Feedback — Text Classification and Sentiment Analysis

Help

You submitted this quiz on **Tue 27 Mar 2012 12:06 AM PDT**. You got a score of **4.00** out of **4.00**.

Question 1

Assume the following probabilities for each word being part of a *positive* or *negative* movie review.

	pos	neg
I	0.13	0.29
always	0.08	0.06
like	0.19	0.03
foreign	0.07	0.21
films	0.11	0.14

Now consider the sentence...

```
I always like foreign films.
```

Using Naïve Bayes and assuming equal prior probability for each class, will we classify this sentence as being part of a *positive* or *negative* review?

Your Answer		Score	Explanation
○ pos			
◉ neg	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

With Naïve Bayes as a language model, we're looking for the class that maximizes the probability of the sentence. (See lecture slide #41 in the slide set on Text Classification and Naïve Bayes.) Because we're looking for the argmax, we can ignore the denominator (the likelihood of the sentence, in general), and the question states that the classes have equal prior probability, so we simply find the product of the likelihoods of each word, i.e., multiply each column and take the larger.

Question 2

Suppose we have the following short movie reviews, each labeled with a genre, either *comedy* or *action*.

1. fun, couple, love, love **comedy**
2. fast, furious, shoot **action**
3. couple, fly, fast, fun, fun **comedy**
4. furious, shoot, shoot, fun **action**
5. fly, fast, shoot, love **action**

Now consider the following new review...

```
fast, couple, shoot, fly
```

Using a simple $Naïve\ Bayes$ approach with $Laplace\ Smoothing$, how would we classify this, *comedy* or *action*? ***(Hint: Don't forget to include the prior.)***

Your Answer		Score	Explanation
◉ action	✔	1.00	
○ comedy			
Total		1.00 / 1.00	

Question Explanation

(See lecture slide #44 in the slide set on Text Classification and Naïve Bayes.) If *r* is our new review, we're looking for the class (genre) that maximizes (argmax) *P(c|r)*, where *P(c|r)* is proportional to the prior probability of the class, *P(c)*, times the product of *P(w|c)* for each word in *r*. *P(c)* is the number of reviews for each genre, divided by the total number of reviews. *P(w|c)* is the count of word *w* in class *c* plus 1 (i.e., smoothed), divided by the total count for *w* plus the vocabulary size (the number of unique words).

Question 3

As part of learning a sentiment lexicon for analyzing movie reviews, you decide to use the $TurneyPolarity$method to compute the (real-valued) polarity of the phrase $''special\ effects''$ using $PointwiseMutualInformation$relative to the sentiment words $''good''$ and $''bad''$.

Here are some actual counts from a corpus of IMDB reviews, where **NEAR** means "within 10 words of".

	count
corpus size (*N*)	1,595,494
unique types (*V*)	46,060
"special effects"	437
"good"	3124
"bad"	1791
"special effects" **NEAR** "good"	36
"special effects" **NEAR** "bad"	18

Your task is to calculate the value of $Turney's Polarity(''special effects'')$ with regard to $''good''$ and $''bad''$.

(signed numerical response rounded to the nearest tenth, e.g., 0.1, –0.7, etc.)

You entered:

0.2

Your Answer		**Score**	**Explanation**
0.2	✔	1.00	Option explanation
Total		1.00 / 1.00	

Question Explanation

See lecture slide #60 in the section "Learning Sentiment Lexicons". For this calculation, total size of corpus, vocabulary size, and standalone count of "special effects" are not needed.
*Polarity("special effects") = log_2[(count("special effects" NEAR "good") * count("bad")) / (count("special effects" NEAR "bad") * count("good"))].*

Question 4

Once again we'll look at sentiment in movie reviews, but this time we'll see if our analysis yields a different result with standard $Naïve Bayes$ vs. a $Binarized(Boolean feature)$ $Naïve Bayes$ approach.

We train our classifier with documents having the following ***counts*** for key sentiment words, with *positive* or *negative* class assigned as noted.

	"good"	*"poor"*	*"great"*	(class)
d1.	3	0	3	*pos*
d2.	0	1	2	*pos*

d3. 1	3	0	*neg*
d4. 1	5	2	*neg*
d5. 0	2	0	*neg*

Now consider the following new review.

```
A good, good plot and great characters, but poor acting.
```

Your task is to assign *positive* or *negative* sentiment first using standard $Naïve\,Bayes$ then with a $Binarized Naïve\,Bayes$ approach. Do the different approaches yield the same result in this case? Which of the following matches the pattern of your results? *(Hint: Use Add-1 smoothing with **both** methods.)*

(response: [*standard*] / [*binarized*])

Your Answer		Score	Explanation
○ pos / neg			
○ neg / pos			
○ neg / neg			
◉ pos / pos	✔	1.00	Option explanation
Total		1.00 / 1.00	

Question Explanation

(See lecture slide on "Binarized (Boolean feature) Multinomial Naïve Bayes".) Use the same method in each case, except that for the $Binarized$ approach, only count each word a maximum of once per document, i.e., transform the table of counts above to 0 or 1 for each cell.