

V = 5

M = 3 (embedding size)

0
0
1
0
0

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

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)

0	X	-0.5	0.5	0.3	=	0.5
0		-0.3	0.2	0.1		1.2
1		0.5	1.2	0.9		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)

Word vector
for study
(1 x 3)

V = 5
M = 3 (embedding size)

0	X	-0.5	0.5	0.3	=	0.5	X	0.0	1.0	2.0	-1.0	-0.5
0		-0.3	0.2	0.1				-2.3	1.2	0.4	5.6	-1.0
1		0.5	1.2	0.9				0.1	-0.2	-0.5	0.4	-0.6
0		0.3	0.5	1.2								
0		0.1	1.0	-0.2								

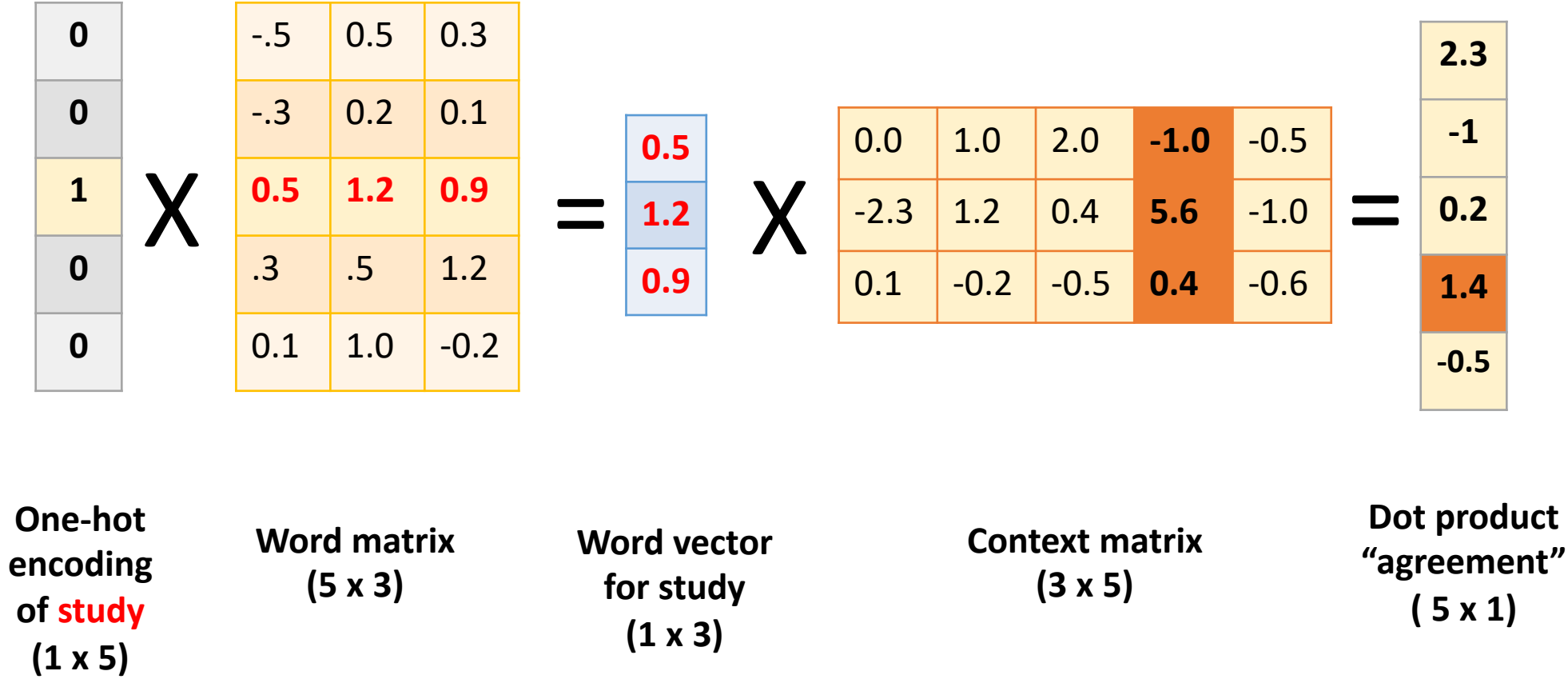
One-hot
encoding
of study
(1 x 5)

