

$V = 5$   
 $M = 3$  (embedding size)

0
0
1
0
0

One-hot  
encoding  
of **study**  
(1 x 5)

→ target  
word

trying to  
predict  
word content  
like

If sentence = 'I like to study' and  $m=2$   
then i/p target  
Study like  
Study to  
Study tonight  
Study END  
word matrix keeps getting updated. ↓  
below process repeated for each of these.

**V = 5**

**M = 3** (embedding size)

0	X	-0.5	0.5	0.3
0		-0.3	0.2	0.1
1		0.5	1.2	0.9
0		0.3	0.5	1.2
0		0.1	1.0	-0.2

One-hot  
encoding  
of **study**  
(1 x 5)

Word matrix  
(5 x 3)

**V = 5**

**M = 3** (embedding size)

The diagram illustrates the calculation of a word vector. On the left, a one-hot encoding vector of size 5 is shown as a column of five cells. The third cell contains the value 1, while the others are 0. This vector is multiplied (indicated by a large 'X') by a word matrix of size 5x3. The matrix contains numerical values for each of the five rows and three columns. The third row of the matrix, corresponding to the '1' in the one-hot vector, is highlighted in yellow. An equals sign follows, leading to a word vector of size 1x3. This vector is shown as a column of three cells containing the values 0.5, 1.2, and 0.9, which are the values from the highlighted row of the matrix. The word vector is enclosed in a blue oval.

0
0
1
0
0

**X**

-0.5	0.5	0.3
-0.3	0.2	0.1
0.5	1.2	0.9
0.3	0.5	1.2
0.1	1.0	-0.2

**=**

0.5
1.2
0.9

One-hot  
encoding  
of **study**  
(1 x 5)

Word matrix  
(5 x 3)

Word vector  
for study  
(1 x 3)

$V = 5$   
 $M = 3$  (embedding size)

0
0
1
0
0

 $\times$ 

-0.5	0.5	0.3
-0.3	0.2	0.1
0.5	1.2	0.9
0.3	0.5	1.2
0.1	1.0	-0.2

 $=$ 

0.5
1.2
0.9

 $\times$ 

0.0	1.0	2.0	-1.0	-0.5
-2.3	1.2	0.4	5.6	-1.0
0.1	-0.2	-0.5	0.4	-0.6

*unique words*

*latent dimensions*

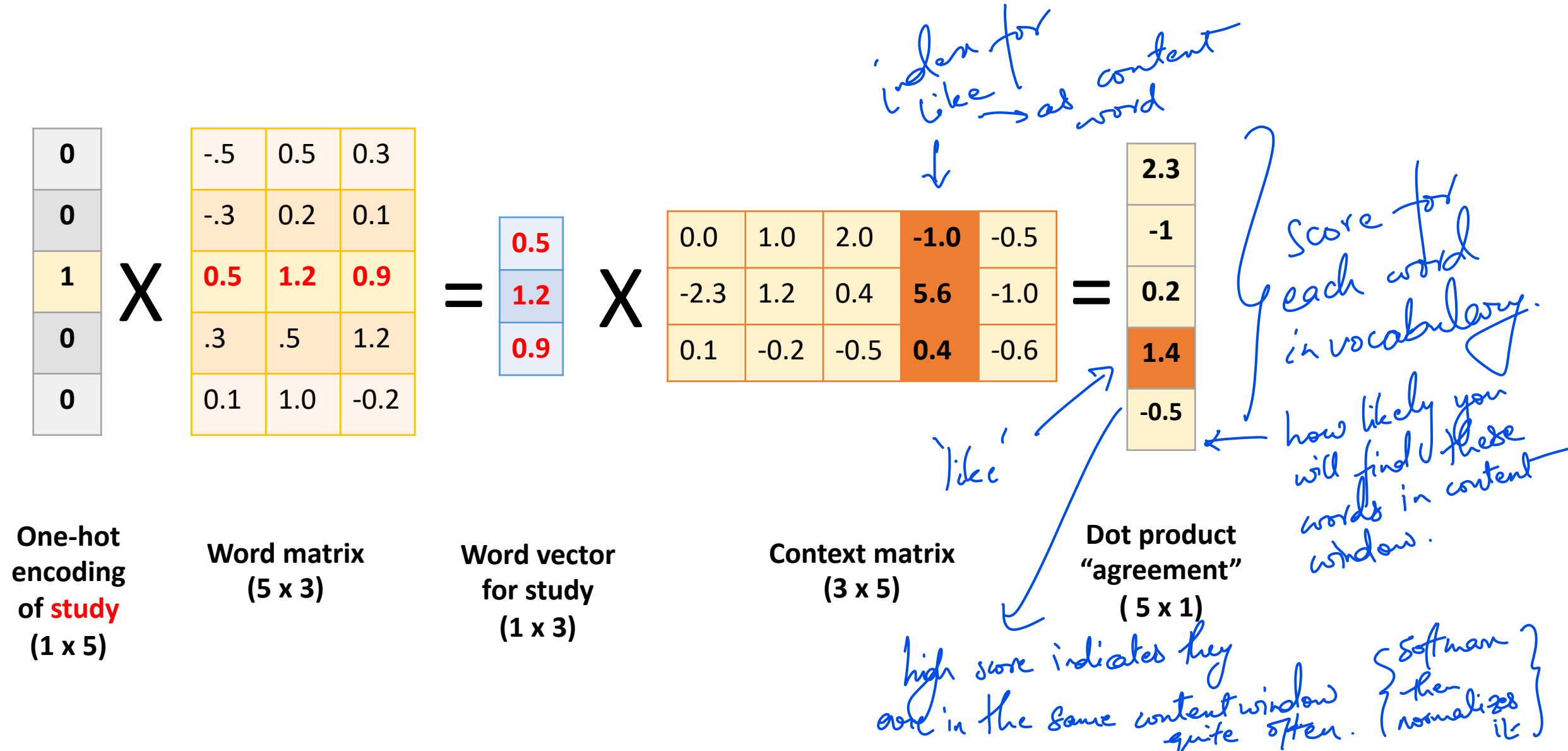
One-hot  
encoding  
of **study**  
(1 x 5)

