Practice Exercises: Naïve Bayes Model

Exercise 1. (30 points) Assume the following likelihoods for each word being part of a positive or negative movie review, and equal prior probabilities for each class.

	pos	neg
I	0.09	0.16
always	0.07	0.06
like	0.29	0.06
foreign	0.04	0.15
films	0.08	0.11

What class will Naive Bayes assign to the sentence "I always like foreign films."?

Answer:

$$P(\text{pos}|\text{I always like foreign films}) \propto P(\text{pos}) \times P(\text{I}|\text{pos}) \times P(\text{always}|\text{pos}) \times P(\text{like}|\text{pos}) \times P(\text{foreign}|\text{pos}) \times P(\text{films}|\text{pos}) = 0.5 \times 0.09 \times 0.07 \times 0.29 \times 0.04 \times 0.08$$

We will use log-space to avoid underflow problem

$$\log\big(P(\text{pos}|\text{I always like foreign films})\big) = \log(0.5) + \log(0.09) + \log(0.07) + \log(0.29) + \log(0.04) + \log(0.08) = -5.534$$

Let's calculate for neg class

$$\log\big(P(\text{neg}|\text{I always like foreign films})\big) \propto \log(P(\text{neg})) + \log(P(\text{I}|\text{neg})) + \log(P(\text{always}|\text{neg})) + \log(P(\text{like}|\text{neg})) + \log(P(\text{foreign}|\text{neg})) + \log(P(\text{films}|\text{neg}))$$
$$= \log(0.5) + \log(0.16) + \log(0.06) + \log(0.06) + \log(0.15) + \log(0.11) = -5.32$$

Since $-5.32 > -5.534$, the class that Naïve Bayes model assign to the sentence is negative.

Note: To be graded, you need to explain your answer. Just saying the class is negative will not be counted.

Exercise 2. (30 points) Given 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**

and a new document D:
fast, couple, shoot, fly

compute the most likely class for D. Assume a naive Bayes classifier and use add-1 smoothing for the likelihoods.

Answer:

Use **c** to denote the category **comedy**, use **a** to denote the category **action**

The prior probabilities for categories are as follows.

$$P(\text{a}) = \frac{3}{5} = 0.6$$
$$P(\text{c}) = \frac{2}{5} = 0.4$$

The vocabulary contains {fun, couple, love, fast, furious, shoot, fly}

So $|V| = 7$

Let's construct the count table from the corpus.

	c=comedy	c=action
fun	3	1

couple	2	0
love	2	1
fast	1	2
furious	0	2
shoot	0	4
fly	1	1
Total	9	11

Let's calculate the likelihoods using add-1 smoothing.

$$P(\text{fun}|\text{a}) = \frac{\text{count}(\text{fun}, a) + 1}{(\sum_{w \in V} \text{count}(w, a)) + |V|} = \frac{1+1}{11+7} = \frac{2}{18} = 0.11$$

We calculate the other words and get the likelihoods for each word being part of category **action** and **comedy** as follows.

	c=comedy	c=action
fun	0.25	0.11
couple	0.1875	0.055
love	0.1875	0.11
fast	0.125	0.167
furious	0.0625	0.167
shoot	0.0625	0.278
fly	0.125	0.11

Let's consider the document D = fast, couple, shoot, fly

$P(\text{c}|\text{fast, couple, shoot, fly}) \propto\propto P(c) \times P(\text{fast}|c) \times P(\text{couple}|c) \times P(\text{shoot}|c) \times P(\text{fly}|c) = 0.4 * 0.125 * 0.1875 * 0.0625 * 0.125 = 7.32\text{e-}5$

$$P(\text{a}|\text{fast, couple, shoot, fly}) \propto P(a) * P(\text{fast}|a) * P(\text{couple}|a) * P(\text{shoot}|a) * P(\text{fly}|a) = 0.6 * 0.167 * 0.055 * 0.278 * 0.11 = 0.000168$$

$P(\text{a}|\text{D}) > P(c|D)$, so, the most likely class for D is **action**

Note: Again, just say the most likely class for D is action without any explanation will not give you any credit.

