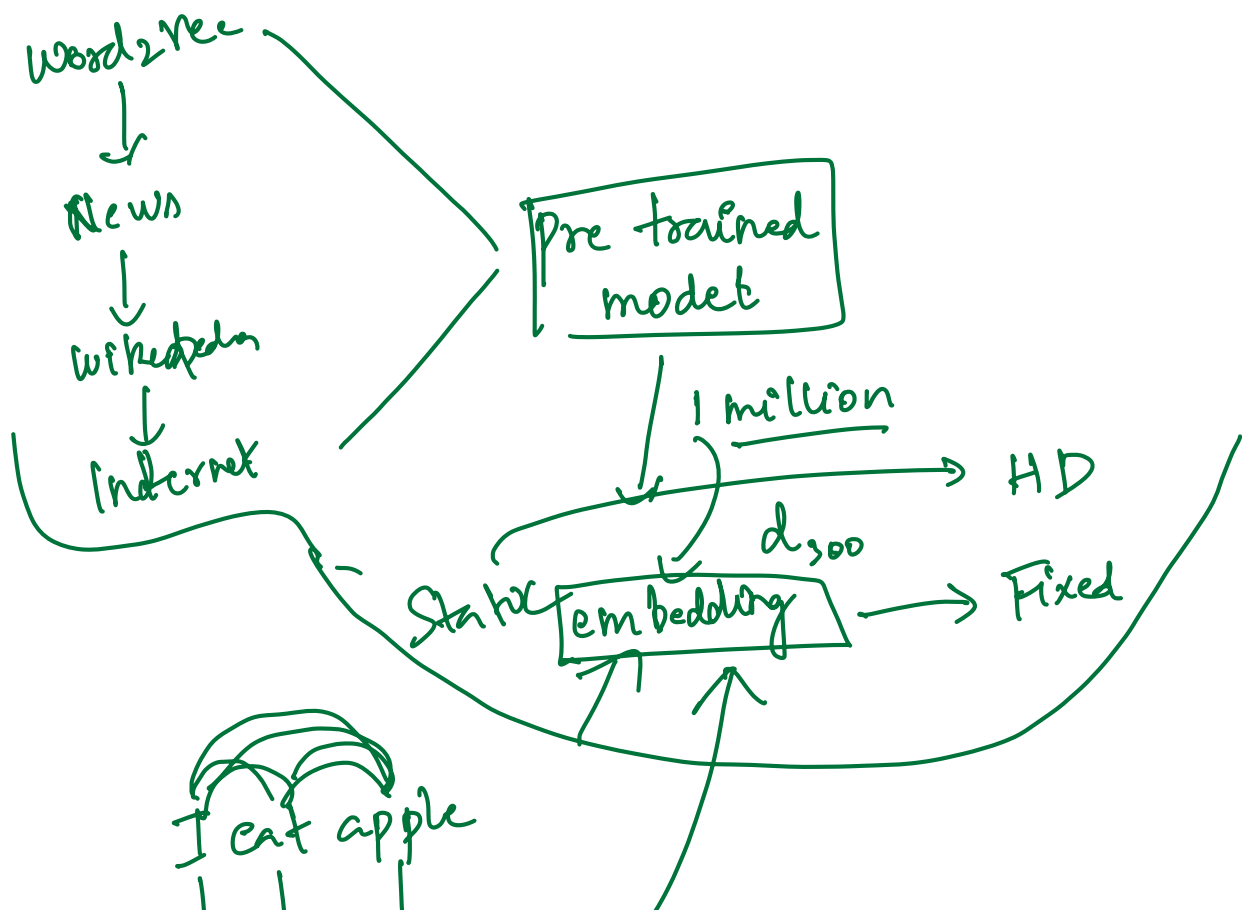


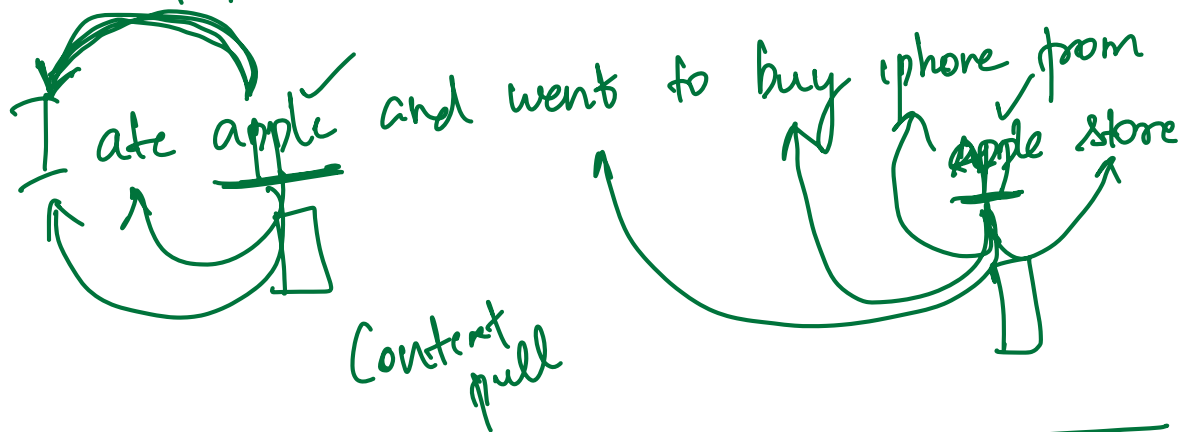
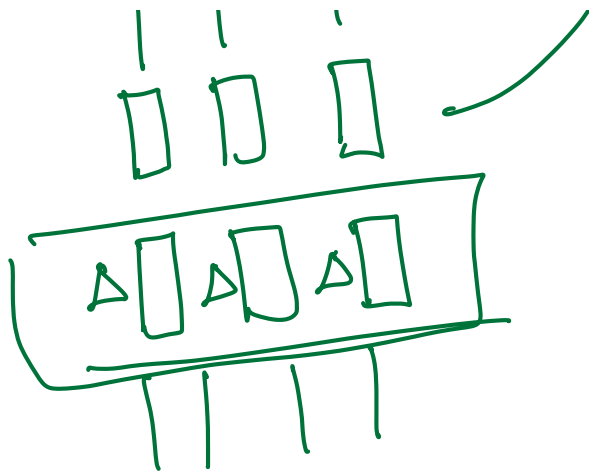
Static Embedding



embedding → Static
↓
Context
embedding
↓
based on Document

We are going to update the Static
Embedding based on the neighbouring
tokens with more Context

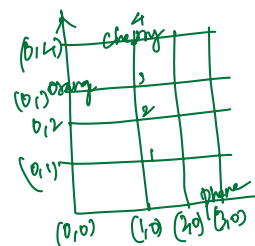




Compare \longleftrightarrow Similarity

$$\text{Cosine Similarity} = \frac{A \cdot B}{\sqrt{A_1^2 + A_2^2} \cdot \sqrt{B_1^2 + B_2^2}}$$

Dot Product



$d=2$

Cosine . dot

Cherry	<div>✓ 1 4</div>	$\frac{0+12}{\sqrt{1+16} \cdot \sqrt{0+9}} = 0.97$	12	scaled dot product $\frac{12}{\sqrt{1}} = \frac{12}{\sqrt{2}} = 8.49$
Orange	0 3			
Cherry	1 4	$\frac{3+0}{\sqrt{1+9} \sqrt{16+0}} = 0.24$	3	$3/\sqrt{2}$ 2.12
Phone	3 0			
Orange	0 3	$\frac{0+0}{\sqrt{0+0}} = 0$	0	$0/\sqrt{2}$ 0
Phone	3 0			

apple and an orange
an apple phone

	Tech	Feature	other
Orange	0	.3	.0
phone	4	0	0
apple	2	.2	0
and	0	0	2
an	0	0	3

n

Similarity		Dot Product			
	orange	phone	Apple	And	An
orange	9	0	6✓	0	0
phone	0	16	8✓	0	0
Apple	6	8	8	0	0
And	0	0	0	4	6
An	0	0	0	6	9