Word matrix  
(5 x 3)

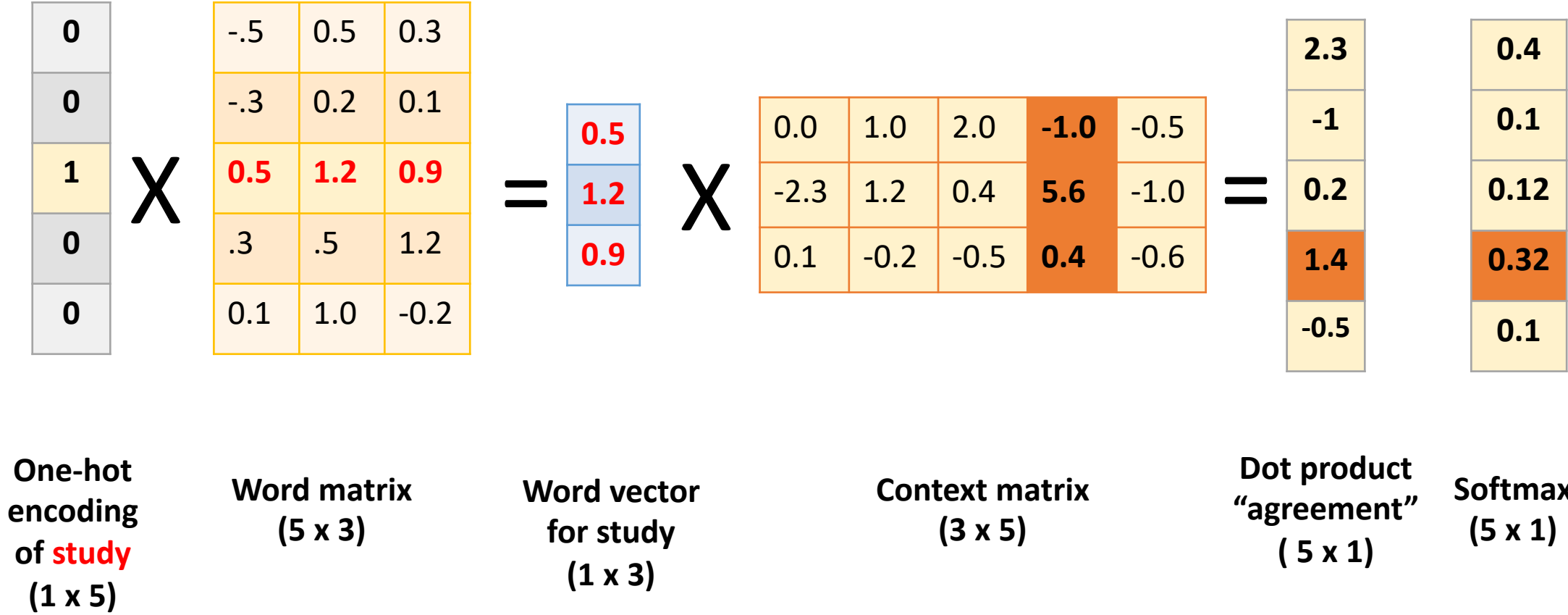
Word vector  
for study  
(1 x 3)

