University of Colorado **Boulder**

Department of Computer Science
CSCI 2824: Discrete Structures
Chris Ketelsen

Discrete Probability
Bayes' Theorem and the Law of Total Probability

Discrete Probability Theory

Conditional Probability: The probability that E occurs given that F occurred:

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}$$

Product Rule: $p(E \cap F) = p(E \mid F)\, p(F)$

Independence: Events A and B are independent if

- $p(A \mid B) = p(A)$
- $p(B \mid A) = p(B)$
- $p(A \cap B) = p(A)P(B)$

The Law of Total Probability

Example: Suppose I have two bags of marbles containing black and white marbles:

	White	Black
Bag 1	6	4
Bag 2	3	7

Now suppose I put the two bags in a box. If I close my eyes, grab a bag from the box, and then grab a marble from the bag, what is the probability that it is black?

The Law of Total Probability

Example: Suppose I have two bags of marbles containing black and white marbles:

	White	Black
Bag 1	6	4
Bag 2	3	7

Same scenario as before, but now suppose that the first bag is much larger than the second bag, so that when I reach into the box I'm twice as likely to grab the first bag as the second. What is the probability of grabbing a black marble?

The Law of Total Probability

The Law of Total Probability: Let A and B be events. Then

$$p(A) = p(A \mid B)p(B) + p(A \mid \overline{B})p(\overline{B})$$

Proof:

Remark: *Incredibly* useful when you don't know $p(A)$ but you do know $p(A \mid B)$ (e.g. the prob of drawing a white marble given it's Bag 1)

The Law of Total Probability

The Generalized LTP: Let $B_1, B_2, \ldots B_n$ be disjoint events such that $B_1 \cup B_2 \cup \cdots \cup B_n = S$. Then

$$p(A) = p(A \mid B_2)p(B_2) + p(A \mid B_2)p(B_2) + \cdots + p(A \mid B_n)p(B_n)$$

Bayes' Theorem

Bayes' Thm is one of the most important theorems in probability

Allows us to do interesting thing with conditional probabilities

- Predict probability of disease based on test results
- Do better at old game shows
- Determine whether email is SPAM or not

The Law of Total Probability

Example: Suppose I have two bags of marbles containing black and white marbles:

	White	Black
Bag 1	6	4
Bag 2	3	7

Same scenario as before. If I close my eyes, grab a bag from the box, reach into the bag and pull out a white marble. What is the probability that I picked Bag 1?

The Law of Total Probability

Example: Suppose I have two bags of marbles containing black and white marbles:

	White	Black
Bag 1	6	4
Bag 2	3	7

Same scenario as before. If I close my eyes, grab a bag from the box, reach into the bag and pull out a white marble. What is the probability that I picked Bag 1?

Bayes' Theorem

This notion of using information (the ball is red) to update our belief about an event (that we drew from Box 1) is the cornerstone of a statistical framework called **Bayesian Reasoning**

The formulas that we derived in the previous example are called Bayes' Theorem

$$p(F \mid E) = \frac{p(E \mid F) \cdot p(F)}{p(E)}$$

Or, using the definition of $p(E)$

$$p(F \mid E) = \frac{p(E \mid F) \cdot p(F)}{p(E \mid F)p(F) + p(E \mid \overline{F})p(\overline{F})}$$

Let's look at a classic example of Bayes' Theorem

Bayes' Theorem

Classic Cancer Test Example

Let's assume we know that 1% of people over the age of 40 have cancer

$$p(\mathrm{C}) = 0.01$$

Bayes' Theorem

Classic Cancer Test Example

Let's assume we know that 1% of people over the age of 40 have cancer

$$p(\mathrm{C}) = 0.01$$

Let's assume that 90% of people who **have cancer** will test positive for cancer in test

$$p(\mathrm{pos} \mid \mathrm{C}) = 0.90$$

Bayes' Theorem

Classic Cancer Test Example

Let's assume we know that 1% of people over the age of 40 have cancer

$$p(\mathrm{C}) = 0.01$$

Let's assume that 90% of people who **have cancer** will test positive for cancer in test

$$p(\mathrm{pos} \mid \mathrm{C}) = 0.90$$

Finally, assume that 8% of people who do **not** have cancer will also test positive

$$p(\mathrm{pos} \mid \overline{C}) = 0.08$$

Bayes' Theorem

What is the probability that a person who tests positive for cancer **actually has cancer**? In other words, what is

$$p(\mathrm{C} \mid \mathrm{pos})$$

Bayes' Theorem

What is the probability that a person who tests positive for cancer **actually has cancer**? In other words, what is

$$p(\mathrm{C} \mid \mathrm{pos})$$

Most people will assume that if they get a positive test, then there is a 90% chance that they actually have cancer.

