CS 480 – Introduction to Artificial Intelligence

Topic: Naïve Bayes

Mustafa Bilgic

http://www.cs.iit.edu/~mbilgic

https://twitter.com/bilgicm

Task

- Classify emails as spam (s) / not-spam (~s) based on the words they contain
- You look at 100 random emails; 40 of them are spam, 60 of them are not-spam
- What is P(s) for a new email?

Features

- Assume you'll look into the emails' contents; you've decided that the word Nigeria[1] seems to correlate well with spam. You group the 100 emails as follows

Nigeria	Spam	Count
t	s	30
f	s	10
t	~s	10
f	~s	50

If the word Nigeria appears in the new email, then what is P(s | Nigeria=t)?

1. Why "Nigeria?" https://www.google.com/search?q=nigeria+scam+emails

NIGERIA=T

7 N B

Nigeria	Spam	Count
t	s	30
f	s	10
t	~s	10
f	~s	50

If the word Nigeria appears in the new email, then what is P(s | Nigeria=t)?

$$P(s \mid N=t) = \frac{P(s, N=t)}{P(N=t)} = \frac{30/100}{(30+10)/100} = \frac{30}{40}$$

7 IV B

Add Admission into your vocabulary

Nigeria	Adm.	Spam	Count
t	t	s	10
t	f	s	20
f	t	s	3
f	f	s	7
t	t	~s	8
t	f	~s	2
f	t	~s	40
f	f	~s	10

What is P(s | N=t, A=f)? What about P(s | N=t, A=t)?

Add Admission into your vocabulary

Nigeria	Adm.	Spam	Count
t	t	s	10
t	f	s	20
f	t	s	3
f	f	s	7
t	t	~s	8
t	f	~s	2
f	t	~s	40
f	f	~s	10

What is P(s | N=t, A=f)? What about P(s | N=t, A=t)?

$$P(s \mid N=t, A=f) = \frac{P(s, N=t, A=f)}{P(N=t, A=f)} = \frac{20/100}{(20+2)/100} = \frac{20}{22}$$

P(s | N=t) was 0.75. P(s | N=t, A=f) is 0.91

Add Admission into your vocabulary

Nigeria	Adm.	Spam	Count
t	t	s	10
t	f	s	20
f	t	s	3
f	f	s	7
t	t	~s	8
t	f	~s	2
f	t	~s	40
f	f	~s	10

What is P(s | N=t, A=f)? What about P(s | N=t, A=t)?

$$P(s \mid N=t, A=t) = \frac{P(s, N=t, A=f)}{P(N=t, A=f)} = \frac{10/100}{(10+8)/100} = \frac{10}{18}$$

P(s | N=t) was 0.75. P(s | N=t, A=f) is 0.91. P(s | N=t, A=t) = 0.56.

Now assume we add 998 more words

W_1	W_2	…	W_{1000}	Spam	Count
t	t	…	t	s	
t	t	…	f	s	
…	…	…	…	…	
f	f	…	f	~s	

Q: How many entries are there in this table?

A: $2^{1001} \approx 2 \times 10^{301}$

We have 100 emails. If all emails are distinct, 100 entries will be 1; The rest will be 0.

Q: What is $P(s \mid W_1=t, W_2=f, \ldots, W_{1000}=t)$?

A: Either 1 or 0 if it is in D, otherwise, it is NaN

Q: How big of a training data do we need?

NB

Naïve Bayes

- Given X_1, X_2, …, X_n, and class Y
- Assume $X_i \perp X_j \mid Y$

$$P(Y|X_1, X_2, \ldots, X_n) = \frac{P(X_1, X_2, \ldots, X_n|Y)P(Y)}{P(X_1, X_2, \ldots, X_n)} = \frac{P(Y)\prod_{i=1}^{n} P(X_i|Y)}{P(X_1, X_2, \ldots, X_n)}$$

We need to estimate $P(Y)$ and $P(X_i \mid Y)$

What is the Bayesian network representation of Naïve Bayes?

Naïve Bayes

Nigeria	Adm.	Spam	Count
t	t	s	10
t	f	s	20
f	t	s	3
f	f	s	7
t	t	~s	8
t	f	~s	2
f	t	~s	40
f	f	~s	10

What is P(S)?

What is P(N|S)?

What is P(A|S)?

P(Spam)
P(N|S)
P(A|S)

MLE
or
LS (+1)

MLE

Naïve Bayes

Nigeria	Adm.	Spam	Count
t	t	s	10
t	f	s	20
f	t	s	3
f	f	s	7
t	t	~s	8
t	f	~s	2
f	t	~s	40
f	f	~s	10

What is P(S)?

Spam	P(S)
s	40/100
~s	60/100

What is P(N|S)?

Nigeria	Spam	P(N,S)	P(N|S)
t	s	30/100	30/40
f	s	10/100	10/40
t	~s	10/100	10/60
f	~s	50/100	50/60

What is P(A|S)?

Adm.	Spam	P(A,S)	P(A|S)
t	s	13/100	13/40
f	s	27/100	27/40
t	~s	48/100	48/60
f	~s	12/100	12/60

Inference in naïve bayes

- What is P(spam | N=t, A=f)?

Zero probabilities

- We have n features, X_1 through X_n
- If $P(X_i|C)$ is zero for any feature and class combination, we would be in trouble
- Example
  - Assume that X_{592} is a weird feature that is rarely *true* in the world. Assume that X_{592} is always *false* in our training data, no matter what the class is
    - $P(X_{592} = f \mid C = t) = 1$; $P(X_{592} = t \mid C = t) = 0$
    - $P(X_{592} = f \mid C = f) = 1$; $P(X_{592} = t \mid C = f) = 0$
  - In one of the objects in our test data, X_{592} is *true*.
    - What is $P(C \mid X_1, X_2, \ldots, X_{592} = t, \ldots X_n)$?
- One solution: Laplace smoothing

Multiplying Several Probability Numbers

- Assume we have 10,000 features
- What is $0.9^{10,000}$ using a computer?
- Try math.pow(0.9, 10000) in Python
- In Naïve Bayes,
  - $a = P(Y = T) \prod P(X_i | Y = T)$
  - $b = P(Y = F) \prod P(X_i | Y = F)$
  - $P(Y = T | X) = \frac{a}{a+b}$
  - If $a = b = 0$, then what?