DSC 40A

Theoretical Foundations of Data Science I

In This Video

- We’ll define the Law of Total Probability and Bayes Theorem.

Getting to Campus

- You conduct a survey:
  - How did you get to campus today? Walk, bike, or drive?
  - Were you late?

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

Getting to Campus

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

What is the probability that a randomly selected person is late?

A. 24%
B. 30%
C. 45%
D. 50%

Getting to Campus

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

- Since everyone either walks, bikes, or drives,

P(Late) = P(Late AND Walk) + P(Late AND Bike) + P(Late AND Drive)

- This is called the **Law of Total Probability**.

Getting to Campus

	Late	Not Late
Walk	6% +	24% = 30%
Bike	3%	7%
Drive	36%	24%

Suppose someone tells you that they walked. What is the probability that they were late?

A. 6%
B. 20%
C. 25%
D. 45%

$$P(\text{late} \mid \text{walk}) = \frac{P(\text{late AND walk})}{P(\text{walk})}$$

$$P(\text{late AND walk}) = P(\text{walk}) \times P(\text{late} \mid \text{walk}) \quad \frac{6\%}{30\%} = \frac{1}{5}$$

Getting to Campus

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

- Since everyone either walks, bikes, or drives,

P(Late) = P(Late AND Walk) + P(Late AND Bike) + P(Late AND Drive)

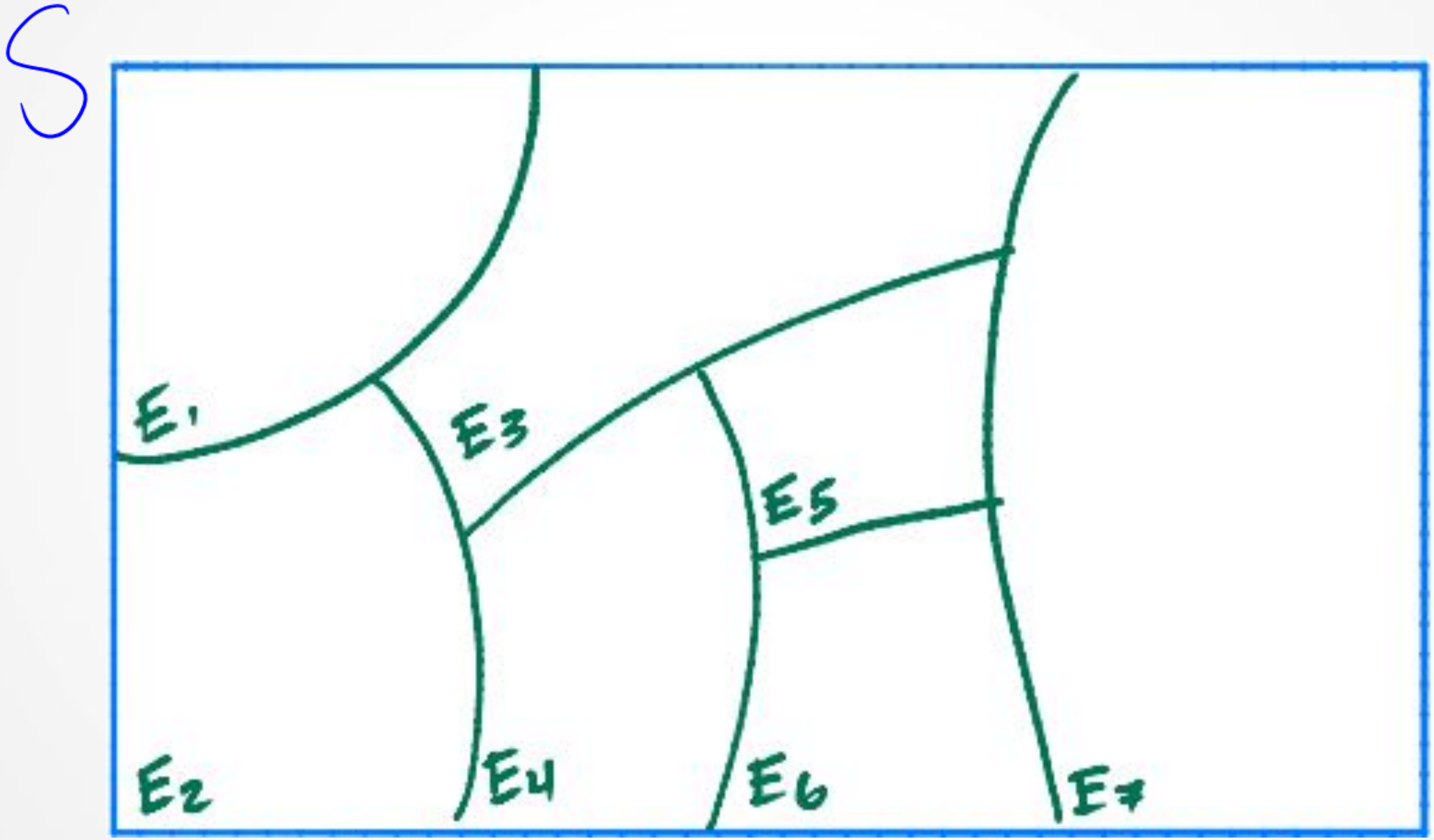
P(Late) = P(Late|Walk)*P(Walk) + P(Late|Bike)*P(Bike)+P(Late|Drive)*P(Drive)