Scaled Dot Product

	orange	phone	Apple	And	An
orange	$9/\sqrt{5}$	$0/\sqrt{5}$	$6/\sqrt{5}$	0	0
phone	0	$16/\sqrt{5}$	$8/\sqrt{5}$	0	0
Apple	$6/\sqrt{5}$	$8/\sqrt{5}$	$8/\sqrt{5}$	0	0
And	0	0	0	$4/\sqrt{5}$	$6/\sqrt{5}$

An	0	0	0	$6/\sqrt{5}$	$9/\sqrt{5}$
----	---	---	---	--------------	--------------

1.7313

Scaled Dot Product

	Orange	phone	Apple	And	An
Orange	5.2	0 [~]	3.4	0 [✓]	0 [✓]
phone	0	9.2	4.6	0	0
Apple	3.4	4.6	4.6	0	0
And	0	0	0	2.3	3.4
An	0	0	0	3.4	5.2

apple and an orange

an apple phone

	apple	orange	and	an
apple	4.6	3.4	0	0
orange	3.4	5.2	0	0
and	0	0	2.3	3.4
an	0	0	3.4	5.2

	apple	phone	an
apple	4.6	4.6	-
phone	4.6	9.2	

Apple and an Orange

$$\begin{aligned}\text{Apple} &= \underline{4.6} \text{ apple} + \underline{3.4} \text{ Orange} \\ \text{Orange} &= 3.4 \text{ apple} + 5.2 \text{ Orange}\end{aligned}$$

an Apple phone

$$\begin{aligned}\text{Apple} &= 4.6 \text{ Apple} + 4.6 \text{ phone} \\ \text{phone} &= 4.6 \text{ Apple} + 9.2 \text{ phone}\end{aligned}$$

document

token



embedding



dot product in table



Scaled by $1/\sqrt{d}$



each token understanding
how it is attending
others token

Next

I want coefficients to add 1

Normalization

$$\text{Apple} = \underline{4.6} \text{ apple} + \underline{3.4} \text{ Orange}$$

$$\text{Orange} = 3.4 \text{ apple} + 5.2 \text{ Orange}$$

$$\frac{4.6 \text{ apple} + 3.4 \text{ Orange}}{4.6 + 3.4}$$

$$4.6 / 4.6 + 3.4$$

$$3.4 / 4.6 + 3.4$$

$$\text{Orange} \Rightarrow 0.57 \text{ apple} + 0.43 \text{ orange}$$

$$\frac{1x - 1y}{1-1}$$

$$x \rightarrow \frac{e^x}{1} \rightarrow 0 \rightarrow \infty$$

Need Coeff to be positive

$$\text{Apple} = \underline{4.6} \text{ apple} + \underline{3.4} \text{ Orange}$$

$$\text{Orange} = 3.4 \text{ apple} + 5.2 \text{ Orange}$$

$$\text{Apple} = \frac{e^{4.6} \text{ apple} + e^{3.4} \text{ Orange}}{e^{4.6} + e^{3.4}}$$

Normalization

$$= \frac{e^{4.6} / e^{4.6}}{e^{4.6} / e^{4.6} + e^{3.4}} + \frac{e^{3.4} / e^{3.4}}{e^{4.6} / e^{4.6} + e^{3.4}}$$

