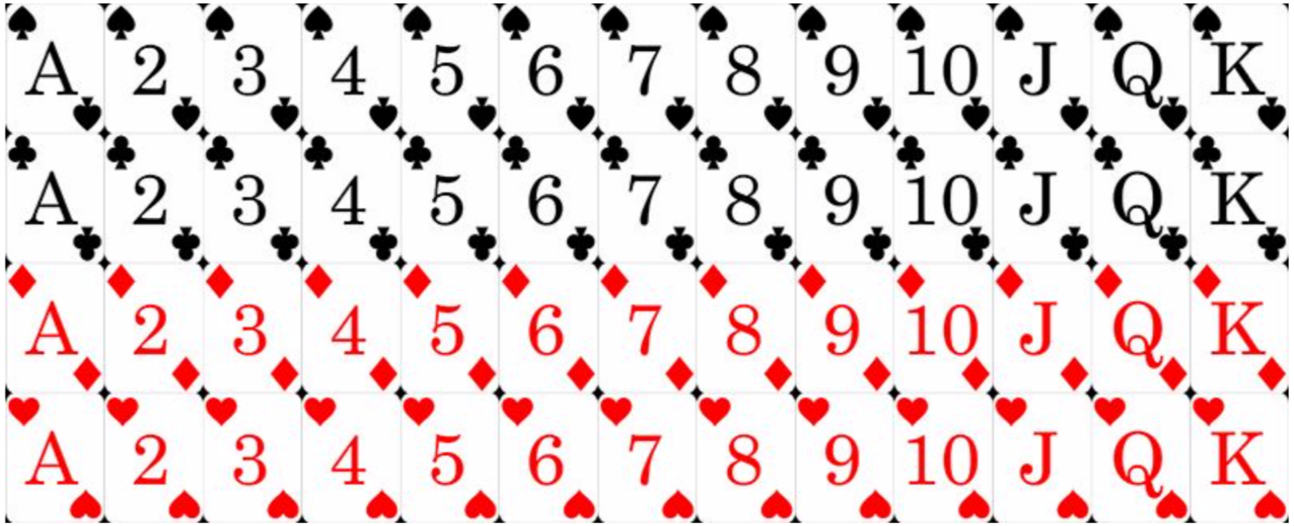
CSCI 3202: Intro to Artificial Intelligence

