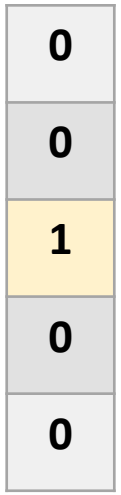
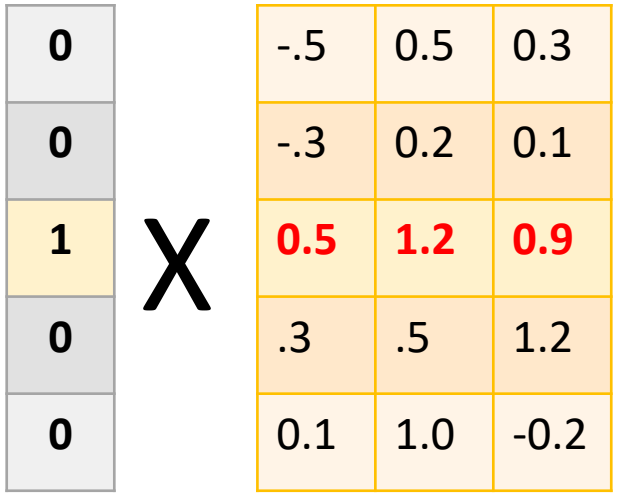
V = 5
M = 3 (embedding size)



One-hot encoding of study (1 x 5)

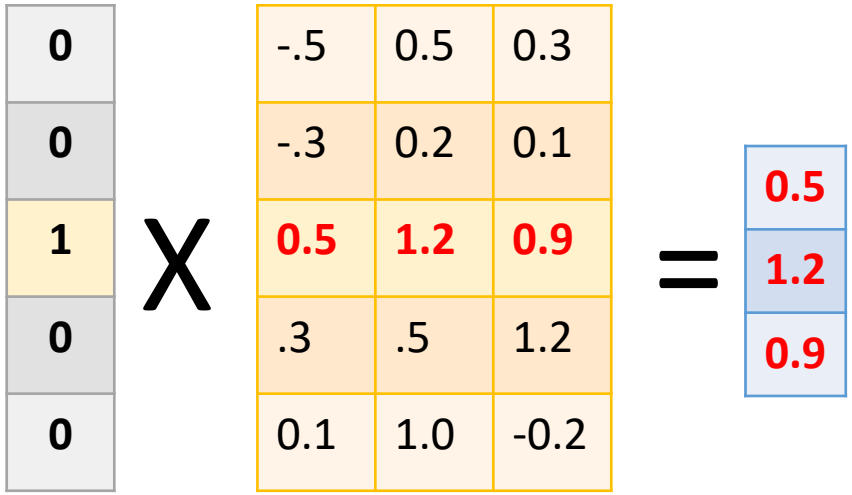
V = 5
M = 3 (embedding size)



One-hot encoding of study (1 x 5)

Word matrix (5 x 3)

V = 5
M = 3 (embedding size)

