

no/m res : b p d_{model}

self, w : | | d_{model}

$w_{pos}[0]$ \leftarrow 0

$v_1 \leftarrow 1$

$v_2 \leftarrow 2$

3

The

cat

is

happy

token

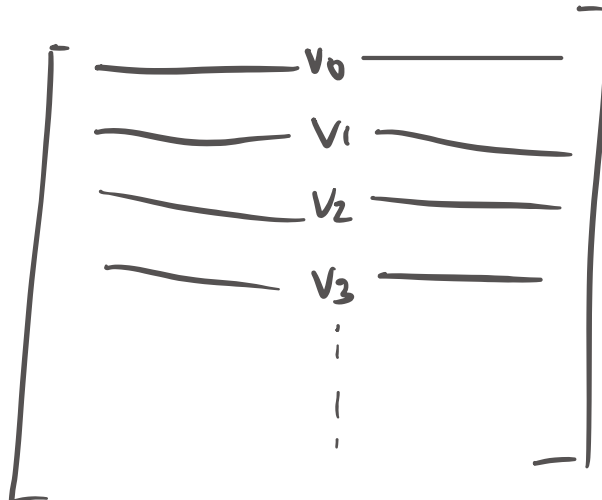
11 \rightarrow [$-d_{model}-$]

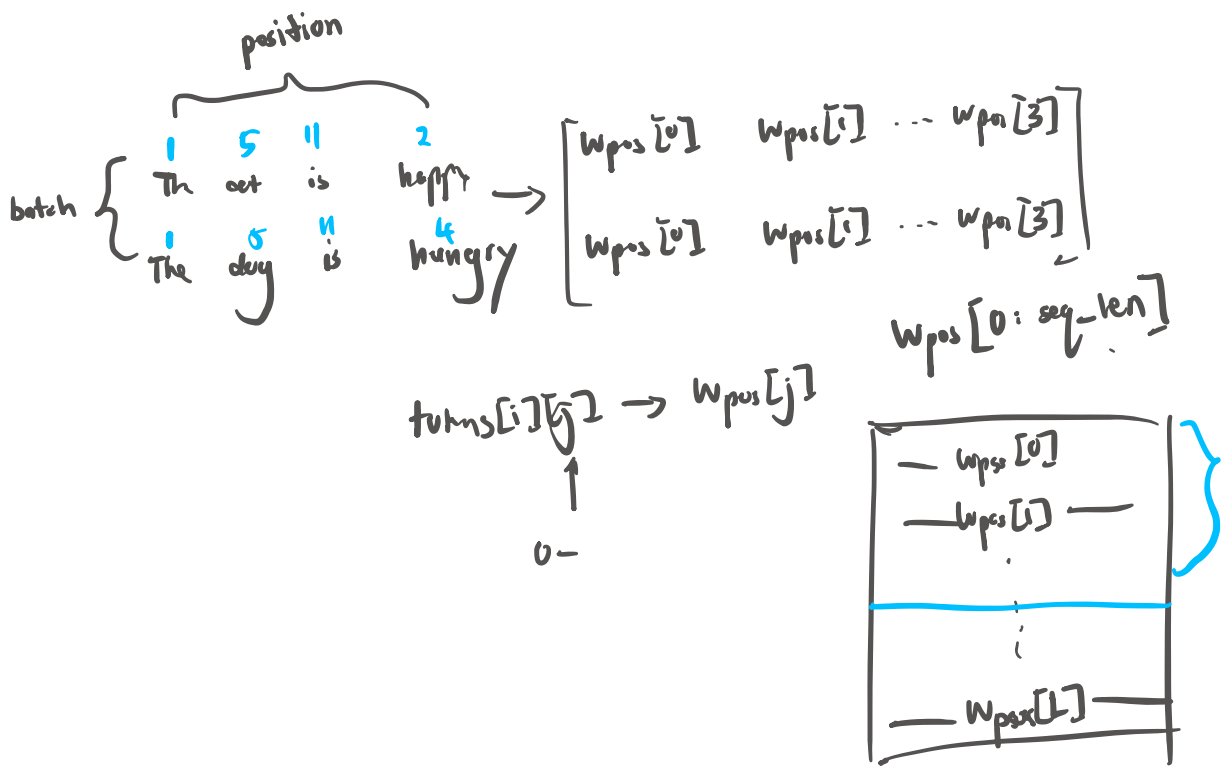
7 \rightarrow [$-d_{model}-$]

1

24

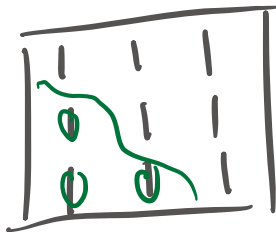
W-pos :