Bayes' Theorem

What is the probability that a person who tests positive for cancer **actually has cancer**? In other words, what is

$$p(\mathrm{C} \mid \mathrm{pos})$$

Most people will assume that if they get a positive test, then there is a 90% chance that they actually have cancer.

But this ignores the incredibly important fact that **not many people have cancer**! Remember:

$$p(\mathrm{C}) = 0.01$$

Bayes' Theorem

Let's do the actual calculation. From Bayes Law, we have

$$p(\mathrm{C} \mid \mathrm{pos}) = \frac{p(\mathrm{pos} \mid \mathrm{C})\, p(\mathrm{C})}{p(\mathrm{pos} \mid \mathrm{C})\, p(\mathrm{C}) + p(\mathrm{pos} \mid \overline{C})\, p(\overline{C})}$$

The only quantity we haven't specified is the probability of not having cancer

$$p(\overline{C}) = 1 - p(\mathrm{C}) = 1 - 0.01 = 0.99$$

Plugging everything into Bayes Law gives ...

Bayes' Theorem

Let's do the actual calculation. From Bayes Law, we have

$$p(\mathrm{C} \mid \mathrm{pos}) = \frac{0.90 \cdot 0.01}{0.90 \cdot 0.01 + 0.08 \cdot 0.99} = 0.10$$

So even if you test positive for cancer, there is only about a 10% chance that you actually **have** cancer.

Bayes' Theorem

What is the probability that a person who tests negative for cancer is **actually cancer free**? In other words, what is

$$p(\overline{C} \mid \text{neg})$$

From Bayes Law, we have

$$p(\overline{C} \mid \text{neg}) = \frac{p(\text{neg} \mid \overline{C})\, p(\overline{C})}{p(\text{neg} \mid \overline{C})\, p(\overline{C}) + p(\text{neg} \mid C)\, p(C)}$$

We need to compute new quantities $p(\text{neg} \mid \overline{C})$ and $p(\text{neg} \mid C)$

$$p(\text{neg} \mid \overline{C}) = 1 - p(\text{pos} \mid \overline{C}) = 1 - 0.08 = 0.98$$

$$p(\text{neg} \mid C) = 1 - p(\text{pos} \mid C) = 1 - 0.90 = 0.10$$

Bayes' Theorem

What is the probability that a person who tests negative for cancer is **actually cancer free**? In other words, what is

$$p(\overline{C} \mid \text{neg})$$

From Bayes Law, we have

$$\begin{aligned} p(\overline{C} \mid \text{neg}) &= \frac{p(\text{neg} \mid \overline{C})\, p(\overline{C})}{p(\text{neg} \mid \overline{C})\, p(\overline{C}) + p(\text{neg} \mid C)\, p(C)} \\ &= \frac{0.98 \cdot 0.99}{0.98 \cdot 0.99 + 0.08 \cdot 0.01} \\ &= 0.99917 \end{aligned}$$

So if you receive a negative test there is a 99.9% chance that you in fact do not have cancer

Bayes' Theorem

Note, we can use Bayes' Theorem in more general cases, for example in the ball-drawing problem when there are more than two boxes. The simpler version of the theorem still holds

$$p(F \mid E) = \frac{p(E \mid F) \cdot p(F)}{p(E)}$$

We just need to compute an alternate definition of $p(E)$

Generalized Bayes' Thm: Suppose that E is an event from sample space S and $F_1, F_2, \ldots, F_n$ are mutually disjoint events such that $\cup_{k=1}^{n} F_k = S$. Then

$$p(F_i \mid E) = \frac{p(E \mid F_i) \cdot p(F_i)}{\sum_{k=1}^{n} p(E \mid F_k) \cdot p(F_k)}$$

Bayesian SPAM Filters

Bayesian SPAM filters use information about previously seen emails to guess whether an incoming email is SPAM or not

Simple versions of a BSF look for occurrences of particular words in the message that might indicate that the message is SPAM, e.g.

$$buy, \quad viagra, \quad money, \quad nigeria$$

Suppose you have catalogued a bunch of emails, some of them SPAM and some of them not

Could you estimate the probability that a word w appears in an email given that you know the email is SPAM?