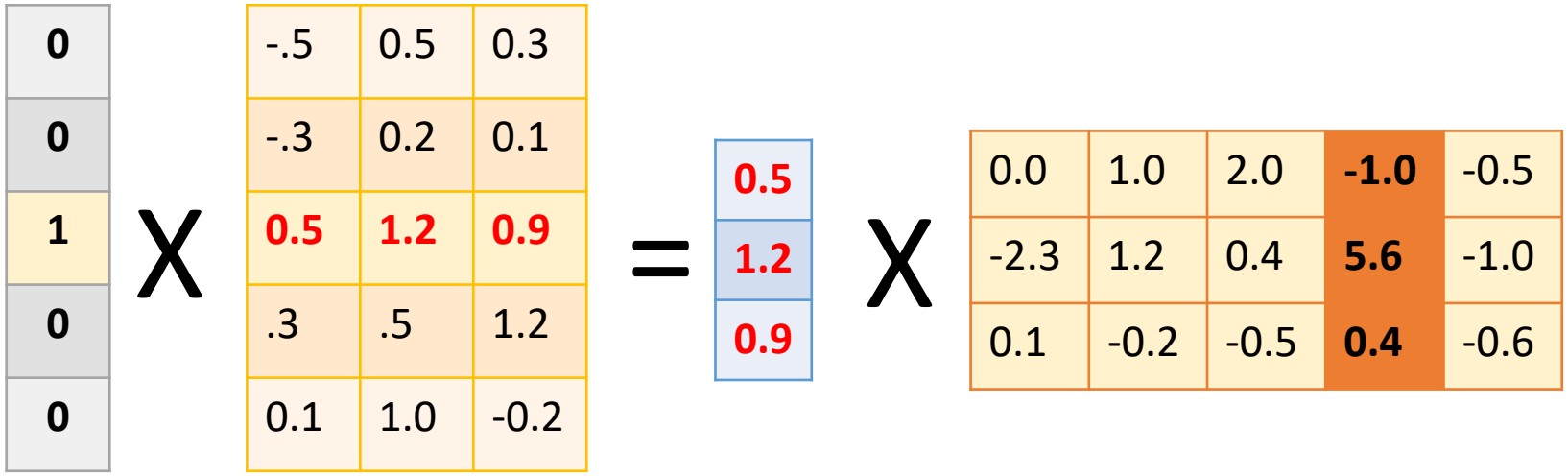


One-hot encoding of study (1 x 5)

Word matrix (5 x 3)

Word vector for study (1 x 3)

V = 5
M = 3 (embedding size)



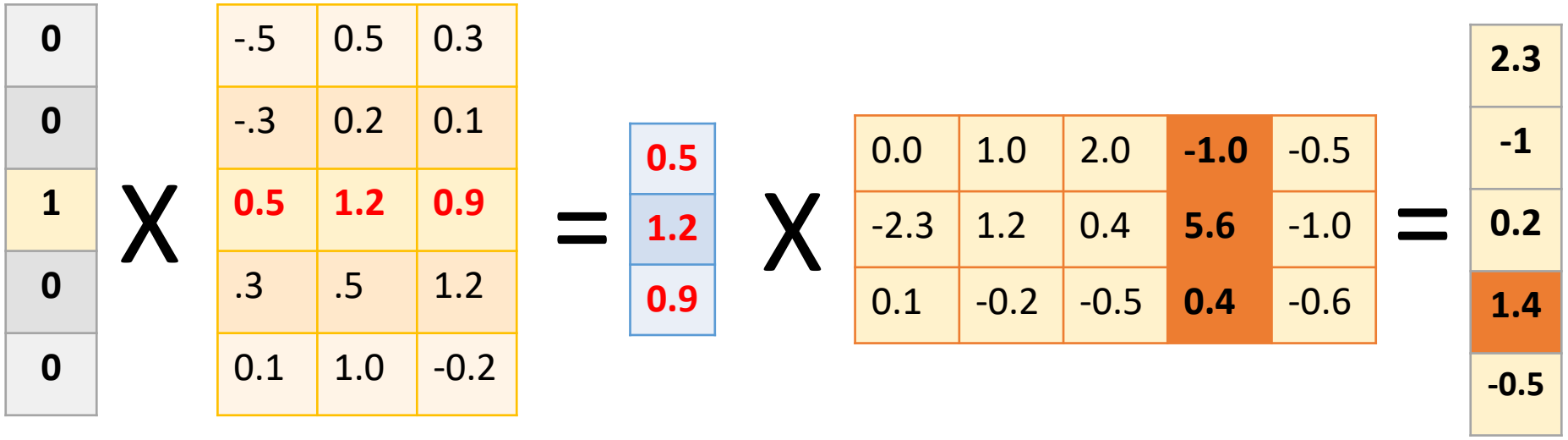
One-hot encoding of study (1 x 5)

Word matrix (5 x 3)

Word vector for study (1 x 3)

Context matrix (3 x 5)

V = 5
M = 3 (embedding size)



One-hot encoding of study (1 x 5)