Context matrix  
(3 x 5)

$V = 5$   
 $M = 3$  (embedding size)



V = 5  
M = 3 (embedding size)



V = 5  
M = 3 (embedding size)

1<sup>st</sup> epoch of training

0	-0.5	0.5	0.3
0	-0.3	0.2	0.1
1	0.5	1.2	0.9
0	0.3	0.5	1.2
0	0.1	1.0	-0.2

X

=

0.5
1.2
0.9

X

0.0	1.0	2.0	-1.0	-0.5
-2.3	1.2	0.4	5.6	-1.0
0.1	-0.2	-0.5	0.4	-0.6

=

2.3
-1
0.2
1.4
-0.5

prediction

0.4
0.1
0.12
0.32
0.1

ground truth

0
0
0
1
0

One-hot  
encoding  
of **study**  
(1 x 5)

Word matrix  
(5 x 3)

Word vector  
for study  
(1 x 3)

Context matrix  
(3 x 5)

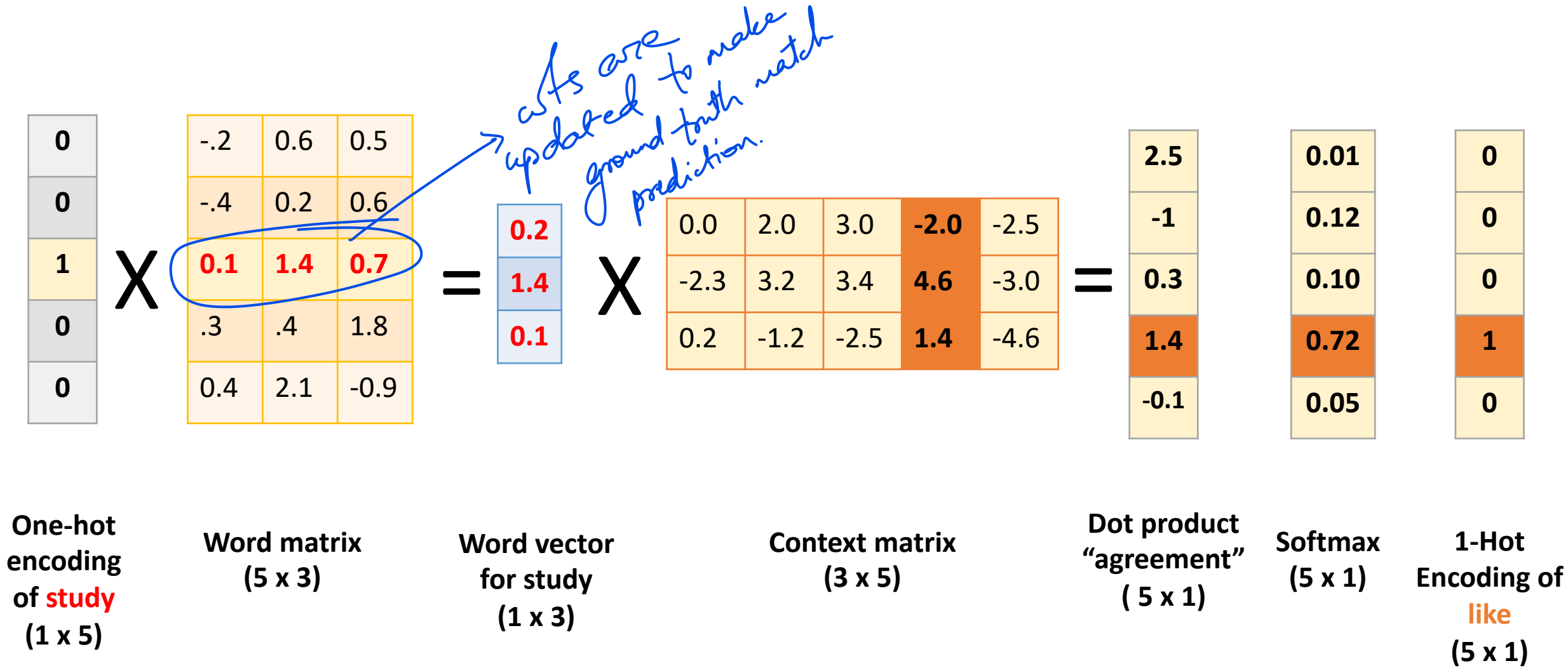
Dot product  
"agreement"  
(5 x 1)

