DSC 40A

Theoretical Foundations of Data Science I

Announcements

- Homework 6 due Monday
- Homework 7 will be released 11/27 and due 12/6.

* Class on Wednesday as usual

Question

Answer at q.dsc40a.com

Remember, you can always ask questions at q.dsc40a.com!

If the direct link doesn't work, click the "Lecture Questions" link in the top right corner of dsc40a.com.

Last Time

- We defined Bayes' Theorem:

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)} = \frac{P(A,B)}{P(A)}$$

- Bayes' Theorem describes how to update the probability of one event given that another has occurred.

Today

- What does it mean for one event not to influence the probability of another?
- Independence and conditional independence.

Independence

Updating Probabilities

- Bayes' Theorem describes how to update the probability of one event given that another has occurred.

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)}$$

if $\frac{P(A|B)}{P(A)} > 1 \Rightarrow$ A occurring increases prob of B

if $\frac{P(A|B)}{P(A)} < 1 \Rightarrow$ A occuring decrease prob of B

Updating Probabilities

- P(A) is our prior belief that A happens.
- P(A | B) is our updated belief that A happens, now that we know B happens.
- Sometimes knowing that B happens doesn't change anything.
- Sometimes, P(B|A) = P(B). Knowing that A occurs doesn't't change anything.

Example

We flip a fair coin twice.

$P(2=H \mid 1=H) = P(2=H)$

- P(Second Flip = Heads) = $\frac{1}{2}$
- P(Second Flip = Heads | First Flip = Heads) = $\frac{1}{2}$

Independent Events

- A and B are **independent events** if one event occurring does not affect the chance of the other event occurring.

$$P(B|A) = P(B) \qquad\qquad P(A|B) = P(A)$$

- Otherwise, A and B are **dependent events**.

If one of the above is true, must the other be true?

A. yes
B. not necessarily

prove using Bayes

Independent Events

- A and B are independent events if one event occurring does not affect the chance of the other event occurring.

assume $\qquad$ prove

$$P(B|A) = P(B) \qquad P(A|B) = P(A)$$

prove $\qquad$ assume

- Using Bayes' Theorem, if one is true, then so is the other.

$$P(B|A) = \frac{P(A|B) * \cancel{P(B)}}{P(A)} = \cancel{P(B)}$$

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

Independent Events

- A and B are independent events if and only if:

$$\underbrace{P(A \text{ and } B)}^{0} = \overbrace{P(A)}^{0} * \overbrace{P(B)}^{\text{anything}}$$

$P(A \cap B)$, $P(A, B)$

$\Rightarrow P(B|A) = P(B)$

- This more general definition allows for the probability of A or B to be zero.

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

might be zero

Mutual exclusivity and Independence

$P(A) > 0$ $P(B > 0)$

Suppose *A* and *B* are two events with non-zero probability.
Is it possible for *A* and *B* to be both mutually exclusive and independent?

A. Yes
B. No
C. It depends on *A* and *B*

$\rightarrow A \cap B = \phi$

$\Rightarrow P(A \cap B) = 0$

Mutual exclusivity and Independence

- When two events are **mutually exclusive**, it is impossible for them to happen together:

$$P(A \cap B) = 0$$

- When two events are **independent**:

$$P(A \cap B) = \mathrm{P}(\overset{\neq 0}{A}) \cdot P(\overset{\neq 0}{B}) \quad \neq 0$$

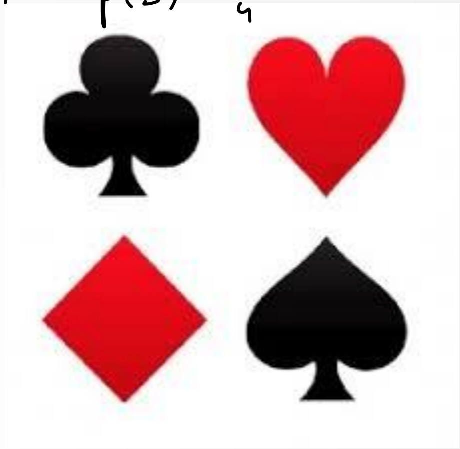
- Thus if they are both **mutually exclusive** and **independent** then at least one of them must have **zero** probability.

Example

1) You draw two cards, one at a time, **with replacement**.
- A is the event that the first card is a heart.
- B is the event that the second card is a club.

2) You draw two cards, one at a time, **without replacement**.
- A is the event that the first card is a heart.
- B is the event that the second card is a club.

S $\quad P(S) = \frac{1}{4}$



$P(A) = \frac{1}{4} = \text{Prob}(\text{suit})$

1) $P(B|A) = P(B) = \frac{1}{4}$

2) $P(B|A) > P(B)$

Are A and B independent?
A. yes in both cases
B. yes with replacement, no without replacement
C. no with replacement, yes without replacement
D. no in both cases

Example

You draw one card from a deck of 52.

