Positive tweets

I am happy because I am Learning NLP

I am happy, not sad.

Negative tweets

I am sad, I am not learning NLP

I am sad, not happy.

Frequency table

word	Pos	Neg
I	3	3
am	3	3
happy	2	1
because	1	0
Learning	1	1
NLP	1	1
Sad	1	2
not	1	2
N_{class}	13	12

word	Pos	Neg
I	$\frac{3}{13} = 0.23$	0.25
am	0.23	0.25
happy	0.15	0.08
because	0.08	0.
Learning	0.08	0.08
NLP	0.08	0.08
Sad	0.08	0.17
not	0.08	0.17

I am happy today; I am learning

$$\prod_{i=1}^{m} \frac{P(W_i \mid Pos)}{P(W_i \mid neg)} = \frac{0.23}{0.25}\ \frac{0.23}{0.25}\ \frac{0.15}{0.08}\ 1\ \frac{0.23}{0.25}\ \frac{0.23}{0.25}\ \frac{0.08}{0.08}$$

$$= 1.34 > 1 \quad \text{positive!}$$

Laplacian Smoothing

usually we compute

$$P(w_i \mid class) = \frac{freq(w_i, class)}{N_{class}}, \quad class \in \{Positive, Negative\}$$

However, if a word does not appear in the traing, it get a probability of 0. to fix this we add smoothing

$$P(w_i \mid class) = \frac{freq(w_i, class) + 1}{N_{class} + V}$$

V : number of unique words in vocabulary

N : number of total frequency in each class

Likelihood ratio

For a word w_i, $\quad ratio(w_i) = \frac{P(w_i \mid Pos)}{P(w_i \mid Neg)}$

$$\approx \frac{freq(w_i, 1) + 1}{freq(w_i, 0) + 1}$$

For $N_{positive} \approx N_{negative}$.

To infer sentiment of a tweet, we can compute

$$\frac{P(Pos)}{P(neg)} \prod_{i=1}^{m} \frac{P(w_i \mid Pos)}{P(w_i \mid neg)}$$

Log likelihood

$$\log \frac{P(Pos)}{P(neg)} + \sum_{i=1}^{m} \log \frac{P(w_i \mid Pos)}{P(w_i \mid neg)} \begin{cases} > 0 & positive \\ < 0 & negative \end{cases}$$