# Naïve Bayes Example

- Reference: https://web.stanford.edu/~jurafsky/slp3/, Chapter 3.

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

| 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 | 0.24 | 0.25 |
| am | 0.24 | 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 |

Let's classify the following tweets as positive or negative:

1. I am not sad.
2. I am learning NLP.

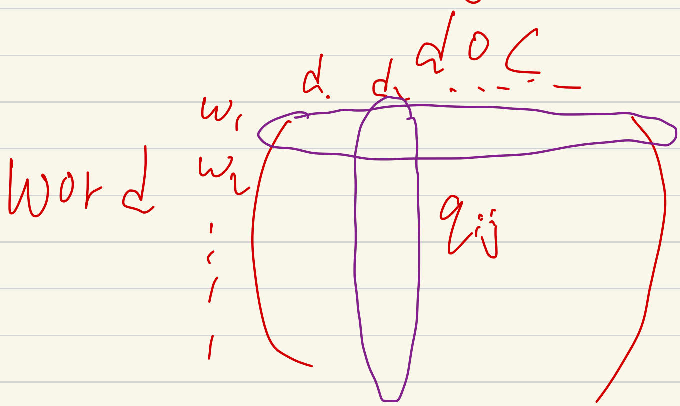
$$LR = \frac{Pr[Pos] \cdot Pr[I \mid Pos] \cdot Pr[am \mid Pos] \cdot Pr[not \mid Pos] \, Pr[sad \mid Pos]}{Pr[Neg] \cdot Pr[I \mid Neg] \cdot Pr[am \mid Neg] \cdot Pr[not \mid Neg] \cdot Pr[sad \mid Neg]}$$

$$= \frac{13}{12} \cdot \frac{0.24}{0.25} \cdot \frac{0.24}{0.25} \cdot \frac{0.08}{0.17} \cdot \frac{0.08}{0.17} < 1$$

Negative

(1) Pre-processing
- Tokenization
- Remove tags/Stop-words
- Stemming

(2) TF-IDF $w_{ij} = tf_{ij} \times \log \frac{N}{df_i}$




(3), N-Gram Model

$$Pr[w_n | w_1 w_2 \dots w_{n-1}] = P$$

Uni-gram: $P = Pr[w_n]$

Bi-gram: $P = Pr[w_n | w_{n-1}]$ (Markov Chain)

N-gram: $P = Pr[w_n | w_{n-N+1}, \dots w_{n-1}]$

$$\Pr[w_n \mid w_{n-1}] = \frac{C(w_{n-1}, w_n)}{C(w_{n-1})} \quad (\text{MLE})$$

$w_{n-1}$ only appears in the testing set.

Laplace Smoothing: $\dfrac{C(w_{n-1} w_n)+1}{C(w_{n-1})+|V|}$

---

Naive Bayes

Input: $d$

Output: $c \in \{c_1, c_2, \cdots c_j\}$

$$\Pr[c \mid d] = \frac{\Pr[d \mid c]\Pr[c]}{\Pr[d]}$$

$d = w_1 w_2 \cdots w_n$

$$\approx \underbrace{\Pr[w_1 \mid c] \Pr[w_2 \mid c] \cdots \Pr[w_n \mid c] \Pr(c)}$$

$$c_{NB} \in \arg\max_c$$

emb_dim

|V| Word Embedding_1

$v_1 \in \mathbb{R}^{emb\_dim}$

$v_2$

$\vdots$

$v_{|V|}$

Output of word embedding

centerword

$\underbrace{O_{-m} \cdots O_{-1}}_{m \text{ words}} \mid C \mid \underbrace{O_1\, O_2 \cdots O_m}_{m \text{ words}}$

context words.

$$\hat{v} = \frac{1}{2m} \cdot \sum_{j=1}^{m} \left( v_{O_j} + v_{O_{-j}} \right)$$

emb_dim

|V| Word_Embedding_2

$u_1^T$

$u_2^T$

$\vdots$

$u_{|V|}^T$

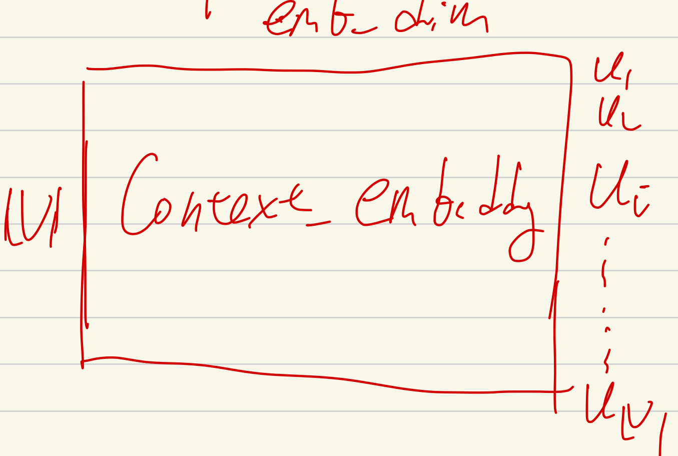
$$\Pr\left[ C \mid O_{-m}, O_{-(m-1)}, \cdots, O_{-1}, O_1, O_2 \cdots O_m \right] = \frac{\exp(u_c^T \cdot \hat{v})}{\sum_{j \in V} \exp(u_j^T \cdot \hat{v})}$$

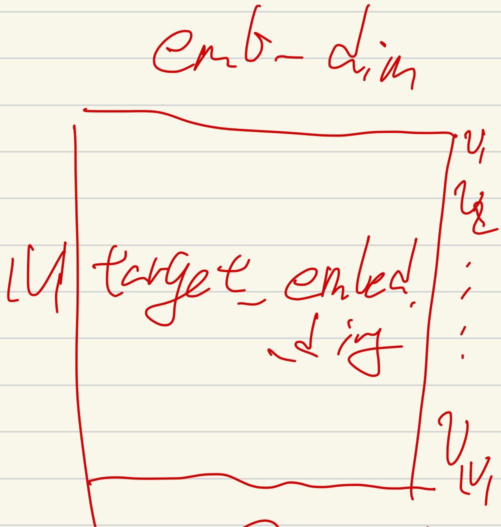
$$\prod_{c \in d} \frac{\exp(u_c^T \cdot v)}{\sum_{j \in V} \exp(u_j^T \cdot v)}$$

depends on c

---

# Skip-Gram




| | | | | | | |
|---|---|---|---|---|---|---|
| $o_{-m}$ | ... | $o_{-1}$ | $c$ | $o_1$ | ... | $o_m$ |

$$\sum_{c \in d} \sum_{\substack{j=-m \\ j \neq 0}}^{m} \left( \log \frac{\exp(u_{o_j}^T \cdot v_c)}{\underbrace{\sum_{i \in V} \exp(u_i^T v_c)}_{\text{computationally too heavy}}} \right)$$

① usually we use target_emb, $v$

② $\frac{u+v}{2}$

③ $(u^T, v^T)^T$

Sample $V_{ns} \subseteq V$

$$\sum_{c \in C_d} \sum_{\substack{j=-m \\ j \neq 0}}^{m} \log \left( \frac{\exp(u_{o_j}^T \cdot v_c)}{\sum_{i \in V_{ns}} \exp(u_i^T \cdot v_c)} \right)$$