Word matrix
(5 x 3)

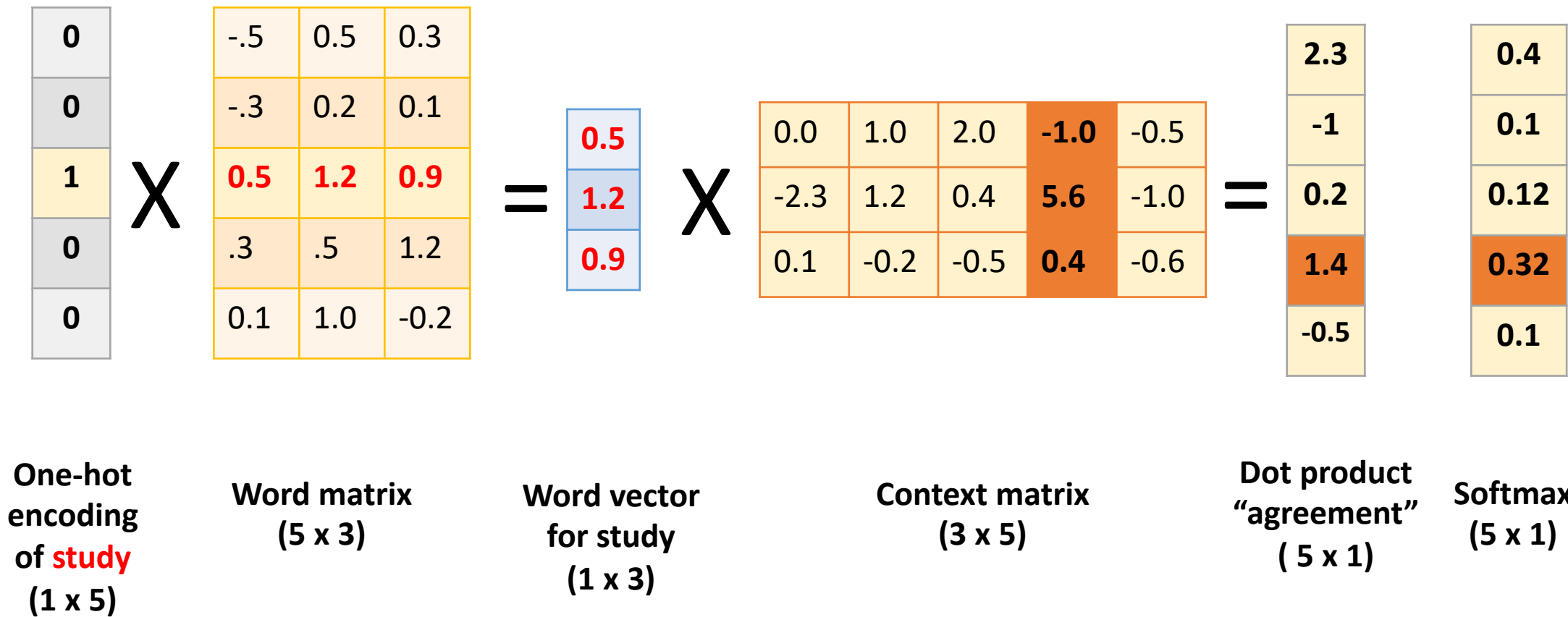
Word vector
for study
(1 x 3)

Context matrix
(3 x 5)

V = 5
M = 3 (embedding size)

