CS 542 (Fall 2023) Written Assignment 1
Bayes' Theorem and Naïve Bayes Classification

Eric Burton Martin

September 7, 2023

1 Bayes' Theorem

You get an email. You know that 90% of your email is legitimate (L) while 10% is spam (S).

a. Assume the following probabilities:

- The probability that an email contains the word "Bitcoin" (B) if it is spam is 96%.
- The probability that an email contains the word "Bitcoin" if it is legitimate is 5%.

What is the probability that your new email is spam given that it contains the word "Bitcoin"? **Show your work!**

b. Assume the following probabilities:

- The probability that an email contains the word "Covid" (C) if it is spam is 50%.
- The probability that an email contains the word "Covid" if it is legitimate is 12%.

What is the probability that your new email is legitimate given that it contains the word "Covid"? **Show your work!**

1.1 Answer to part a

a. We can use Baye's theorem as follows

$$P(\mathbf{A}|\mathbf{B}) = \frac{P(\mathbf{B}|\mathbf{A})P(\mathbf{A})}{P(\mathbf{B})} \tag{1}$$

We are given the following:

$$P(\textbf{Legit}|\textbf{Email}) = 0.90 \tag{2}$$

$$P(\textbf{Spam}|\textbf{Email}) = 0.10 \tag{3}$$

$$P(\textbf{Bitcoin}|\textbf{Spam}) = 0.96 \tag{4}$$

$$P(\textbf{Bitcoin}|\textbf{Legit}) = 0.05 \tag{5}$$

We want to solve the following equation:

$$P(\textbf{S}|\textbf{B}) = \frac{P(\textbf{B}|\textbf{S})P(\textbf{S})}{P(\textbf{B})} \tag{6}$$

We need to solve for P(B) though since that is not directly given. Since an email is either spam or legit, the probability that an email contains the word Bitcoin is therefore:

$$\begin{aligned} P(\textbf{B}) = \\ P(\textbf{B} \wedge \textbf{S}) + P(\textbf{B} \wedge \textbf{L}) = \\ (P(\textbf{B}|\textbf{S}) * P(\textbf{S}) + (P(\textbf{B}|\textbf{L}) * P(\textbf{L})) = \\ (0.96 * 0.10) + (0.05 * 0.90) = \\ \textbf{0.141} \end{aligned} \tag{7}$$

We can now solve for P(S given B) which gives:

$$P(\textbf{S}|\textbf{B}) = \frac{(0.96 * 0.10)}{(0.141)} = \textbf{0.681} \tag{8}$$

1.2 Answer to part b

b. We know the following:

$$P(\textbf{Legit}|\textbf{Email}) = 0.90 \tag{9}$$

$$P(\mathbf{Spam}|\mathbf{Email}) = 0.10 \tag{10}$$

$$P(\mathbf{Covid}|\mathbf{Spam}) = 0.50 \tag{11}$$

$$P(\mathbf{Covid}|\mathbf{Legit}) = 0.12 \tag{12}$$

We want to find the probability that a new email is legitimate given that it contains the word "Covid" P(L given C) which we can represent as:

$$P(\mathbf{L}|\mathbf{C}) = \frac{P(\mathbf{C}|\mathbf{L})P(\mathbf{L})}{P(\mathbf{C})} \tag{13}$$

We must first solve for P(C) by doing the following:

$$\begin{aligned} P(\mathbf{C}) = \\ P(\mathbf{C} \wedge \mathbf{S}) + P(\mathbf{C} \wedge \mathbf{L}) = \\ (P(\mathbf{C}|\mathbf{S}) * P(\mathbf{S}) + (P(\mathbf{C}|\mathbf{L}) * P(\mathbf{L})) = \\ (0.50 * 0.10) + (0.12 * 0.90) = \\ \mathbf{0.158} \end{aligned} \tag{14}$$

Now we can complete the Baye's formula and solve for P(L given C):

$$P(\mathbf{L}|\mathbf{C}) = \frac{(0.12) * (0.90)}{(0.158)} = \mathbf{0.684} \tag{15}$$

2 Naïve Bayes

The following problem is from the Jurafsky and Martin book, Exercise 4.2, reproduced below.