$$\text{Apple} = 0.55 \text{ Apple} + 0.45 \text{ Orange}$$

$$\text{Orange} = 0.44 \text{ apple} + 0.56 \text{ Orange}$$

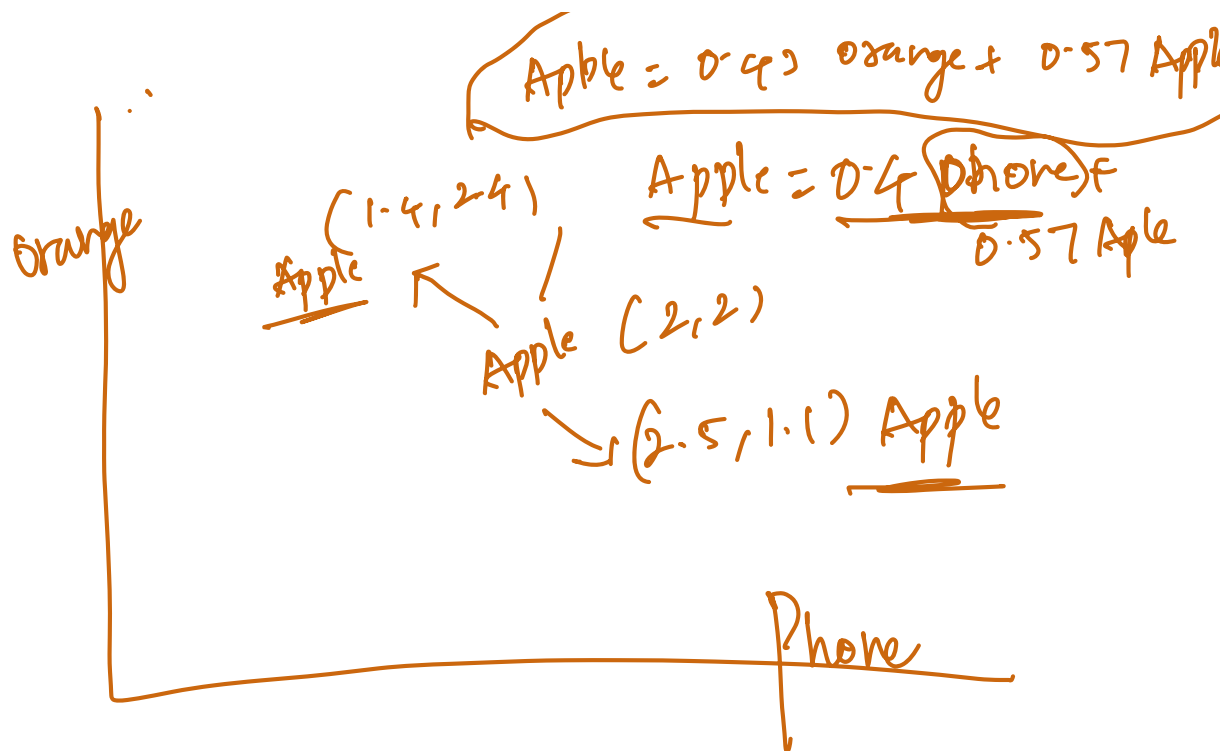
an Apple phone

$$\text{Apple} = 4.6 \text{ Apple} + 4.6 \text{ phone}$$

$$\text{phone} = 4.6 \text{ Apple} + 9.2 \text{ phone}$$