Softmax  
(5 x 1)

1-Hot  
Encoding of  
**like**  
(5 x 1)

V = 5  
M = 3 (embedding size)

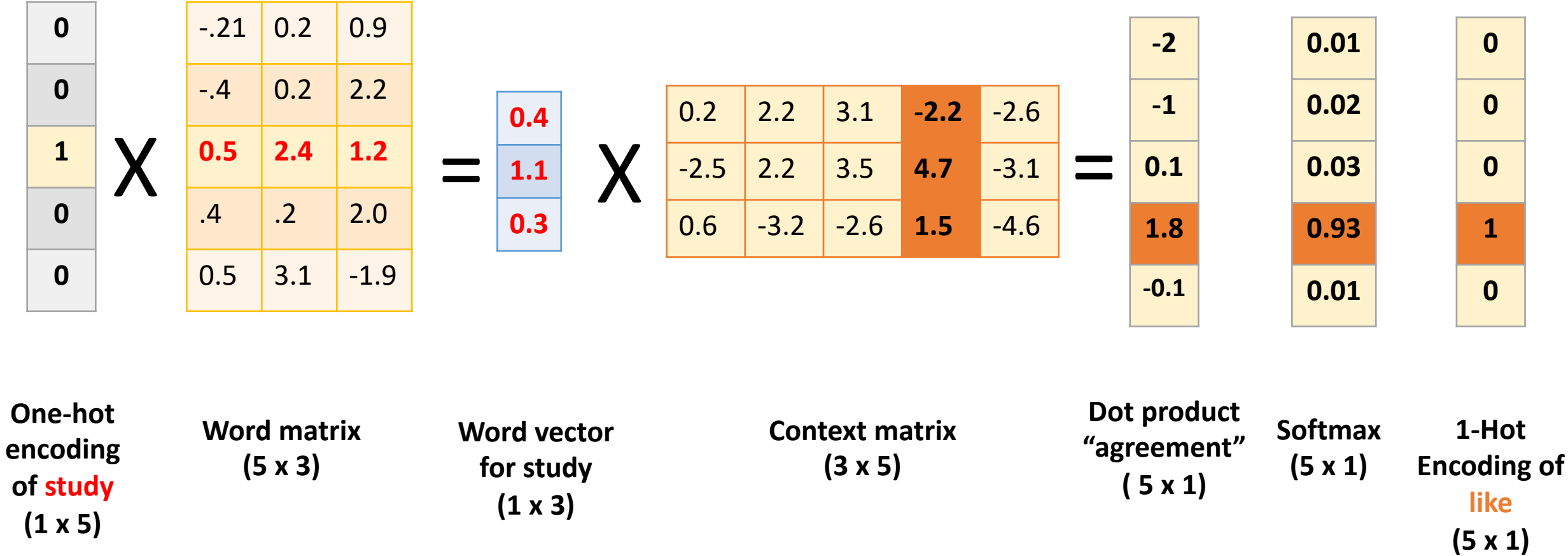
2<sup>nd</sup> epoch of training





V = 5  
M = 3 (embedding size)

3<sup>rd</sup> epoch of  
training  
(convergence)



$V = 5$

$M = 3$  (embedding size)

*vocab size*

cat	-.21	0.2	0.9
dog	-.4	0.2	2.2
study	<b>0.5</b>	<b>2.4</b>	<b>1.2</b>
like	.4	.2	2.0
tonight	0.5	3.1	-1.9

*representation of word for future models.*

This matrix now becomes our “word embeddings”. Each word in our vocabulary is now represented as a vector of numbers!

Word matrix  
(5 x 3)

*latent dimensions*