Name: **Fatima Usman**   Quiz 2   Roll #: **21L-5605**

Question: Calculate the PMI of the words "cat" and "mat". Apply Preprocessing and write all steps. Clearly write assumptions you make. Consider sliding window size = ±1. First create a co-occurrence matrix with respect to window size.
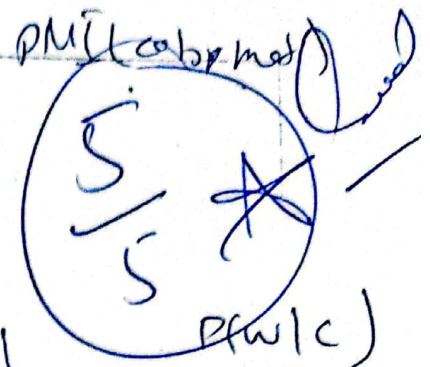
CORPUS:
D1: THE CAT SAT ON THE MAT.
D2: THE DOG SAT ON THE MAT.
D3: THE CAT AND DOG SAT BESIDE EACH OTHER ON THE MAT.
D4: CATS AND DOGS ARE OFTEN CONSIDERED AS PETS BY HUMANS

$\text{PMI}(cat, mat)$

$$\frac{5}{5} \quad \bcancel{\quad}$$

Adding due to required values

	THE	SAT	AND	CAT	MAT	
CAT	2	1	2	0	0	5
MAT	3	0	0	0	0	3
	5	1	2	0	0	8

$P(w|c)$

$c = 33$

P(M

$$\text{PMI}(cat, mat) = \log_e\left(\frac{P(cat, mat)}{P(cat) \cdot P(mat)}\right) = \frac{0/8}{(5/8)(8/8)}$$

$\log_2(0) = $ very minor value.

is very low,

PMI(cat, mat) = 0   As they both are target words they are not in the content of each other (according to given corpus corpus and sliding window size.

use
max(formul, 0)