$$\text{Apple} = 0.43 \text{ phone} + 0.57 \text{ Apple}$$

$$\text{phone} = 0.57 \text{ phone} + 0.43 \text{ Apple}$$



Reiterate

Given document

||

Tokenization

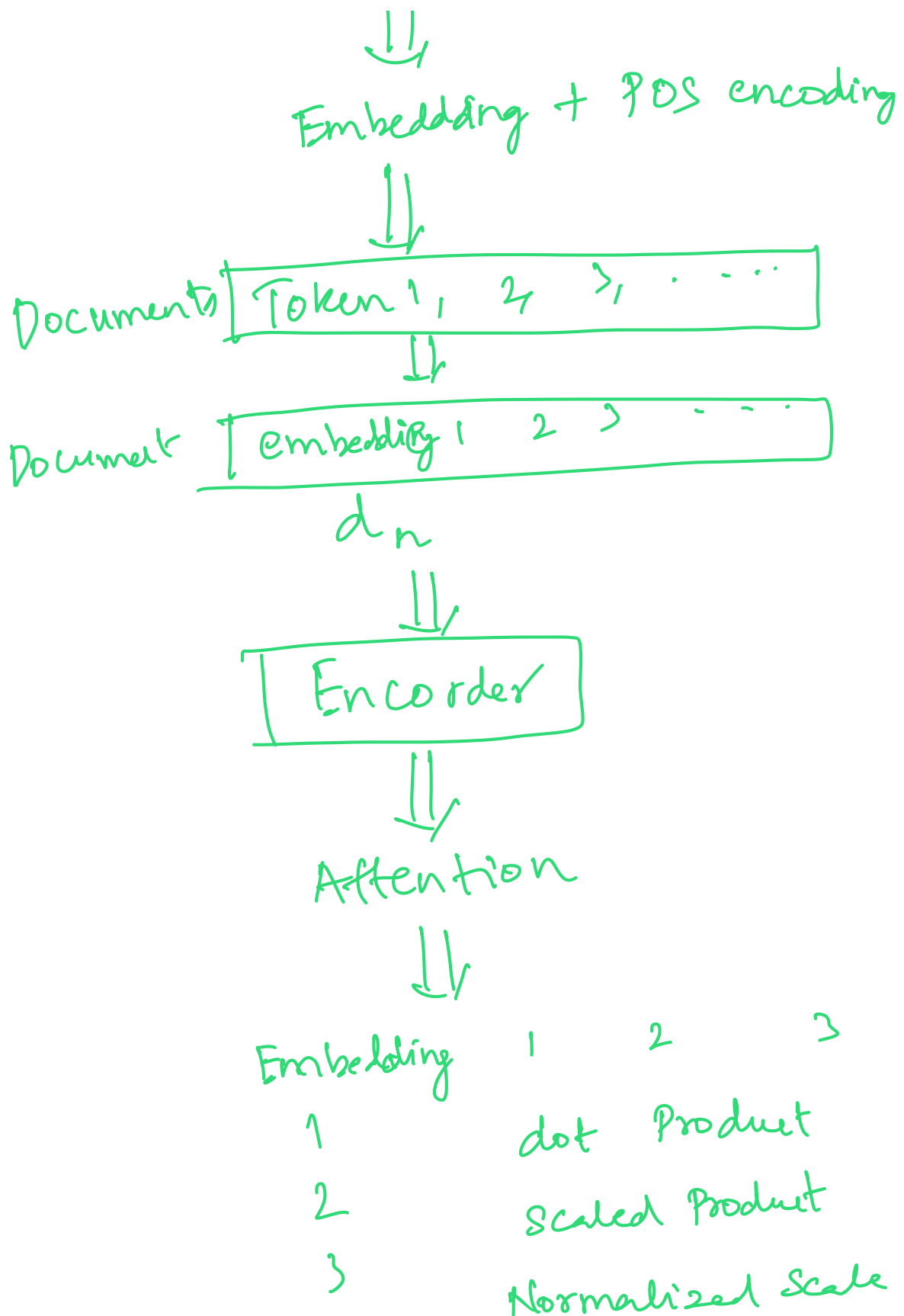
||

Token ID

||

Embedding

||

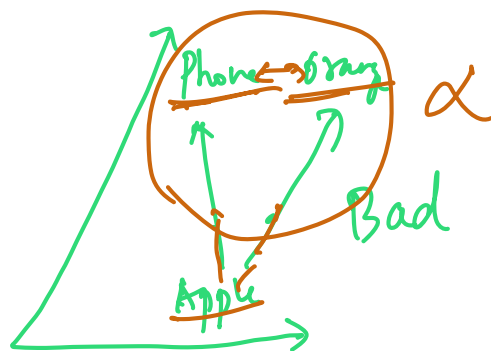
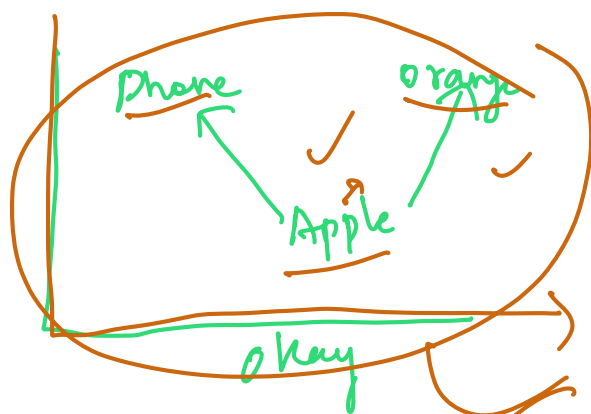