- A is the event that the card is a heart. $P(A) = \frac{1}{4}$
- B is the event that the card is a face card (J, Q, K).

Are A and B independent?

A. yes

B. no

♥:	2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A
♦:	2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A
♣:	2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A
♠:	2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

$$P(B) = \frac{\#\text{face}}{\#\text{cards}} = \frac{12}{52} = \frac{3}{13}$$

$$P(B|A) = \frac{3}{13} = P(B)$$

\# card both face and heart

$$P(A \cap B) = \frac{3}{52} = P(A) \cdot P(B) = \frac{1}{4} \cdot \frac{3}{13}$$

$$P(B) = \frac{\#\text{outcomes in } B}{\#\text{outcomes in } S}$$

$$P(B|A) = \frac{\#\text{outcomes in } A \text{ and } B}{\#\text{outcomes in } A}$$

Assuming Independence

- Sometimes we assume that events are independent to make calculations easier.
- Real-world events are almost never exactly independent, but may be close.
- Example: A is event that a student is a data science major, B is the event that they bike to campus.

Example

1% of UCSD students are data science majors. 25% of UCSD students eat avocado toast for breakfast. Assuming that being a DSC major and eating avocado toast for breakfast are independent:

a) What percentage of DSC majors eat avocado toast for breakfast?

$P(\text{avocado} \mid \text{DSC}) = P(\text{avocado}) = 25\%$

a) What percentage of UCSD students are DSC majors who eat avocado toast for breakfast?

$P(\text{avocado} \cap \text{DSC}) = P(\text{avocado}) \cdot P(\text{DSC}) = 0.25\%$

Conditional Independence

Conditional Independence

- Sometimes, events that are dependent *become* independent upon learning some new information.
- Or independent events can become dependent, given additional information.

Example

Oops, you lost the King of Clubs! Draw one card from a deck of 51.

- A is the event that the card is a heart. $P(A)$
- B is the event that the card is a face card (J, Q, K). $P(B) = \frac{11}{51}$

♥: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A
♦: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A
♣: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, A
♠: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

Are A and B independent?

a) yes

$\boxed{\text{b) no}}$

$P(B|A) = P(\text{face given } ♡) = \frac{3}{13}$

$P(B|A) \neq P(B)$

Example

Oops, you lost the King of Clubs! Draw one card from a deck of 51.

- A is the event that the card is a heart. $P(A) = \frac{13}{51}$
- B is the event that the card is a face card (J, Q, K).

* C the card is red

♥:	2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A
♦:	2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A
♣:	2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, A
♠:	2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A

Now suppose you learn that the card is red. Are A and B independent given this new information?

A. yes B. no

$P(A \cap B \mid C) = \frac{3}{26}$ $P(A \mid C) = \frac{1}{2}$ $P(B \mid C) = \frac{6}{26}$

$\Rightarrow P(A \cap B \mid C) = P(A \mid C)\, P(B \mid C)$

Conditional Independence

- Recall that A and B are independent if

$$P(A \text{ and } B) = P(A) * P(B)$$

- A and B are conditionally independent given C if and only if

$$P((A \text{ and } B)|C) = P(A|C) * P(B|C)$$

and $P(C)>0$

- Given that C occurs, this says that A and B are independent of one another.

Assuming Conditional Independence

- Sometimes we assume that events are conditionally independent to make calculations easier.
- Real-world events are almost never conditionally independent, but may be close.

Example

Suppose that 80% of UCSD students use TikTok and 25% of UCSD students eat avocado toast for breakfast. What is the probability that a random UCSD student uses TikTok and eats avocado toast for breakfast, assuming that these events are conditionally independent given that a person is a UCSD student?

$$P(\text{avocado} \cap \text{tiktok} \mid UCSD) = P(\text{avo} \mid UCSD) \cdot P(\text{tiktok} \mid UCSD)$$

$$= 80\% \cdot 25\% = 20\%$$

Independence vs. Conditional Independence

- Is it reasonable to assume conditional independence of
  - using TikTok
  - eating avocado toast for breakfast

  given that a person is a UCSD student?
- Is it reasonable to assume independence of these events in general, among all people?

Which assumptions do you think are reasonable?

A. both
B. conditional independence only
C. independence (in general) only
D. neither

Independence vs. Conditional Independence

- In general, there is **no relationship** between independence and conditional independence.

Summary

- Events are independent when knowledge of one event does not change the probability of the other event.
- Events that are not independent *can* be conditionally independent given new information (and the opposite is true).
- **Next time:** Solving the classification problem using Bayes' Theorem and an assumption of conditional independence.