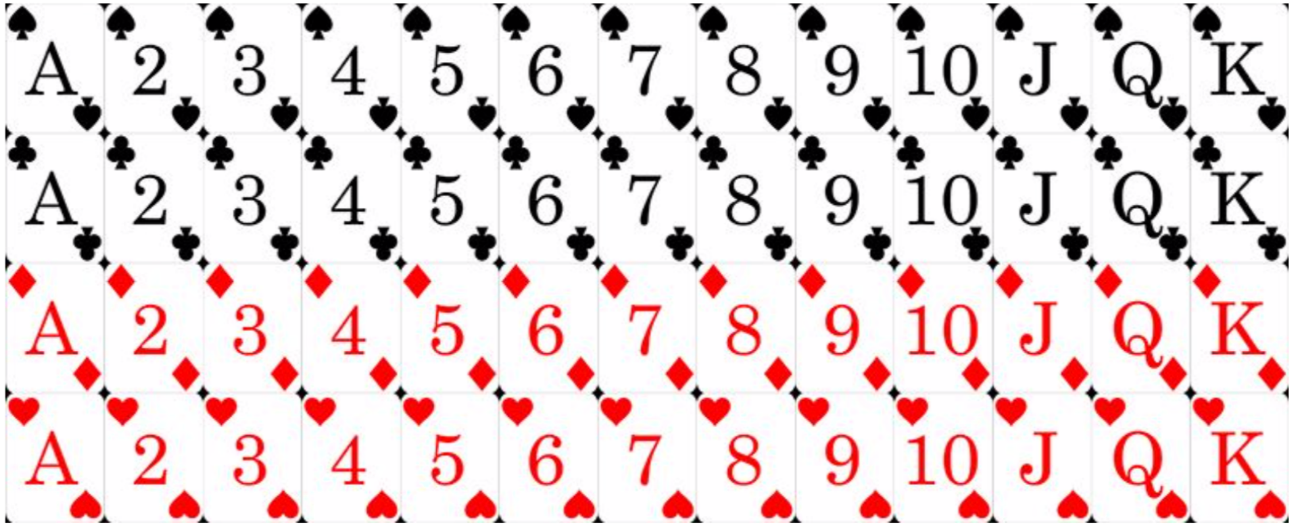
CSCI 3202: Intro to Artificial Intelligence

Lecture 14: Probability Review

Rachel Cox
Department of Computer Science

- Exam grades - posted on Gradescope
  Solutions - posted on Moodle

  Regrade requests - use the Regrade Request function in G.S.



