Homework 2 //

Group 30, Marko Horvatić, Soumic Akkaraju

2.1. Theory

$$P(w_1, w_2, \dots w_n) = P(w_1)P(w_2)\dots P(w_n) = \prod_{i=1}^{n} P(w_i)$$

$$P(w_i \mid Pos) = \frac{C(w_i, Pos) + k}{\sum_{w \in V} C(w, Pos) + kV}$$

$$Class(R) = \underset{C}{\operatorname{argmax}}\ P(C \mid R)$$

$P(boring \mid Pos) = ?$

$$P(boring \mid Pos) = \frac{C(boring, Pos) + 1}{\sum_{w \in V} C(w, Pos) + V}$$

$$= \frac{0 + 1}{5 \cdot 1 + 3 \cdot 1 + 7 \cdot 5 + 25 \cdot 1 + 49} = \frac{1}{92} //$$

$P(\text{"intriguing yet disappointing "} \mid Pos)$

$= P(\text{"intriguing"} \mid Pos) \cdot P(\text{"yet "} \mid Pos) \cdot P(\text{"disappointing"} \mid Pos)$

$$= \frac{1+1}{92} \cdot \frac{1}{92} \cdot \frac{1}{92} = \frac{4}{(92)^3} = \frac{1}{194672} //$$

$P(\text{"intriguing yet disappointing "} \mid Neg)$

$$= \frac{1}{34+49} \cdot \frac{1}{34+49} \cdot \frac{2}{34+49} = \frac{1}{83} \cdot \frac{1}{83} \cdot \frac{2}{83} = \frac{2}{(83)^3}$$

$P(Neg) > P(Pos) \Rightarrow C = \text{"Neg"} //$

3.1 Theory

$$P(w_1, w_2 \dots w_n) = P(w_1 \mid \langle S \rangle) * \prod_{i=1}^{n} P(w_i \mid w_{i-1})$$

↳ Bigram model

Q: How do compute $P(f_i \mid t)$ when f_i is a bigram?

A:

$$P(f_i \mid t) = \frac{C(f_i, t) + k}{\sum_{f_i \in F} C(f_i, t) + k \cdot F}$$, where F is a set of all the bigrams (bigram vocabulary)

Q: $P(movie \mid great)$, $P(enough \mid familiar)$ being positive

A: $P(movie \mid great) = \frac{1+1}{2+1} = \frac{2}{3}$ $\quad |F_{BIGRAM}| = 35$

$$P(enough \mid familiar) = \frac{0+1}{1+1} = \frac{1}{2}$$

$$P(\text{"great movie"} \mid +) = \frac{1+1}{36+39} = \frac{2}{75}$$

$$P(\text{"familiar enough"} \mid +) = \frac{0+1}{36+39} = \frac{1}{75}$$

Q: $P(\text{"the movie is intriguing"} \mid Pos) = ?$

$P(\text{"the movie is intriguing"} \mid Pos) =$

$= P(\text{"the movie"} \mid Pos) \cdot P(\text{"movie is"} \mid Pos) \cdot P(\text{"is intriguing"} \mid Pos)$

$$= \frac{1+1}{75} \cdot \frac{0+1}{75} \cdot \frac{1+1}{75} = \frac{4}{(75)^3}$$