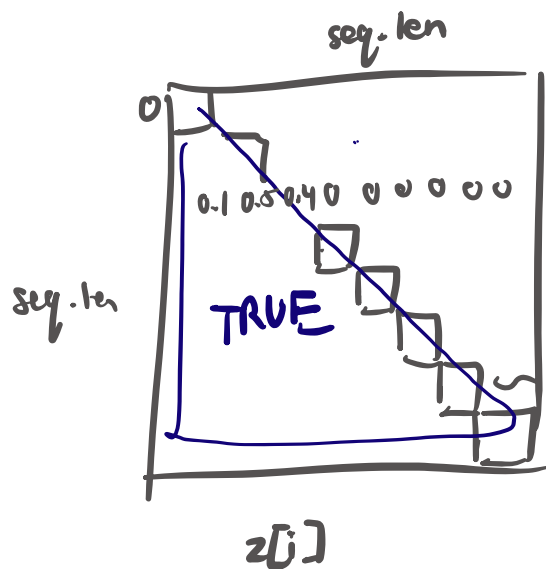


$[w_{pos}[0:seq_len] \text{ for } _ \text{ in range(num_batch)}]$

$[Tensor \quad Tensor \quad \dots \quad Tensor]$



Tensor



q =

k =

" t1 shape , t2 shape \rightarrow output shape "

avg
x

-4.04

$$-\log(0.9) - \log(0.1)$$

The eat is happy
 ↓ ↓ ↓ ↓
 [1 1 2 3]

[] [] is: 0.9 happy: 0.1 $-\log(0.1)$

hungry: 0.3

the: 0

happy: 1

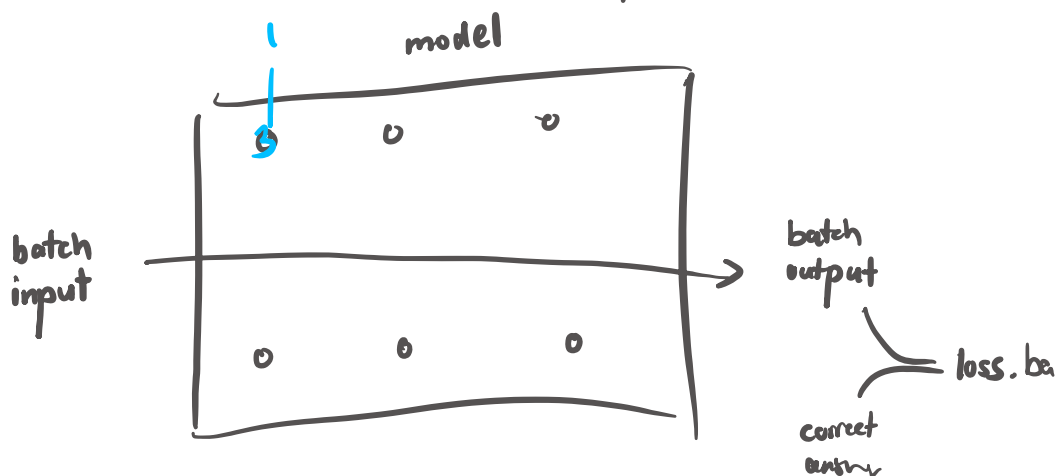
happy: 0

$$-\log(1) = 0$$

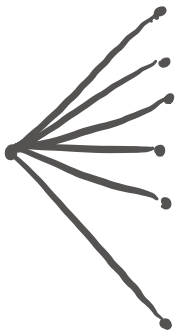
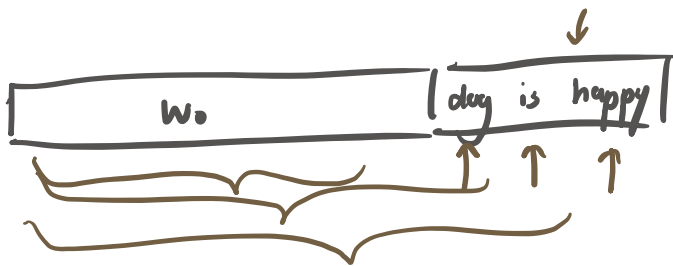
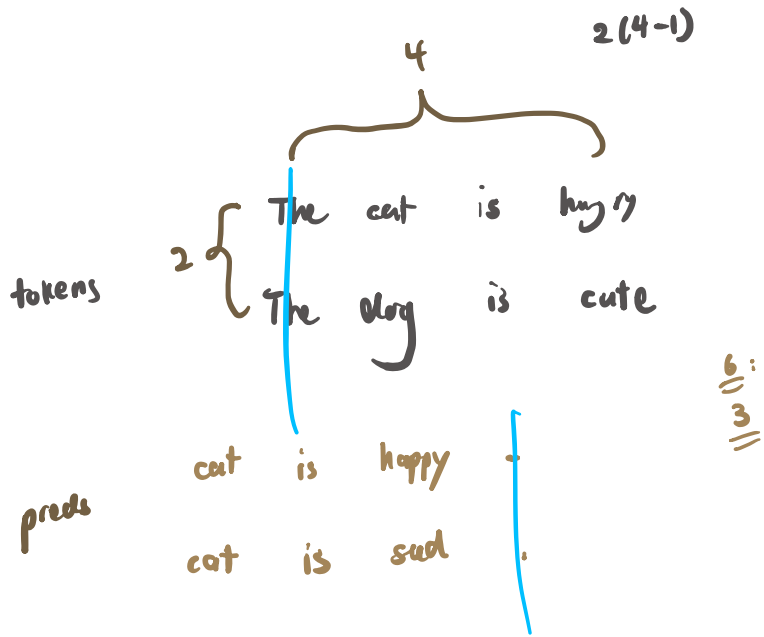
$$-\log(0) = -\infty$$


batch seq-len vocab-len

, dim=2



$$\text{batch}[\text{tokens}], \text{shape}[0] \times (\dots, \text{shape}[1]);$$



The nice 

~~The dog~~

The nice woman

the dog is happy

batch seq-len d-vocab

tokens: seq-len

logits: batch d-vocab