第 7 节练习：情感分析

From Languages to Information CS124
—— Group Work on Naïve Bayes and Sentiment Analysis
http://web.stanford.edu/class/cs124/lec/nbsection19.html

拉普拉斯平滑（Laplace Smoothing）

拉普拉斯平滑，又称 Add -1 Smoothing

- 核心思想是取代MLE，估计$P(w) = \frac{C(w)}{N}$

拉普拉斯平滑

- 拉普拉斯平滑是拉普拉斯在研究太阳升起问题时提出的
- 其中，第 $n+1$ 天，我们观察到之前太阳升起了 s 次
- 则今天太阳升起的概率为：$p_{Lap} =$

$$(S_{n+1}=1|S_1+\cdots+S_n=s)=\frac{s+1}{n+2}$$

拉普拉斯平滑

- 拉普拉斯平滑既然被称作加一平滑，自然是每个计数上都要加一
- 我们以一个词典为例，从一个大小为 V 的词典里，独立地取出 N 个词，则有一元概率：$P(w_i) = \frac{C(w_i)}{\sum_j C(w_j)} = \frac{C(w_i)}{N}$
- 每个计数上加一： $P(w_i) = \frac{C(w_i)+1}{\sum_j (C(w_j)+1)} = \frac{C(w_i)+1}{N+V}$

拉普拉斯平滑结果

Original:

	i	want	to	eat	chinese	food	lunch	spend
i	5	827	0	9	0	0	0	2
want	2	0	608	1	6	6	5	1
to	2	0	4	686	2	0	6	211
eat	0	0	2	0	16	2	42	0
chinese	1	0	0	0	0	82	1	0
food	15	0	15	0	1	4	0	0
lunch	2	0	0	0	0	1	0	0
spend	1	0	1	0	0	0	0	0

Smoothed:

	i	want	to	eat	chinese	food	lunch	spend
i	6	828	1	10	1	1	1	3
want	3	1	609	2	7	7	6	2
to	3	1	5	687	3	1	7	212
eat	1	1	3	1	17	3	43	1
chinese	2	1	1	1	1	83	2	1
food	16	1	16	1	2	5	1	1
lunch	3	1	1	1	1	2	1	1
spend	2	1	2	1	1	1	1	1

练习

We want to build a naïve bayes sentiment classifier using add -1 smoothing, as described in the lecture (not binary naïve bayes, regular naïve bayes). Here is our training corpus:

Training Set:

```
- just plain boring
- entirely predictable and lacks energy
- no surprises and very few laughs
+ very powerful
+ the most fun film of the summer
```

Test Set:

```
predictable with no originality
```

Questions

1. Compute the prior for the two classes + and -, and the likelihoods for each word given the class (leave in the form of fractions).

2. Then compute whether the sentence in the test set is of class positive or negative (you may need a computer for this final computation).

Questions

3. Would using binary multinomial Naïve Bayes change anything?

4. Why do you add $|V|$ to the denominator of add-1 smoothing, instead of just counting the words in one class?

Questions

5. What would the answer to question 2 be without add-1 smoothing?

Answers

1. **Compute the prior for the two classes + and -, and the likelihoods for each word given the class (leave in the form of fractions).**

|V| = 20, n- = 14, n+ = 9
P(-) = 3/5, P(+) = 2/5
P(and | -) = (2 + 1) / (14 + 20) = 3/34
P(any_other_vocab_word_in_-_sentence | -) = (1 + 1) / (14 + 20) = 2/34, e.g. P('plain' | -)
P(any_vocab_word_not_in_-_sentence | -) = (0 + 1) / (14 + 20) = 1/34, e.g. P('powerful' | -), P('with' | -)
P(the | +) = (2 + 1) / (9 + 20) = 3/29
P(any_other_vocab_word_in_+_sentence | +) = (1 + 1) / (9 + 20) = 2/29, e.g. P('powerful' | +)
P(any_vocab_word_not_in_+_sentence | +) = (0 + 1) / (9 + 20) = 1/29, e.g. P('plain' | +), P('with' | +)

Answers

2. **Then compute whether the sentence in the test set is of class positive or negative (you may need a computer for this final computation).**

C = {+, -}
*P(c | "predictable with no originality") ∝ P(c) * P ("predictable with no originality" | c) = P(c) * P(predictable | c) * P(with | c) * P(no | c) * P(originality | c) ~= P(c) * P(predictable | c) * P(no | c), 'with' and 'originalty' are unknown*
*P(- | "predictable with no originality") = (3/5) * (2/34) * (2/34) = 0.002076*
*P(+ | "predictable with no originality") = (2/5) * (1/29)^2 = 0.0004756*
P(- | "predictable with no originality") is greater, so the test set sentence is classified as class negative.

Answers

3. **Would using binary multinomial Naïve Bayes change anything?**

*No, using binary NB would not change anything - under this scheme n+ = 8, P(+ | "predictable with no originality") = (2/5) * (1/28)^2 = 0.0005102, which is still less than P(- | "predictable with no originality").*

4. **Why do you add $|V|$ to the denominator of add-1 smoothing, instead of just counting the words in one class?**

In add-1 smoothing we assume we have seen each word once regardless of whether they appear in the original class or not and thus add |V| to the denominator. Note that words that do not appear in the train set are 'unk' and are not included in the vocab.

Answers

5. **What would the answer to question 2 be without add-1 smoothing?**

P(c | "predictable with no originality") = 0 for class positive because at least one of the words in the test set does not appear in the positive train examples; P('predictable' | +) = P('no' | +) = 0.

Thank you!