Exercise 3. (40 points) Assume that we train two models, multinomial naive Bayes and binarized naive Bayes, both with add-1 smoothing, on the following

document counts for key sentiment words, with positive or negative class assigned as noted.

doc	"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

Use both naive Bayes models to assign a class (pos or neg) to this sentence:
`A good, good plot and great characters, but poor acting.`

Do the two models agree or disagree?

Answer

From the corpus, let's build the count table that show the number of occurrences of words in each class using multinomial naive Bayes and binarized naive Bayes.

	NB counts		**Binary counts**	
	pos	neg	pos	neg
good	3	2	1	2
poor	1	10	1	3
great	5	2	2	1
Total	9	14	4	6

In our vocabulary, there is 3 words {good, poor, great} so |V|=3

Prior probabilities

$$P(\text{pos}) = \frac{2}{5} = 0.4$$
$$P(\text{neg}) = \frac{3}{5} = 0.6$$

Let's using the count table to calculate likelihoods of each word being a part of each class using add-1 smooth.
Multinomial naive Bayes:

$$P(\text{good}|\text{pos}) = \frac{\text{count}(\text{good}, \text{pos}) + 1}{(\sum_{w \in V} \text{count}(w, \text{pos})) + |V|} = \frac{3+1}{9+3} = \frac{1}{3} = 0.33$$

$$P(\text{good}|\text{neg}) = \frac{\text{count}(\text{good}, \text{neg}) + 1}{(\sum_{w \in V} \text{count}(w, \text{neg})) + |V|} = \frac{2+1}{14+3} = 0.176$$

P(poor|pos) = (1+1)/(9+3) = 0.167
P(poor|neg) = (10+1)/(14+3) = 0.647

P(great|pos) = (5+1)/(9+3) = 0.5
P(great|neg) = (2+1)/(14+3) = 0.176

Using Multinomial naive Bayes, let's calculate the most likely for the sentence D as follows.
`A good, good plot and great characters, but poor acting.`

We will ignore words which are not included in our vocabulary.

P(pos|D) ∝ P(pos) * P(good|pos) * P(good|pos) * P(great|pos) * P(poor|pos)
= 0.4 * 0.33 * 0.33 * 0.5 * 0.167 = 3.637 * 1e-3

P(neg|D) ∝ P(neg) * P(good|neg) * P(good|neg) * P(great|neg) * P(poor|neg)
= 0.6 * 0.176 * 0.176 * 0.176 * 0.647 = 2.11 * 1e-3

P(pos|D) > P(neg|D), so the most likely class the D using multinomial naïve Bayes is **positive**!

Binarized naive Bayes

$$P(\text{good}|\text{pos}) = \frac{\text{count}(\text{good}, \text{pos}) + 1}{(\sum_{w \in V} \text{count}(w, \text{pos})) + |V|} = \frac{1+1}{4+3} = 0.286$$

$$P(\text{good}|\text{neg}) = \frac{\text{count}(\text{good}, \text{neg}) + 1}{(\sum_{w \in V} \text{count}(w, \text{neg})) + |V|} = \frac{2+1}{6+3} = 0.33$$

P(poor|pos) = (1+1)/(4+3) = 0.286
P(poor|neg) = (3+1)/(6+3) = 0.44

P(great|pos) = (2+1)/(4+3) = 0.428
P(great|neg) = (1+1)/(6+3) = 0.22

Using Binarized naive Bayes, let's calculate the most likely for the sentence D as follows.

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

Note that with binarized NB, we keep a single instance for each word in D.

$P(pos|D) \propto P(pos) * P(good|pos) * P(great|pos) * P(poor|pos)$
$= 0.4 * 0.286 * 0.428 * 0.286 = 0.014$

$P(neg|D) \propto P(neg) * P(good|neg) * P(great|neg) * P(poor|neg)$
$= 0.6 * 0.33 * 0.22 * 0.44 = 0.019$

We have $P(neg|D) > P(pos|D)$, so the most likelily class of D is neg!

So, the two models disagree!

Note: You will get credit for the exercise if you just answer without any explanation. The explanation could be manual calculation or using program shown in Google Colab.