In other words, could you estimate $p(w \mid SPAM)$?

Bayesian SPAM Filters

Estimating $p(w \mid SPAM)$:

Simply count the number of SPAM messages that contain word w, then divide by the total number of SPAM messages

$$p(w \mid SPAM) = \frac{\text{\# SPAM messages containing } w}{\text{Total \# of SPAM messages}}$$

Similarly, we could estimate the probability that the word w appears in a not SPAM message (often called a HAM message)

$$p(w \mid HAM) = \frac{\text{\# HAM messages containing } w}{\text{Total \# of HAM messages}}$$

Bayesian SPAM Filters

Example: Given the following set of messages, estimate $p(\text{money} \mid SPAM)$ and $p(\text{money} \mid HAM)$

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*money*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

Bayesian SPAM Filters

Example: Given the following set of messages, estimate $p(\text{money} \mid SPAM)$ and $p(\text{money} \mid HAM)$

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	**money**	*fly*
buy	**money**	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

$$p(\text{money} \mid SPAM) = \frac{2}{3}, \quad p(\text{money} \mid HAM) = \frac{1}{2}$$

Bayesian SPAM Filters

OK, but how does this help us decide if an email is SPAM or HAM?

What quantity do we really want to compute?

Bayesian SPAM Filters

OK, but how does this help us decide if an email is SPAM or HAM?

What quantity do we really want to compute?

Really, we want $p(SPAM \mid \text{money})$ and $p(HAM \mid \text{money})$

This asks the question: What is the probability that this email is SPAM given that I've observed the word money in it?

Fortunately, we can get at these quantities using Bayes' Thm

Bayesian SPAM Filters

For a general word w we have

$$p(SPAM \mid w) = \frac{p(w \mid SPAM) \cdot p(SPAM)}{p(w)}$$

or better yet $p(SPAM \mid w) =$

$$\frac{p(w \mid SPAM) \cdot p(SPAM)}{p(w \mid SPAM)p(SPAM) + p(w \mid HAM)p(HAM)}$$

What additional terms in this formula do we need to estimate?

Bayesian SPAM Filters

For a general word w we have

$$p(SPAM \mid w) = \frac{p(w \mid SPAM) \cdot p(SPAM)}{p(w)}$$

or better yet $p(SPAM \mid w) =$

$$\frac{p(w \mid SPAM) \cdot p(SPAM)}{p(w \mid SPAM)p(SPAM) + p(w \mid HAM)p(HAM)}$$

What additional terms in this formula do we need to estimate?

We're missing $p(SPAM)$ and $p(HAM)$

Bayesian SPAM Filters

How could we estimate $p(SPAM)$ and $p(HAM)$?

Bayesian SPAM Filters

How could we estimate $p(SPAM)$ and $p(HAM)$?

We just divide the number of each type of message in our data set by the total number of messages

$$p(SPAM) = \frac{\text{\# SPAM messages}}{\text{Total \# messages}}$$

$$p(HAM) = \frac{\text{\# HAM messages}}{\text{Total \# messages}}$$

Bayesian SPAM Filters

Example: Given the same set of messages, estimate $p(SPAM)$ and $p(HAM)$

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*money*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

Bayesian SPAM Filters

Example: Given the same set of messages, estimate $p(SPAM)$ and $p(HAM)$

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*money*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

$$p(SPAM) = \frac{3}{5}, \quad p(HAM) = \frac{2}{5}$$

Bayesian SPAM Filters

Example: Now put it all together. Given the following set of messages, estimate $p(SPAM \mid \text{money})$ and $p(HAM \mid \text{money})$

ham	$spam$	$spam$	$spam$	ham
$work$	$nigeria$	fly	$money$	fly
buy	$opportunity$	buy	buy	$home$
$money$	$viagra$	$nigeria$	fly	$nigeria$