Self Attention

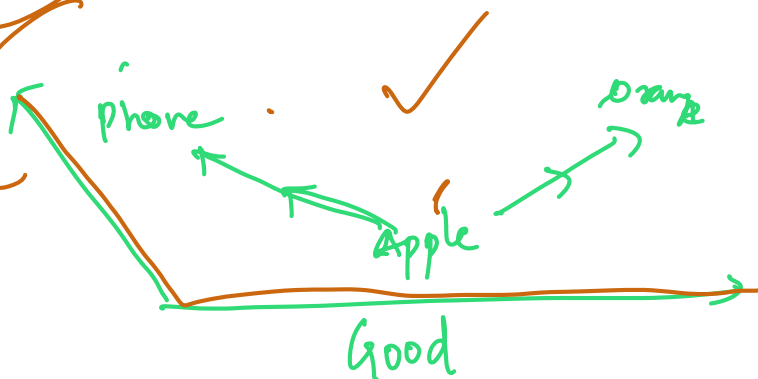
Dot product

Exponential Normalized
Scaled Dot product

update the new
embedding
for each
tokens



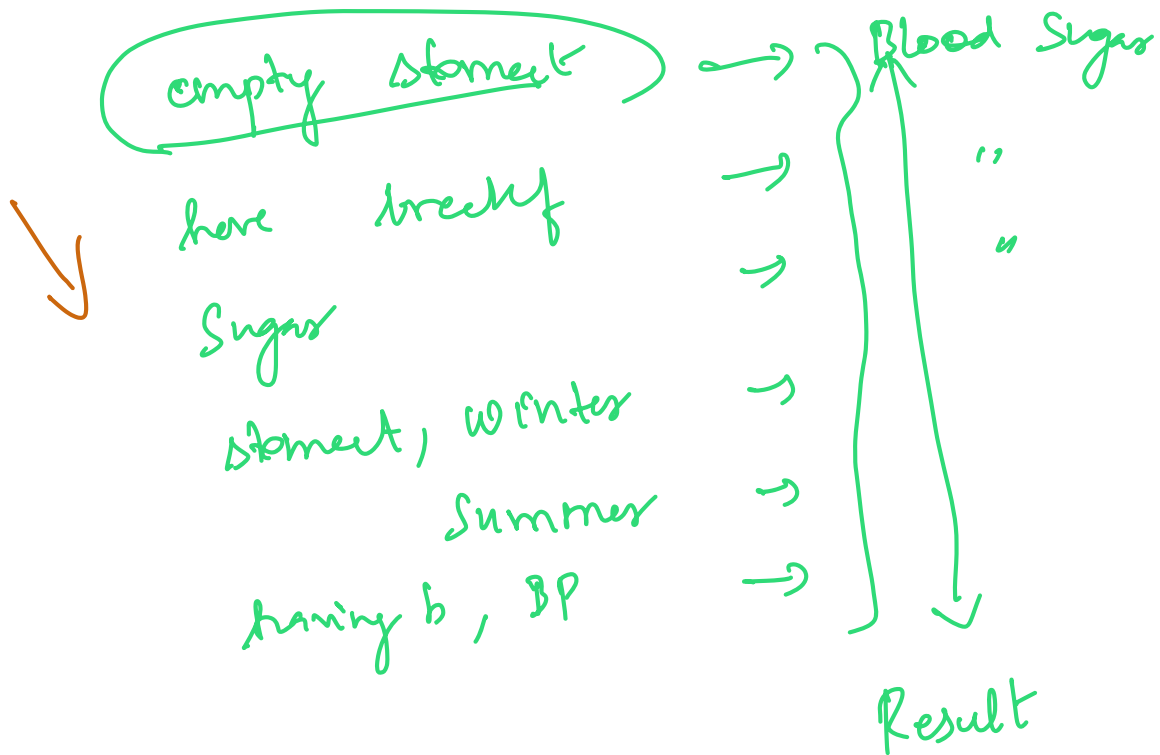
linear
transformation
of embedding



one Testing — Fit → Refence

↳ taking
fitness
↳ sports

diabetes:



Multi head

↳ linear
transformer

Attention

↓
one given document

1 /
✓
Always.