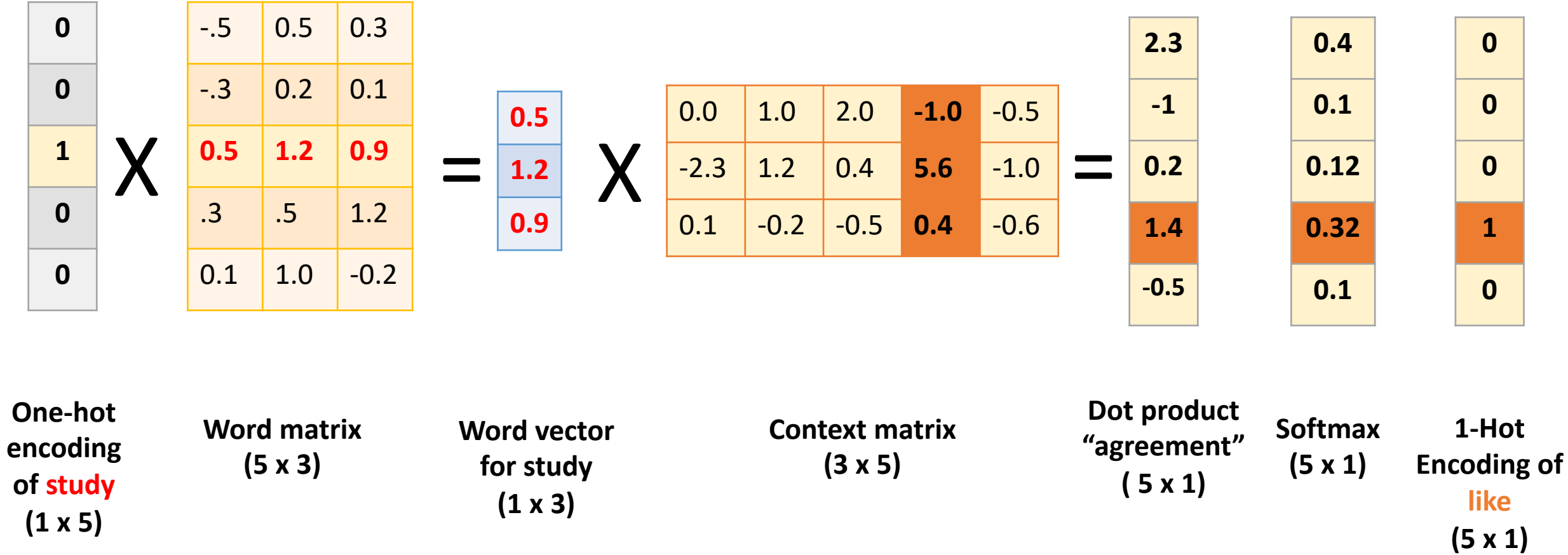


V = 5
M = 3 (embedding size)



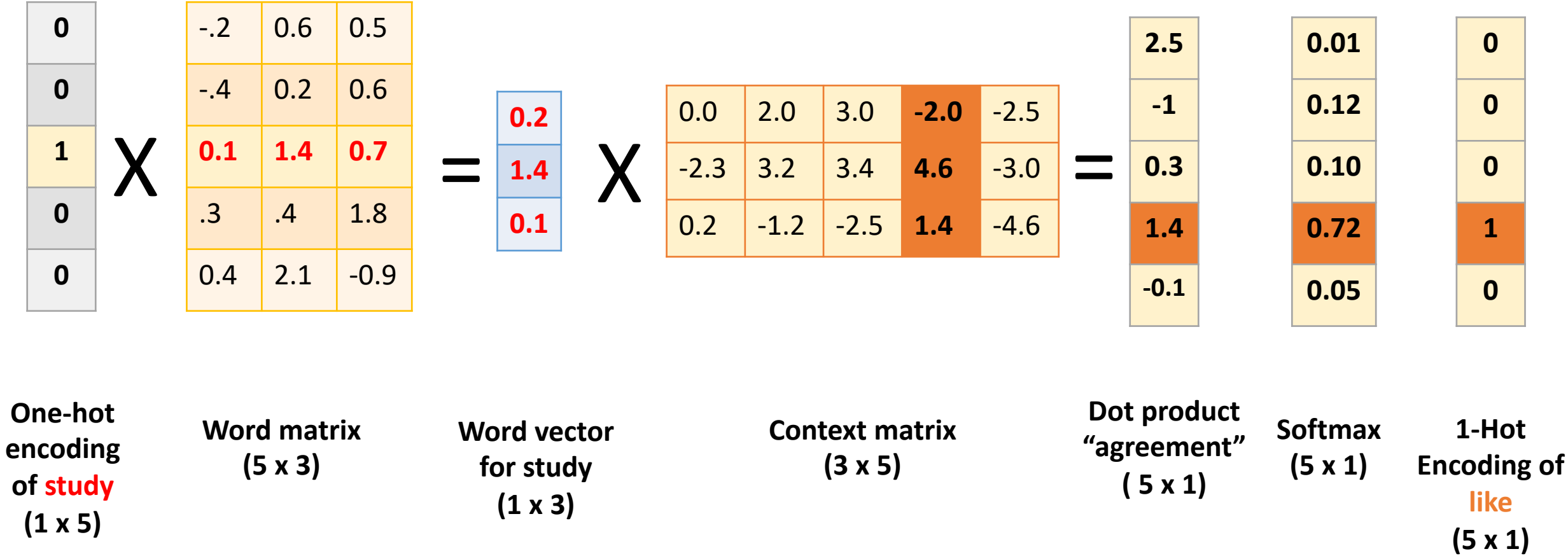
V = 5
M = 3 (embedding size)

