DSC 40A

Theoretical Foundations of Data Science I

Last Time

- We defined Bayes' Theorem:

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

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

In This Video

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

Updating Probabilities

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

ratio

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

- Sometimes, P(B|A) = P(B). Knowing that A occurs doesn't change anything.

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

Independent Events

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

assume

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

similar

conclude

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

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

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

ratio = 1

Independent Events

mult. rule

$P(A \cap B) = P(A) * P(B|A)$

$= P(A) * P(B)$

when A, B independent

- A and B are independent events if

if $P(A) = 0$, 0 0 anything

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

→ satisfied

⇒ if $P(A) = 0$, and B is any event, then A, B are ind.

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

problem: if $P(A) = 0$,

$P(B|A) = \dfrac{P(A \text{ and } B)}{P(A)}$ ← divide by zero

$P(B|A) = P(B)$ not sufficient

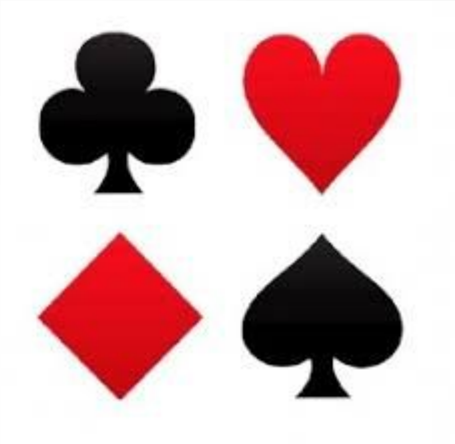
Example

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.

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.



$P(B|A) = P(B)$ with replacement

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

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.
- B is the event that the card is a face card (J, Q, K).

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

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

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

Are A and B independent?

A. yes

B. no

S

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

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

B

$$\underset{\frac{3}{13}}{P(B|A)} = \underset{\frac{12}{52}}{P(B)}$$

all outcomes in S are equally likely, so

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

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

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

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?

$P(avo \mid DSC) = P(avo) = 25\%$

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

$P(avo \cap DSC) = P(avo) * P(DSC) = 0.25\%$

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.
- B is the event that the card is a face card (J, Q, K).

A S

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

B

Are A and B independent?

NO

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

$P(B) = \frac{11}{51}$

not same

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.
- B is the event that the card is a face card (J, Q, K).

C = card is red

	A		B	
♥:	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?

$\frac{6}{26} = \frac{3}{13}$ YES

$$\underbrace{P(A\cap B\mid C)}_{3/26} = \underbrace{P(A\mid C)}_{1/2} * \underbrace{P(B\mid C)}_{6/26}$$

(cond. ind.)

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

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

- 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 like Harry Potter and 25% of UCSD students eat avocado toast for breakfast. What is the probability that a random UCSD student likes Harry Potter and eats avocado toast for breakfast, assuming that these events are conditionally independent given that a person is a UCSD student?

$$P((HP \cap avo) | UCSD) = P(HP | UCSD) * P(avo | UCSD)$$

$$= 80\% * 25\%$$

$$= 20\%$$

Independence vs. Conditional Independence

- Is it reasonable to assume conditional independence of
  - liking Harry Potter
  - 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.