Discussion 7

23 Oct 23

Announcements

- Homework 5 due this Wednesday
- OH from 5 to 7pm at CDS B64
- Median score for midterm 30/39

Review of Conditional Probability

The conditional probability of an event B is the probability that the event will occur given the knowledge that an event A has already occurred.

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

A Thought Experiment - Intuition behind Bayes

Cara lives in the Boston and loves to read. She has an eye for detail and is creative. She loves to paint.

Given this description, what do you think is more likely:

1. Cara is an artist
2. Cara works as a SDE in Silicon Valley

A Thought Experiment

Given Cara's interests, it seems logical that she should be an artist.

Here are some stats for our problem:
0.6% of women are artists in Boston.
65% of women in Boston work in tech.

What do you think now is more likely to be Cara's profession?

Bayes Theorem - A Thought Experiment

Say about 50% of artists fit the description, and 10% of software developers fit the given criteria.

Even if this were true, it seems more likely that Cara works in tech, even if her description is similar to one an artist might have.

Based on this description, let us try to define Bayes theorem in our own words.

Bayes Theorem - Intuition

Based on the experiment, we can conclude that a new evidence cannot determine our beliefs in a vacuum; it should update prior beliefs.

So, at its core, we need to find the probability that our hypothesis is true given an event occured.

Bayes Theorem - Formula

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

A, B	= events
$P(A\|B)$	= probability of A given B is true
$P(B\|A)$	= probability of B given A is true
$P(A), P(B)$	= the independent probabilities of A and B

Problem 1: Rare disease

A person is not feeling well. He goes to the doctor, who suggests to run a battery of tests. Turns out he tested positive for a very rare disease that affects about 0.1% of the population. So he asks the doctor "You know, how certain is it that I have this disease?" and she says "Well, the test will correctly identify 99% of people that have the disease and only incorrectly identify 1% of people who don't have the disease".

What are the chances that he actually has this disease?

Solution

P(D = person has the disease) = 0.001 (0.1% of the population has the disease).

P(D') = 0.999 (1 - P(D)).

P(P|D) = 0.99 (the test correctly identifies 99% of those who have the disease).

P(P|D') = 0.01 (the test incorrectly identifies 1% of those who do not have the disease)

We want to find P(D|P), the probability that the person has the disease given a positive test result.

$$P(D|P) = \frac{P(D) \cdot P(P|D)}{P(D) \cdot P(P|D) + P(D') \cdot P(P|D')}$$

$$= \frac{0.001 \cdot 0.99}{0.001 \cdot 0.99 + 0.999 \cdot 0.01}$$

$$= \frac{0.00099}{0.00099 + 0.00999} = \frac{0.00099}{0.01} = 0.099$$

Diachronic Bayes

- Let's say you live somewhere where cats are sometimes spotted.
- Your mother calls you, and tells you that she saw a cat and asks you to bring cat food.
- You don't know which cat, so you don't know which cat food you have to bring with you.
- So you tell yourself: "We have a lot of orange cats here, so it was probably an orange cat, so I'll bring oranges".
- Once you're there, you see black cat hair in the middle of the garden, and you realize that it’s black cats.

Diachronic Bayes

You have priors on the value of parameters, here the type of cat, but once you witness some evidence, the data, you can update your priors using Bayes' theorem and hopefully reasonably approximate the true value of the parameter. If multi-colored cats also leave only black hair, you'll need some other evidence, and so on.

$$P(H \mid D) = \frac{P(H)P(D \mid H)}{P(D)}$$

Prior: P(H) = probability of the hypothesis before we see the data

Posterior: P(H|D) = probability of the hypothesis after we see the data

Likelihood: P(D|H) = probability of the data under the hypothesis

Total prob of data: P(D) = total probability of the data under any hypothesis

Problem 2

Imagine you have two bags of colored marbles. Bag 1 contains 30 red marbles and 20 green marbles, while Bag 2 contains 40 red marbles and 10 green marbles. You randomly choose one of the bags and then randomly pick a marble from that bag. If you pick a green marble, what is the probability that you chose Bag 2?

Solution

Event	Probability	Explanation
$P(B_1)$	0.5	Prior probability of choosing Bag 1
$P(B_2)$	0.5	Prior probability of choosing Bag 2
$P(G\|B_1)$	0.4	Likelihood of picking a green marble from Bag 1
$P(G\|B_2)$	0.2	Likelihood of picking a green marble from Bag 2

The first step is to define a prior.

Prior probability of choosing Bag 1, P(B1) = 0.5

Prior probability of choosing Bag 2, denoted as P(B2) = 0.5

Bag	Prior $P(B)$
Bag 1	0.5
Bag 2	0.5

Now,

Likelihood	Value
$P(G\|B_1)$	$\frac{\text{Number of green marbles in Bag 1}}{\text{Total number of marbles in Bag 1}} = \frac{20}{50} = 0.4$
$P(G\|B_2)$	$\frac{\text{Number of green marbles in Bag 2}}{\text{Total number of marbles in Bag 2}} = \frac{10}{50} = 0.2$

So, we can update our table as follows:

Bag	Prior $P(B)$	Likelihood $P(G\|B)$
Bag 1	0.5	0.4
Bag 2	0.5	0.2

And finally,

Bag	Prior $P(B)$	Likelihood $P(G\|B)$	Unnormalized $P(B) \cdot P(G\|B)$
Bag 1	0.5	0.4	$0.5 \cdot 0.4 = 0.2$
Bag 2	0.5	0.2	$0.5 \cdot 0.2 = 0.1$

Thank you