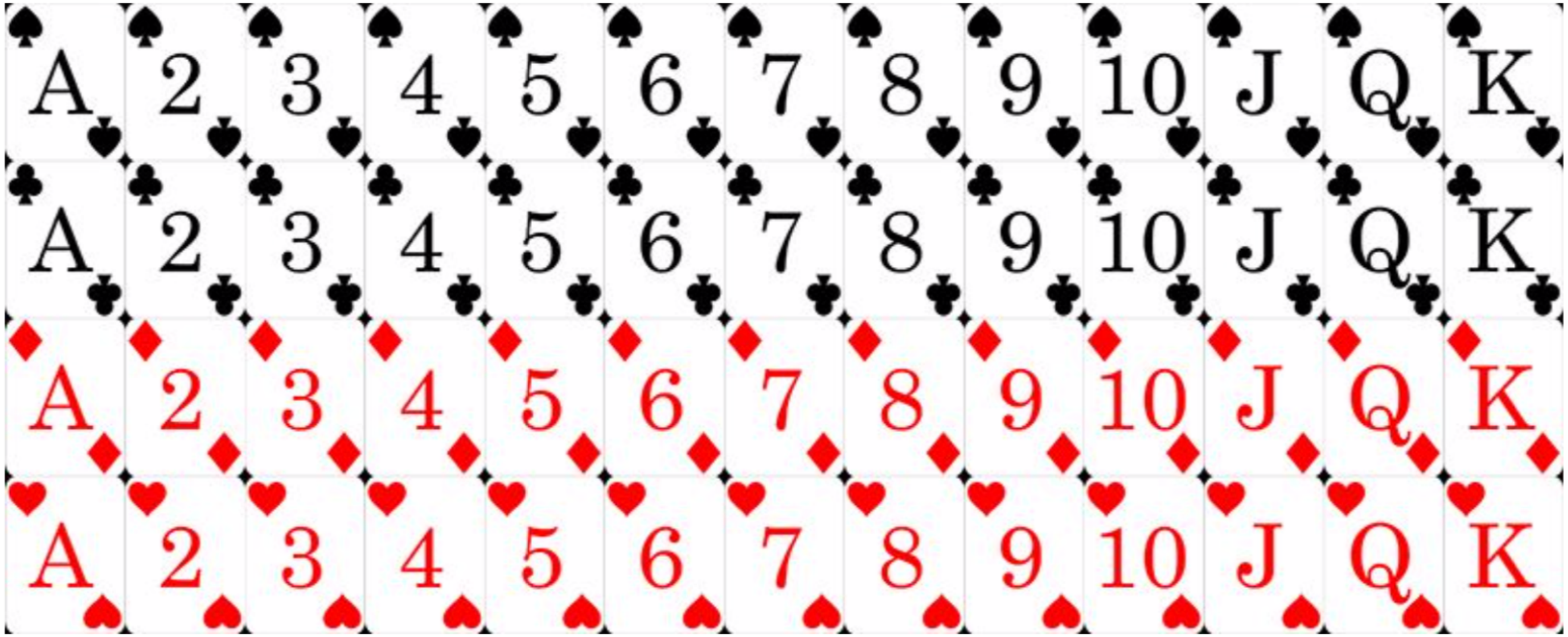
Lecture 20: Probability

Christopher Godley

Department of Computer Science



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 $V = v, C = c,\ p(V = v, C = c),$ is the probability that the card value is v and the card color is c simultaneously.

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

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

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

Example: $p(V = 6 \mid C = red) =$

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

Probability Review

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

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

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

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

In shorthand:

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

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 $p(C)$).

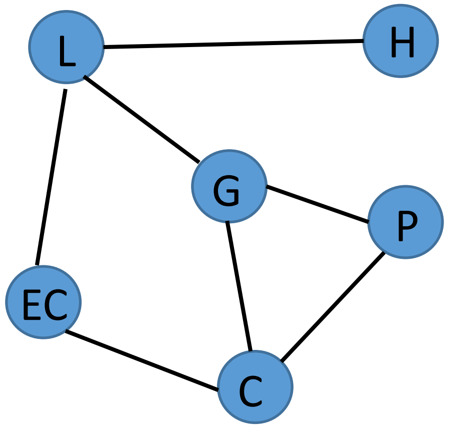
We can calculate the marginal probability using our old friend:

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

Probability Review

Example: Consider a random walk on the graph shown. Let the nodes represent locations on campus defined as follows:

EC = Engineering Center
C = Coffee Shop
G = Gym
L = Library
P = Physic Building
H = Home



Find the probability of going to the Coffee Shop on either your first, second, or third stop. Note: You must move from location to location on each move, i.e. you may not stay put in your current location. Assume that your starting location is the Engineering Center. Assume that all successor nodes/locations have an equal probability of being moved to.

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

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?

Probability Review

Example: (continued)

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

Probability Review

Example: (continued)

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

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?

Probability Review

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

$$p(T_2, T_1 \mid U) = p(T_2 \mid U)\, p(T_1 \mid U)$$

- 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:

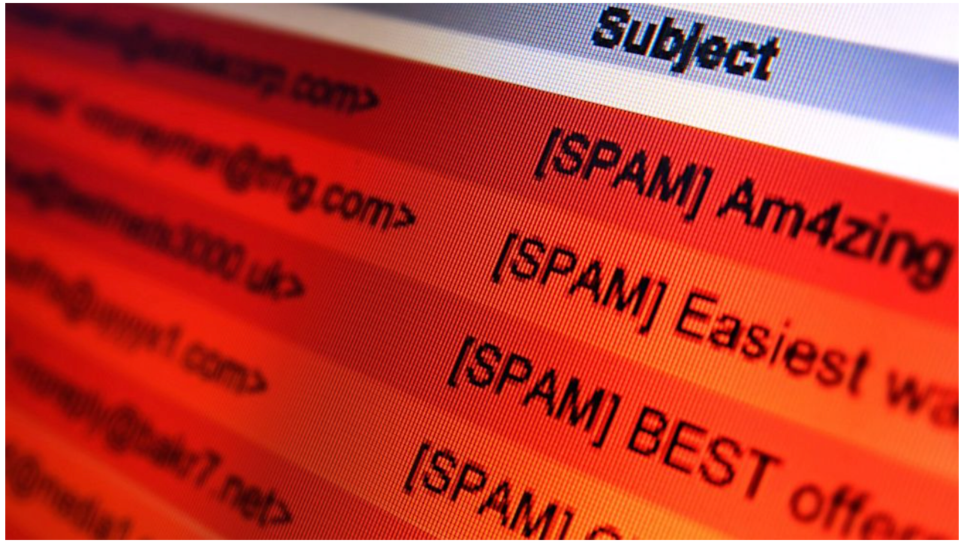
$$p(x = [buy, pills, deal] \mid y = \text{spam}) = p(\, buy \mid \text{spam})\, p(pills \mid \text{spam})\, p(deal \mid \text{spam})$$

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

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

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, $p(y)$:

- 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: $p(\text{ham}) = \dfrac{\text{\# ham emails in training data}}{\text{\# emails total in training data}}$

Probability Review

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

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:

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

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*	

Probability Review

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

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

Next Time

- *More Bayesian Reasoning*