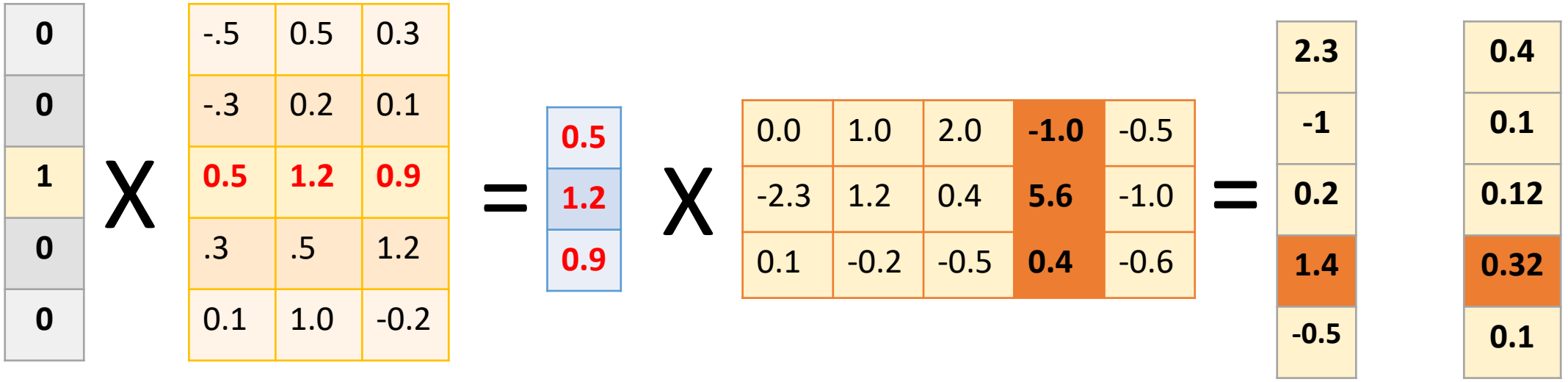
Word matrix (5 x 3)

Word vector for study (1 x 3)

Context matrix (3 x 5)

Dot product "agreement" (5 x 1)

V = 5
M = 3 (embedding size)



One-hot encoding of study (1 x 5)

Word matrix (5 x 3)

Word vector for study (1 x 3)

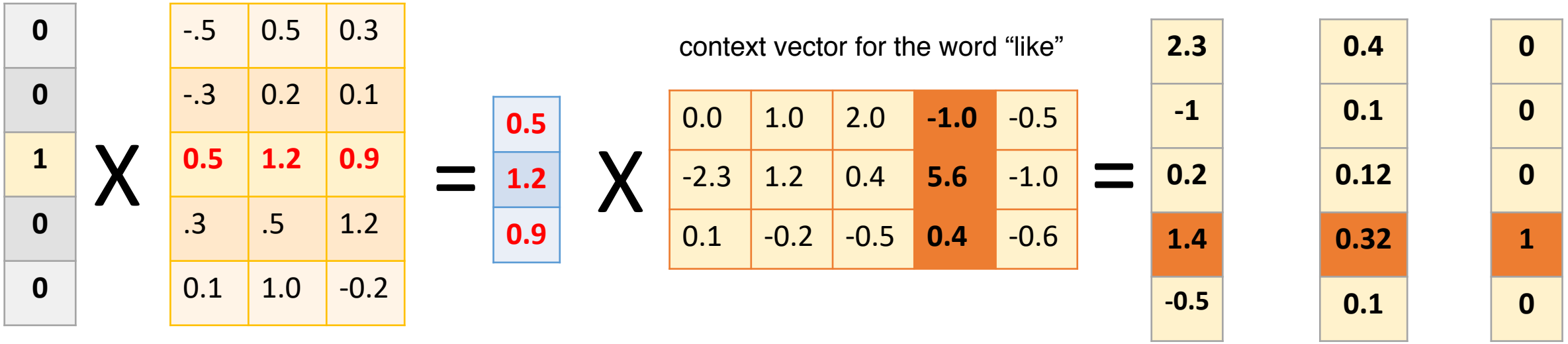
Context matrix (3 x 5)

Dot product "agreement" (5 x 1)

Softmax (5 x 1)

V = 5
M = 3 (embedding size)

1st epoch of training



how much “study” agrees with “like”
appear together

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)