Partitions

- A set of events $E_1, E_2, \ldots, E_k$ is a **partition** of S if
  - $P(E_i \cap E_j) = 0$ for all i,j ← mutually exclusive, no overlap
  - $P(E_1) + P(E_2) + \ldots + P(E_k) = 1$

every $s \in S$ is in exactly one of $E_1, \ldots, E_k$

Partitions



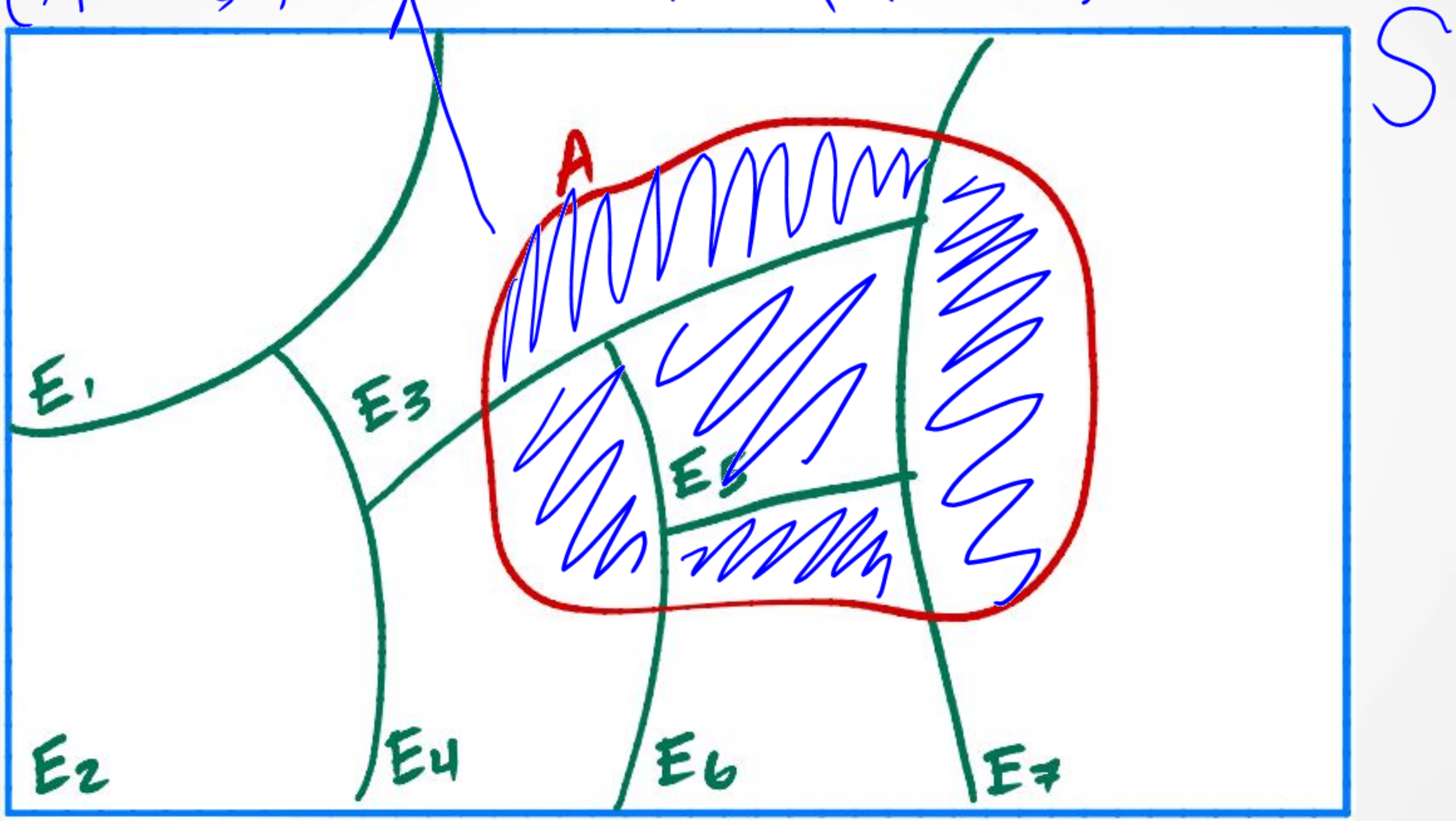
Law of Total Probability

- If A is an event and E_1, E_2, …, E_k is a **partition** of S, then

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \dots + P(A \cap E_k)$$
$$= \sum_{i=1}^{k} P(A \cap E_i)$$