Bayesian SPAM Filters

Example: Now put it all together. Given the following set of messages, estimate $p(SPAM \mid \text{money})$ and $p(HAM \mid \text{money})$

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*opportunity*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

$$
\begin{aligned}
p(SPAM \mid \text{money}) &= \frac{p(\text{money} \mid SPAM) \cdot p(SPAM)}{p(\text{money} \mid SPAM)p(SPAM) + p(\text{money} \mid HAM)p(HAM)} \\
&= \frac{2/3 \cdot 3/5}{2/3 \cdot 3/5 + 1/2 \cdot 2/5} = \frac{2}{3}
\end{aligned}
$$

Bayesian SPAM Filters

Example: Now put it all together. Given the following set of messages, estimate $p(SPAM \mid \text{money})$ and $p(HAM \mid \text{money})$

ham	$spam$	$spam$	$spam$	ham
$work$	$nigeria$	fly	$money$	fly
buy	$opportunity$	buy	buy	$home$
$money$	$viagra$	$nigeria$	fly	$nigeria$

$$\begin{aligned} p(SPAM \mid \text{money}) &= \frac{p(\text{money} \mid SPAM) \cdot p(SPAM)}{p(\text{money} \mid SPAM)p(SPAM) + p(\text{money} \mid HAM)p(HAM)} \\ &= \frac{2/3 \cdot 3/5}{2/3 \cdot 3/5 + 1/2 \cdot 2/5} = \frac{2}{3} \end{aligned}$$

$$\begin{aligned} p(HAM \mid \text{money}) &= \frac{p(\text{money} \mid HAM) \cdot p(HAM)}{p(\text{money} \mid HAM)p(HAM) + p(\text{money} \mid SPAM)p(SPAM)} \\ &= \frac{1/2 \cdot 2/5}{1/2 \cdot 2/5 + 2/3 \cdot 3/5} = \frac{1}{3} \end{aligned}$$

Bayesian SPAM Filters

OK, we found that

$$p(SPAM \mid \text{money}) = \frac{2}{3}, \quad p(HAM \mid \text{money}) = \frac{1}{3}$$

How should we classify this email? As SPAM or as HAM?

Bayesian SPAM Filters

OK, we found that

$$p(SPAM \mid \text{money}) = \frac{2}{3}, \quad p(HAM \mid \text{money}) = \frac{1}{3}$$

How should we classify this email? As SPAM or as HAM?

Since $p(SPAM \mid \text{money}) > 0.5$ it's probably reasonable to classify it as SPAM.

But you should worry about the implications of making a mistake

Marking a SPAM message as HAM is annoying, but not a big deal

Classifying a truly HAM message as SPAM could be disastrous

Might be worth it to set the threshold higher than 0.5

Bayesian SPAM Filters

OK, so we've come up with a Bayesian Spam Filter that can filter messages based on the appearance of certain words

Buuuut, it seems like we might want to do this with all of the words in the email simultaneously

It turns out that we can do this if we're willing to make one fairly strong assumption about the relationship of words in SPAM emails

Bayesian SPAM Filters

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*opportunity*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

Suppose we receive an email with just the words $\{\text{buy}, \text{nigeria}\}$

We would like to compute the probabilities

$$p(SPAM \mid \{\text{buy}, \text{nigeria}\}) \text{ and } p(HAM \mid \{\text{buy}, \text{nigeria}\})$$

which requires knowing the flipped probabilities

$$p(\{\text{buy}, \text{nigeria}\} \mid SPAM) \text{ and } p(\{\text{buy}, \text{nigeria}\} \mid HAM)$$

Bayesian SPAM Filters

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*opportunity*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

$$p(\{\text{buy}, \text{nigeria}\} \mid SPAM) \text{ and } p(\{\text{buy}, \text{nigeria}\} \mid HAM)$$

Technically, we could estimate these probabilities by counting the number of emails that contain **both** buy and nigeria

But this would be very hard if had to re-estimate these probabilities every time we receive a new email

Assumption: Words are **independent** given their category

Bayesian SPAM Filters

$$p(\{\text{buy}, \text{nigeria}\} \mid SPAM) \;\text{ and }\; p(\{\text{buy}, \text{nigeria}\} \mid HAM)$$

Assumption: Words are **independent** given their category

Example:

$$p(\{\text{buy}, \text{nigeria}\} \mid SPAM) = p(\text{buy} \mid SPAM)p(\text{nigeria} \mid SPAM)$$

Under this assumption, we can compute probabilities for individual words exactly one time.