1st epoch of training



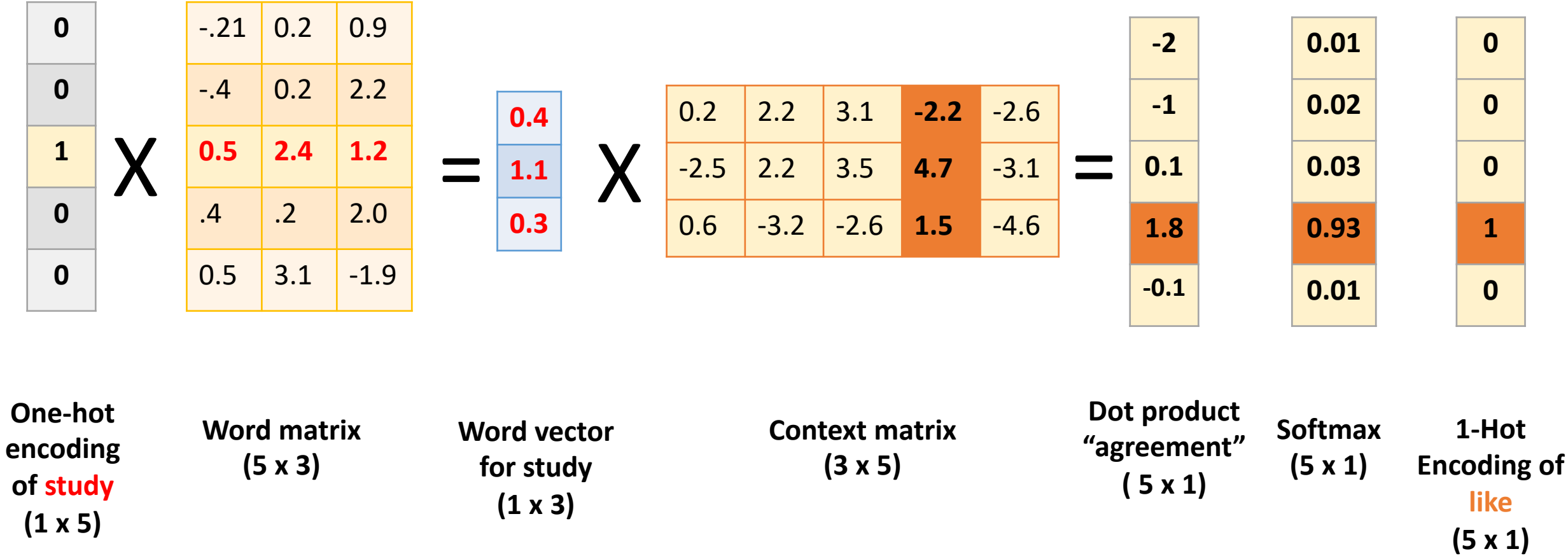
V = 5
M = 3 (embedding size)

2nd epoch of
training



V = 5
M = 3 (embedding size)

3rd epoch of
training
(convergence)



V = 5

M = 3 (embedding size)

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

Word matrix
(5 x 3)

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