- HW3 due Friday Mar 6 @ 11:59pm

HW3

#3c) update

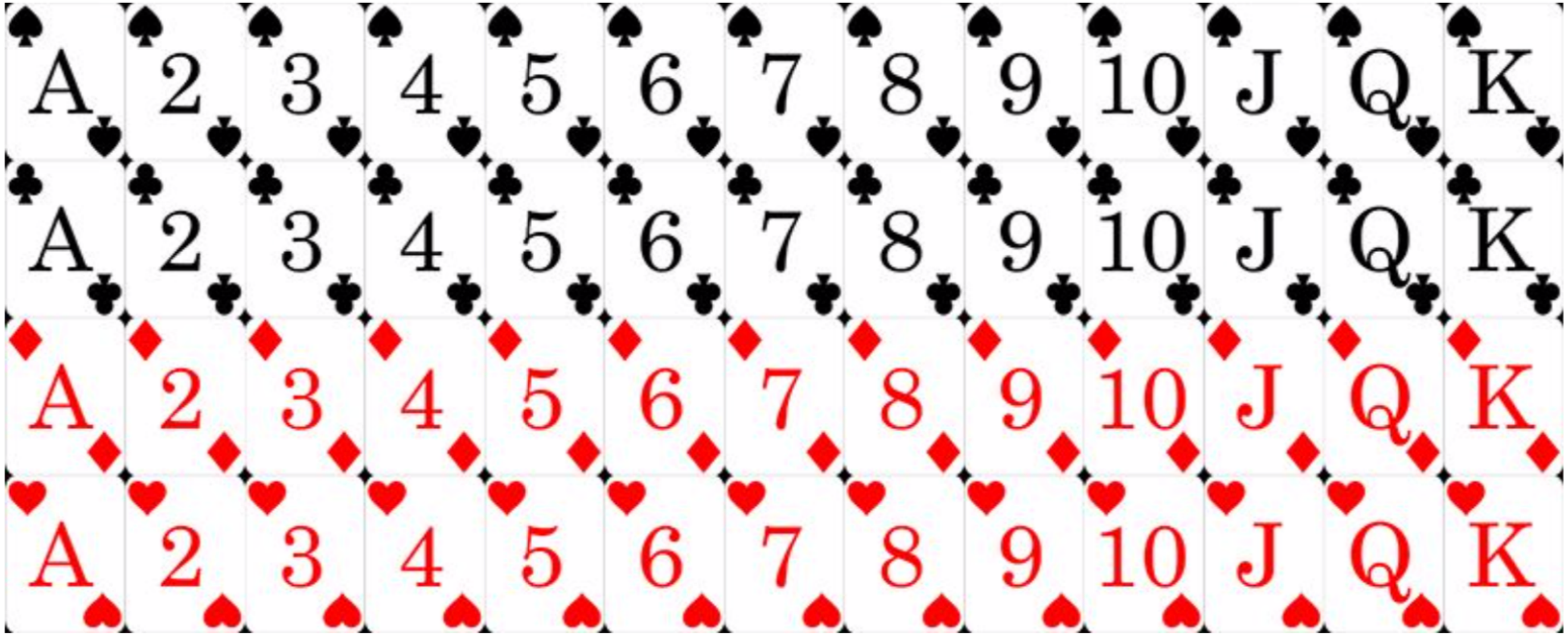
- Report whether you minimized or maximized.
- Report how many ingredients you did match
- Report how many iterations that took
- specifying the initial pop:

[36, 0, 0, 0, ... 0]

[0, 36, 0, 0, ... 0]

⋮

Probability Review



- 52 cards
- 2 colors
- 4 suits
- 13 possible values

Probability Review

Let V be a random variable representing card values.

Let C be a random variable representing card colors.

The **joint probability** of $V = v, C = c$, $\boxed{p(V = v, C = c)}$, is the probability that the card value is v and the card color is c simultaneously.

$p(6 \cap \text{Red})$

Example: $p(V = 6, C = red\ ) =$ 6♥ and 6♦

$$= \frac{1}{52} + \frac{1}{52}$$

$$= \frac{2}{52} \boxed{= 1/26}$$

Probability Review

The **conditional probability** of $V = v$ given $C = c$, $p(V|C)$, is the probability that the card value is v if you know that the card color is c. This is given by

$$\star \quad p(V|C) = \frac{p(V,C)}{p(C)}$$

$$p(V|C) = \frac{p(V \cap C)}{p(C)}$$

Example: $p(V = 6 \mid C = red) = \dfrac{p(6, \text{Red})}{p(\text{Red})} = \dfrac{1/26}{1/2} \boxed{= \dfrac{1}{13}}$

"given"

Example: $p(V = 6, C = red\,) = \dfrac{1}{13} \cdot \dfrac{1}{2} \boxed{= \dfrac{1}{26}}$

Note: $\star \; \boxed{p(V,C) = p(V|C)\,p(C)}$

Probability Review

$p(A, B) = p(B, A)$

$p(A \cap B) = p(B \cap A) = p(B|A)\,p(A)$

$p(A \cap B \cap C)$

Probability Chain Rule: We can use the product rule (twice) to show that for 3 random variables $A, B,$ and C:

$$p(A, B, C) = \boxed{p(C\,|A, B)}\, p(A, B) = \boxed{p(C\,|A, B)}\, \underbrace{p(B\,|A)\, p(A)}$$

$$p(A, B, C) = p(C, A, B) = p(C \mid A, B)\, p(A, B)$$

And if we have D random variables, $X_{1:D} = X_1, X_2, \dots, X_D$, then this becomes

$$p(X_1, X_2, \dots, X_D) = p(X_D \mid X_1, X_2, \dots, X_{D-1})\, p(X_{D-1} \mid X_1, X_2, \dots, X_{D-2}) \cdots p(X_2 \mid X_1)\, p(X_1)$$

