One-hot encoding of study (1 x 5)

0		5	0.5	0.3
0		3	0.2	0.1
1	X	0.5	1.2	0.9
0		.3	.5	1.2
0		0.1	1.0	-0.2

One-hot encoding of study (1 x 5)

Word matrix (5 x 3)

0		5	0.5	0.3		
0		3	0.2	0.1		0.5
1	X	0.5	1.2	0.9	=	1.2
0		.3	.5	1.2		0.9
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)

0		5	0.5	0.3								
0		3	0.2	0.1		0.5		0.0	1.0	2.0	-1.0	-0.5
1	X	0.5	1.2	0.9	=	1.2	X	-2.3	1.2	0.4	5.6	-1.0
0		.3	.5	1.2		0.9	/ \	0.1	-0.2	-0.5	0.4	-0.6
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)

0		5	0.5	0.3										2.3
0		3	0.2	0.1		0.5		0.0	1.0	2.0	-1.0	-0.5		-1
1	X	0.5	1.2	0.9	=	1.2	X	-2.3	1.2	0.4	5.6	-1.0	=	0.2
0		.3	.5	1.2		0.9		0.1	-0.2	-0.5	0.4	-0.6		1.4
0		0.1	1.0	-0.2									-	-0.5

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)

0		5	0.5	0.3										2.3	0.4
0		3	0.2	0.1		0.5		0.0	1.0	2.0	-1.0	-0.5		-1	0.1
1	X	0.5	1.2	0.9	=	1.2	X	-2.3	1.2	0.4	5.6	-1.0	=	0.2	0.12
0		.3	.5	1.2		0.9	/ \	0.1	-0.2	-0.5	0.4	-0.6		1.4	0.32
0		0.1	1.0	-0.2										-0.5	0.1
		0.1	1.0	0.2										-0.5	0.1

One-hot encoding of study (1 x 5)

Word matrix (5 x 3)

Word vector for study (1 x 3)

Context matrix (3×5)

Dot product "agreement" (5 x 1)

Softmax (5 x 1)

0		5	0.5	0.3				conte	xt vecto	or for th	ne word	l "like"		2.3	0.4		0	
0		3	0.2	0.1		0.5		0.0	1.0	2.0	-1.0	-0.5		-1	0.1	,	0	
1	X	0.5	1.2	0.9	=	1.2	X	-2.3	1.2	0.4	5.6	-1.0	=	0.2	0.12	,	0	
0		.3	.5	1.2		0.9	/ \	0.1	-0.2	-0.5	0.4	-0.6		1.4	0.32		1	
0		0.1	1.0	-0.2									I	-0.5	0.1		0	
	how much "study" agrees with "like"									l								

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)

appear together

Dot product

"agreement"

(5 x 1)

Softmax (5 x 1)

1-Hot Encoding of like

(5 x 1)

One-hot

encoding

of study

(1 x 5)

Word matrix

(5 x 3)

Word vector

for study

(1 x 3)

Softmax

(5 x 1)

1-Hot

Encoding of

like

(5 x 1)

Dot product

"agreement"

(5 x 1)

0		2	0.6	0.5										2.5	0.01		0	
0		4	0.2	0.6		0.2		0.0	2.0	3.0	-2.0	-2.5		-1	0.12		0	
1	X	0.1	1.4	0.7	=	1.4	X	-2.3	3.2	3.4	4.6	-3.0	=	0.3	0.10	0	0	
0		.3	.4	1.8		0.1		0.2	-1.2	-2.5	1.4	-4.6		1.4	0.72		1	
0		0.4	2.1	-0.9									l	-0.1	0.05		0	
																Ĺ		

Context matrix

 (3×5)

3rd epoch of training (convergence)

0		21	0.2	0.9										-2	0.01		0	
0		4	0.2	2.2		0.4		0.2	2.2	3.1	-2.2	-2.6		-1	0.02		0	
1	X	0.5	2.4	1.2	=	1.1	X	-2.5	2.2	3.5	4.7	-3.1	=	0.1	0.03		0	
0		.4	.2	2.0		0.3	/ \	0.6	-3.2	-2.6	1.5	-4.6		1.8	0.93		1	
0		0.5	3.1	-1.9										-0.1	0.01		0	
	4				•											l		

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!