Then we can multiply them when we receive specific emails

Final Trick: Note that denominators are the same

$$p(SPAM \mid \text{email}) = \frac{p(\text{email} \mid SPAM)p(SPAM)}{p(\text{email})}$$

Bayesian SPAM Filters

$$p(\{\text{buy}, \text{nigeria}\} \mid SPAM) \;\text{ and }\; p(\{\text{buy}, \text{nigeria}\} \mid HAM)$$

Assumption: Words are **independent** given their category

Example:

$$p(\{\text{buy}, \text{nigeria}\} \mid SPAM) = p(\text{buy} \mid SPAM)p(\text{nigeria} \mid SPAM)$$

Under this assumption, we can compute probabilities for individual words exactly one time.

Then we can multiply them when we receive specific emails

Final Trick: Note that denominators are the same

$$p(HAM \mid \text{email}) = \frac{p(\text{email} \mid HAM)p(HAM)}{p(\text{email})}$$

Bayesian SPAM Filters

$$p(\{\text{buy}, \text{nigeria}\} \mid SPAM) \;\; \text{and} \;\; p(\{\text{buy}, \text{nigeria}\} \mid HAM)$$

Assumption: Words are **independent** given their category

Example:

$$p(\{\text{buy}, \text{nigeria}\} \mid SPAM) = p(\text{buy} \mid SPAM)p(\text{nigeria} \mid SPAM)$$

Under this assumption, we can compute probabilities for individual words exactly one time.

Then we can multiply them when we receive specific emails

Final Trick: Note that denominators are the same

Instead of comparing entire probability we can just compare numerators and make classification decision based on these

Bayesian SPAM Filters

Example: Given the following set of messages, predict whether the email $\{\text{buy, nigeria}\}$ is SPAM or HAM

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*opportunity*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

Need to compare

$$p(SPAM \mid \{\text{buy, nigeria}\})$$

$$p(HAM \mid \{\text{buy, nigeria}\})$$

Bayesian SPAM Filters

Example: Given the following set of messages, predict whether the email $\{\text{buy}, \text{nigeria}\}$ is SPAM or HAM

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*opportunity*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

Applying Bayes' and ignoring the denominatory, need to compare

$$p(\{\text{buy}, \text{nigeria} \mid SPAM)p(SPAM)$$

$$p(\{\text{buy}, \text{nigeria} \mid HAM)p(HAM)$$

Bayesian SPAM Filters

Example: Given the following set of messages, predict whether the email $\{\text{buy}, \text{nigeria}\}$ is SPAM or HAM

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*opportunity*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

Applying independence assumption, need to compare

$$p(\text{buy} \mid SPAM)p(\text{nigeria} \mid SPAM)p(SPAM)$$

$$p(\text{buy} \mid HAM)p(\text{nigeria} \mid HAM)p(HAM)$$

Bayesian SPAM Filters

Example: Given the following set of messages, predict whether the email $\{\text{buy}, \text{nigeria}\}$ is SPAM or HAM

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*opportunity*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

Plugging in numbers, we have

$$\text{SPAM:} \quad \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{3}{5} \approx 0.267$$

$$p(\text{buy} \mid HAM)p(\text{nigeria} \mid HAM)p(HAM)$$

Bayesian SPAM Filters

Example: Given the following set of messages, predict whether the email $\{\text{buy}, \text{nigeria}\}$ is SPAM or HAM

ham	*spam*	*spam*	*spam*	*ham*
work	*nigeria*	*fly*	*money*	*fly*
buy	*opportunity*	*buy*	*buy*	*home*
money	*viagra*	*nigeria*	*fly*	*nigeria*

Plugging in numbers, we have

$$\text{SPAM:} \quad \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{3}{5} \approx 0.267$$

$$\text{HAM:} \quad \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{5} = 0.1$$

Bayesian SPAM Filters

Since $0.267 > 0.1$ we predict that email $\{\text{buy}, \text{nigeria}\}$ is SPAM

Bayesian SPAM Filters