In shorthand:

$$p(X_{1:D}) = p(X_D \mid X_{1:D-1})\, p(X_{D-1} \mid X_{1:D-2}) \cdots p(X_2 \mid X_1)\, p(X_1)$$

Further breakdown of previous slide

$$p(x_1, x_2, x_3, \ldots, x_{D-1}, x_D)$$

$$= p(x_D \mid x_1, x_2, x_3, \ldots, x_{D-1}) \cdot \underbrace{p(x_1, x_2, x_3, \ldots, x_{D-1})}$$

$$= p(x_D \mid x_1, \ldots, x_{D-1}) \cdot P(x_{D-1} \mid x_1, \ldots x_{D-2}) \cdot \underbrace{P(x_1, \ldots, x_{D-2})}$$

$$= p(x_D \mid x_1, \ldots x_{D-1}) \cdot P(x_{D-1} \mid x_1, \ldots x_{D-2})$$

$$\cdot P(x_{D-2} \mid x_1, \ldots, x_{D-3})$$

$$\cdot P(x_1, \ldots, x_{D-3})$$

$$\vdots$$

$$\cdot P(x_2 \mid x_1) P(x_1)$$

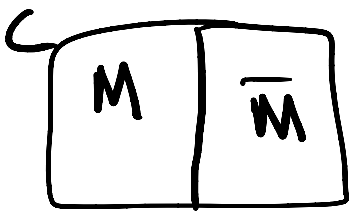
Probability Review

Given the joint density $p(V, C)$, the **marginal probability** of V is simply $p(V)$. (Similarly, the marginal probability of C is just $p(C)$).

We can calculate the marginal probability using our old friend:

Law of Total Probability: $p(V) = \sum_c p(V|C = c)p(C = c)$

C covers all possibilities.



Probability Review

Another old friend:

Thomas Bayes

Portrait purportedly of Bayes used in a 1936 book,[1] but it is doubtful whether the portrait is actually of him.[2] No earlier portrait or claimed portrait survives.

Bayes' Rule: $p(A \mid B) = \dfrac{p(B \mid A)\, p(A)}{p(B)}$

often use LTP

Probability Review

Example: Suppose we know that a particular test for Mad Cow Disease (MCD) will give a positive result for a cow known to be infected with MCD with a 70% reliability rate, and if the cow is healthy the test will return a false positive 10% of the time. Suppose that among the general cow population, 2% of cows are infected with MCD.

Two questions:

1. What is the probability that my cow will test positive?
2. What is the probability that a cow actually has MCD, given that they test positive?

T: let this be the event that the cow tests positive.

M: cow actually has MCD.

$p(T \mid M) = 0.7$, $p(T \mid \overline{M}) = 0.1$, $p(M) = 0.02$

Probability Review

Example: (continued)

1. What is the probability that my cow will test positive?

Law of Total Probability

$$P(T) = P(T|M)P(M) + P(T|\bar{M})P(\bar{M})$$

$$= (0.7)(0.02) + \underbrace{(0.1)\cdot(1-.02)}$$

$$\boxed{= 0.112}$$

Probability Review

Example: (continued)

2. What is the probability that a cow $\underbrace{\text{actually has MCD}}_{M}$, given that they test $\overset{T}{\text{positive}}$?

conditional prob

$$p(M \mid T) = \frac{p(T \mid M)\,p(M)}{p(T)}$$

$$= \frac{(0.7)(.02)}{0.112}$$

$$= 0.125$$

$\boxed{\text{Bayes' Rule}}$

Probability Review

Example: Suppose you have three coins in your pocket. Two of them are fair coins and one of them has tails on both sides. You pull a coin out of your pocket and flip it twice. It comes up tails both times. What is the probability that you are flipping the two-tails coin?

$\vec{x}$ - T_1 and T_2 - the events that we observed

This is our data.

y - let this represent the class of the coin.

