

Processamento de Linguagem Natural – Teste de avaliação

Nome completo:

RA:

1. Calcule:

- **PPMI**(w=information, c=data)

- **PPMI**(w=digital, c=data)

dada a seguinte matriz termo-contexto.

	computer	data	pinch	result	sugar
apricot	0	0	1	0	1
pineapple	0	0	1	0	1
digital	2	1	0	1	0
information	1	6	0	4	0

$$pmi_{ij} = \log_2 \frac{P_{ij}}{P_{i*}P_{*j}}$$

$$P_{ij} = \frac{f_{ij}}{\sum_{i=1}^W \sum_{j=1}^C f_{ij}}$$

$$P_{i*} = \frac{\sum_{j=1}^C f_{ij}}{\sum_{i=1}^W \sum_{j=1}^C f_{ij}}$$

$$P_{*j} = \frac{\sum_{i=1}^W f_{ij}}{\sum_{i=1}^W \sum_{j=1}^C f_{ij}}$$

$$ppmi_{ij} = \begin{cases} pmi_{ij} & \text{if } pmi_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$