Given the following short movie reviews, each labeled with a genre, either comedy or action:

document	class
fly fast shoot love	action
fun couple love love	comedy
fast furious shoot	action
couple fly fast fun fun	comedy
furious shoot shoot fun	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.

Show your work! In particular, show all of the probability distributions involved in the model (namely, P(class) and P(feature|class)) and all of the steps used to calculate them. Create (conditional) probability tables such as those shown below.

class	P(class)
action	
comedy	

P(feature|class)		feature			
		fast	couple	shoot	fly
class	action				
	comedy				

Perform Laplace Smoothing to account for words that do not appear in one class.

2.1 Answer

We will first find the class priors for Action and Comedy. We find the percentage of the documents in our training set are in each class. Let Nc be the number of documents in our training data with class c and Ndoc be the total number of documents. Then:

$$\hat{P}(c) = \frac{N_c}{N_{doc}} \tag{16}$$

With that equation, the prior of each class is calculated below:

$$P(\textbf{Action}) = \frac{3}{5} \tag{17}$$

$$P(\textbf{Comedy}) = \frac{2}{5} \tag{18}$$

Which can be represented in the table below:

class	P(class)
action	0.6
comedy	0.4

We now can compute the likelihoods of the words "fast", "couple", "shoot", and "fly" using the add-one (Laplace) smoothing equation from the Jurafsky and Martin book below:

$$\hat{P}(w_i|c) = \frac{count(w_i, c) + 1}{\sum_{w \in V}(count(w_i, c) + 1)} = \frac{count(w_i, c) + 1}{(\sum_{w \in V}(count(w_i, c)) + |V|} \tag{19}$$

Where the total word count in each category is:

$$count_{(words,action)} = 11 \tag{20}$$

$$count_{(words,comedy)} = 9 \tag{21}$$

And vocabulary V consists of the unique words in all classes.

$$|V| = 7 \tag{22}$$

So, now we can compute the individual probabilities for "fast", "couple", "shoot", and "fly" using Eq. 19.

$$P(\text{“}fast\text{”}|Action) = \frac{2+1}{11+7} = 0.1667 \tag{23}$$

$$P(\text{“}fast\text{”}|Comedy) = \frac{1+1}{9+7} = 0.1250 \tag{24}$$

$$P(\text{“}couple\text{”}|Action) = \frac{0+1}{11+7} = 0.0556 \tag{25}$$

$$P(\text{“}couple\text{”}|Comedy) = \frac{2+1}{9+7} = 0.1875 \tag{26}$$

$$P(\text{“}shoot\text{”}|Action) = \frac{4+1}{11+7} = 0.2778 \tag{27}$$

$$P(\text{“}shoot\text{”}|Comedy) = \frac{0+1}{9+7} = 0.0625 \tag{28}$$

$$P(\text{“}fly\text{”}|Action) = \frac{1+1}{11+7} = 0.1111 \tag{29}$$

$$P(\text{“}fly\text{”}|Comedy) = \frac{1+1}{9+7} = 0.1250 \tag{30}$$

Which can be represented in the table below:

P(feature|class)		feature			
		fast	couple	shoot	fly
class	action	0.1667	0.0556	0.2778	0.1111
	comedy	0.1250	0.1875	0.0625	0.1250

To apply the naive Bayes classifier we can use equation 4.9 from the Jurafsky and Martin book to get the following:

$$P(action)P(sentence|Action) = \frac{3}{5} \cdot \frac{3 \cdot 1 \cdot 5 \cdot 2}{18^4} = 2.86e-4 \tag{31}$$

$$P(comdedy)P(sentence|comedy) = \frac{2}{5} \cdot \frac{2 \cdot 3 \cdot 1 \cdot 2}{16^4} = 1.83e-4 \tag{32}$$

So we can see that the classification of the features “fast”, “couple”, “shoot”, and “fly” using Laplace Smoothing will be **Action**.

Submission Instructions

Please submit your solutions (in PDF format - printed and scanned images are OK) to the drop box on Canvas.