$y = u$ for unfair coin

$y = F$ for fair coin

$$\underbrace{p(y \mid \vec{x})}_{} \overset{\text{in this problem}}{=} p(u \mid T_1, T_2) \overset{\text{Bayes}}{=} \frac{①\, p(T_1, T_2 \mid u)\, ②\, p(u)}{③\, p(T_1, T_2)}$$

$p(T_1, T_2 \mid u) = 1$

$p(u) = \frac{1}{3}$

$$p(T_1, T_2) = p(T_1, T_2 \mid F)\,P(F) + P(T_1, T_2 \mid u)\,p(u)$$
$$= \tfrac{1}{2} \cdot \tfrac{1}{2} \cdot \tfrac{2}{3} + 1 \cdot \tfrac{1}{3}$$

Probability Review

Example: Suppose you have three coins in your pocket. Two of them are fair coins and one of them has tails on both sides. You pull a coin out of your pocket and flip it twice. It comes up tails both times. What is the probability that you are flipping the two-tails coin?

Continued

$$p(u \mid T_1, T_2) = \frac{1 \cdot 1/3}{1 \cdot 1/3 + 1/2 \cdot 1/2 \cdot 2/3}$$

$$= \frac{2}{3}$$

Probability Review

In the coin-flipping example, we made. the assumption that

$$\underbrace{p(T_2, T_1 \mid U)} = \underbrace{p(T_2 \mid U)\, p(T_1 \mid U)}$$

- We assumed that T_2 and T_1 are conditionally independent, given the class of the coin, U.

For coin-flipping, this is actually valid. But we make this naïve assumption for other problems as well.

Class Conditional Independence: Assume that the features of **x** are conditionally independent, given the class **y**.

Probability Review

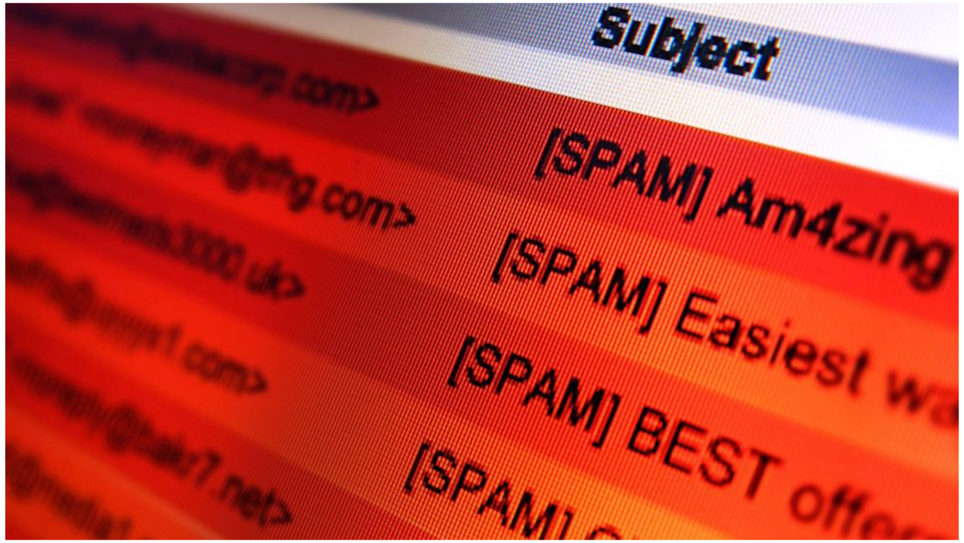
Example: Suppose we receive an email containing the words *buy*, *pills*, and *deal*. Should we classify the email as spam or ham?

Naïve assumption:

$$p(x = [buy, pills, deal] \mid y = \text{spam}) = p(buy \mid \text{spam})\, p(pills \mid \text{spam})\, p(deal \mid \text{spam})$$

Naïve Bayes Classifier:

$$p(y \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid y)\, p(y)}{p(\boldsymbol{x})}$$



In the context of spam filtering:

$$p(\text{spam} \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid \text{spam})\, p(\text{spam})}{p(\boldsymbol{x})}$$

Probability Review

Likelihood

$$\underbrace{p(y \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid y)\, p(y)}{p(\boldsymbol{x})}}_{\text{Bayes' Rule}}$$

posterior prob.