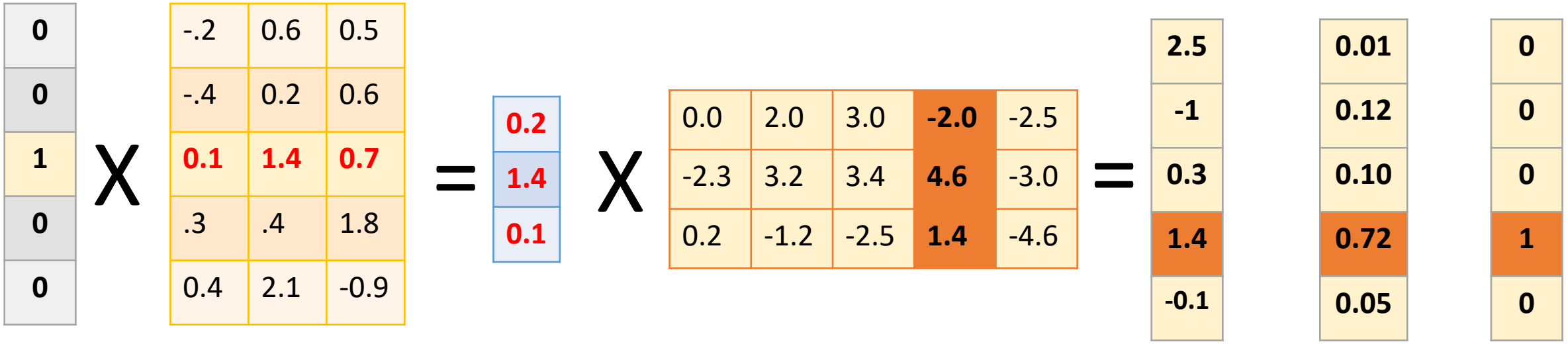
Dot product “agreement” (5 x 1)

Softmax (5 x 1)

1-Hot Encoding of like (5 x 1)

V = 5
M = 3 (embedding size)

2nd epoch of training



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)

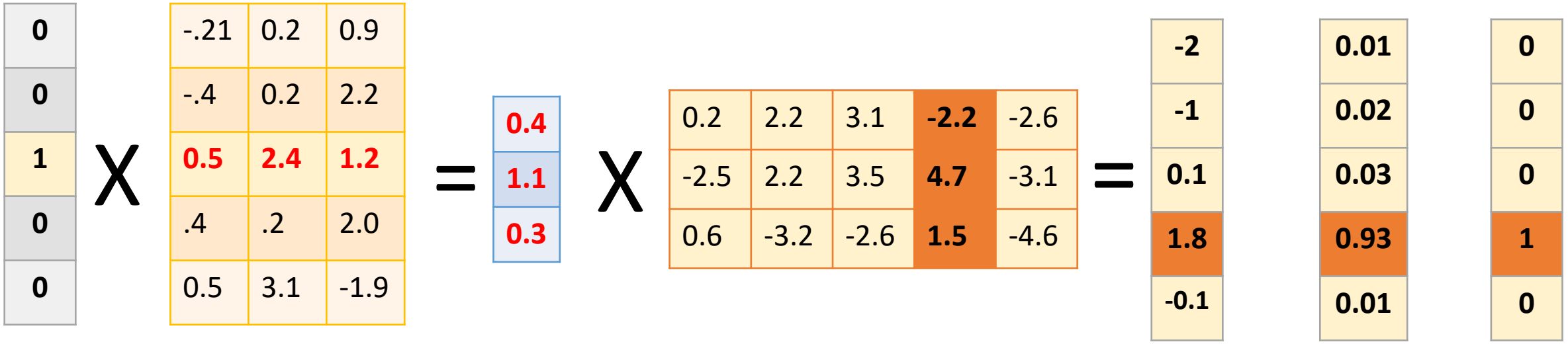
Dot product "agreement" (5 x 1)

Softmax (5 x 1)

1-Hot Encoding of like (5 x 1)

V = 5
M = 3 (embedding size)

3rd epoch of training (convergence)



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)

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!