Partitions

$P(A \cap E_1) + P(A \cap E_2) + P(A \cap E_3) + P(A \cap E_4) + P(A \cap E_5) + \ldots$



Law of Total Probability

- If A is an event and $E_1, E_2, \ldots, E_k$ is a **partition** of S, then

$$P(A) = \underline{P(A \cap E_1)} + P(A \cap E_2) + \ldots + P(A \cap E_k)$$
$$= \sum_{i=1}^{k} P(A \cap E_i)$$

from mult. rule or conditional prob.

- Written another way,

$$P(A) = \underline{P(A \mid E_1) \cdot P(E_1)} + \ldots + P(A \mid E_k) \cdot P(E_k)$$
$$= \sum_{i=1}^{k} P(A \mid E_i) \cdot P(E_i)$$

Getting to Campus

	Late	Not Late
Walk	6%	24%
Bike	3%	7%
Drive	36%	24%

45%

Suppose someone is late. What is the probability that they walked? Choose the best answer.

A. Close to 5%
B. Close to 15%
C. Close to 30%
D. Close to 40%

$$\frac{6}{45} \approx 0.133 \approx 13\%$$

$$P(\text{walk} \mid \text{late}) = \frac{P(\text{walk AND late})}{P(\text{late})}$$

Getting to Campus

- Suppose all you know is
  - P(Late) = 45%
  - P(Walk) = 30%
  - P(Late|Walk) = 20%
- Can you still find P(Walk|Late)?

$$P(Walk|Late) = \frac{P(Walk\ AND\ Late)}{P(Late)}$$

$$= \frac{P(Late|Walk) \times P(Walk)}{P(Late)} = \frac{0.2 \times 0.3}{0.45}$$

$$\approx 0.133$$

Bayes' Theorem

Bayes' Theorem follows from the multiplication rule, or conditional probability.

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

Bayes' Theorem:

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

← can use law of total prob

Bayes' Theorem

Bayes' Theorem follows from the multiplication rule, or conditional probability.

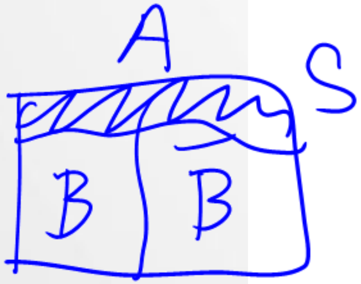
$$P(A) * P(B|A) = P(A \text{ and } B) = P(B) * P(A|B)$$

Bayes' Theorem:

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

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

not B



Bayes' Theorem: Example

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

A manufacturer claims that its drug test will **detect steroid use 95% of the time**. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What's the probability that he used steroids?

What is your first guess?

A. Close to 95%
B. Close to 85%
C. Close to 40%
D. Close to 15%

Bayes' Theorem: Example

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

A manufacturer claims that its drug test will **detect steroid use 95% of the time**. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What's the probability that he used steroids?

Now, calculate it and choose the best answer.

A. Close to 95%
B. Close to 85%
C. Close to 40%
D. Close to 15%

Bayes' Theorem: Example

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

use steroids   test pos

A manufacturer claims that its drug test will **detect steroid use 95% of the time**. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What's the probability that he used steroids?

Solution:

$P(B|A) \leftarrow ?$

B: used steroids

$P(A|B) = 0.95$ $P(A|\overline{B}) = 0.15$

A: tested positive

$P(B) = 0.10$

$P(\overline{B}) = 0.90$

Bayes' Theorem: Example

$$P(B|A) = \frac{P(A|B) * P(B)}{P(B) * P(A|B) + P(\overline{B}) * P(A|\overline{B})} = \boxed{\frac{0.95 * 0.1}{0.1 * 0.95 + 0.9 * 0.15} \approx 0.41}$$

A manufacturer claims that its drug test will **detect steroid use 95% of the time.** What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What's the probability that he used steroids?

Solution:

B: used steroids

A: tested positive

Despite manufacturer's claims, only **41% chance** that cyclist used steroids.

Preview: Bayes' Theorem for Classification

Bayes' Theorem is very useful for classification problems, where we want to predict a class based on some features.

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

B = belonging to a certain class
A = having certain features

$$P(\text{class}|\text{features}) = \frac{P(\text{features}|\text{class}) * P(\text{class})}{P(\text{features})}$$

Summary

- When a set of events partitions the sample space, the law of total probability applies.

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \dots + P(A \cap E_k)$$
$$= \sum_{i=1}^{k} P(A \cap E_i)$$

- Bayes Theorem says how to express P(B|A) in terms of P(A|B).
- **Next time:** independence and conditional independence