Likelihood, $p(\boldsymbol{x} \mid y)$:

- Given a class c, what is the probability that we would observe the features $\boldsymbol{x}$?
- Given that an email is spam (or ham), how likely is it that I would receive email x?
- This can be called the **class-conditional probability** for these classifier problems.

Posterior probability, $p(y \mid \boldsymbol{x})$:

- What is the probability that a new datum belongs to class $y = c$, given its observed features, **x**?
- What is the probability that an email is spam, given its content?

Probability Review

$p(y \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid y)\, p(y)}{p(\boldsymbol{x})}$

Prior probability, $p(y)$: posterior $p(y \mid \vec{x})$

- The general probability of encountering a particular class.
- How likely is it that any arbitrary incoming email is spam (or ham)?

➢ We can estimate $p(y)$ a few different ways:

- Ask an expert.
- Estimate from the data: $p(\text{ham}) = \frac{\text{\# ham emails in training data}}{\text{\# emails total in training data}}$

Probability Review

$p(y \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid y)\, p(y)}{p(\boldsymbol{x})}$

prior

posterior

likelihood

Evidence, $p(\boldsymbol{x})$:

- The probability of encountering the data x, independent of class labels.

- How likely is it that we would. receive an email containing the words x?

- We could calculate $p(x)$ using the Law of Total Probability:

$$p(x) = \sum_{c} p(x \mid y = c)\, p(y = c)$$

➢ But generally, we won't because it doesn't help us make decisions / often cancels out.

Probability Review

$p(y \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid y)\, p(y)}{p(\boldsymbol{x})}$

Law of Total Probability:

$$p(\boldsymbol{x}) = \sum_c p(\boldsymbol{x} \mid y = c)\, p(y = c)$$

➢ But generally, we won't because it doesn't help us make decisions / often cancels out…observe:

$$p(spam \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid spam)\, p(spam)}{p(\boldsymbol{x})} \qquad p(ham \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid ham)\, p(ham)}{p(\boldsymbol{x})}$$

Instead of computing $p(\vec{x})$

we can just compare the numerators

compare these.

$p(x \mid spam)\, p(spam)$ $\qquad$ $p(x \mid ham)\, p(ham)$

"spam score" on "ham score"

Probability Review

Example: Given these emails, calculate the ham score for the new email.

Email 1	Email 2	Email 3	Email 4	Email 5	New Email
ham	spam	spam	spam	ham	???
work	*nigeria*	*fly*	*money*	*fly*	*money*
buy	*money*	*buy*	*buy*	*home*	*nigeria*
money	*pills*	*nigeria*	*fly*	*nigeria*	

$p(\text{ham} \mid \text{money}, \text{nigeria})$ "=" $p(\text{money}, \text{nigeria} \mid \text{ham})\, p(\text{ham})$

not computing denominator

$= p(\text{money} \mid \text{ham}) \cdot p(\text{nigeria} \mid \text{ham})\, p(\text{ham})$

$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{10} = 0.1$

Probability Review

spam score > ham score.

Example: Given these emails, calculate the spam score for the new email.

Email 1	Email 2	Email 3	Email 4	Email 5	New Email
ham	spam	spam	spam	ham	???
work	*nigeria*	*fly*	*money*	*fly*	*money*
buy	*money*	*buy*	*buy*	*home*	*nigeria*
money	*pills*	*nigeria*	*fly*	*nigeria*	

$$p(\text{spam} \mid \text{money}, \text{nigeria}) \text{ "=" } p(\text{money}, \text{nigeria} \mid \text{spam})\, p(\text{spam})$$

$$= p(\text{money} \mid \text{spam}) \cdot p(\text{nigeria} \mid \text{spam})\, p(\text{spam})$$

$$= \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{3}{5} = \frac{4}{15} \approx 0.267$$

Next Time

- *Intro to Bayesian Reasoning*