EECS 391
Intro to AI

Reasoning with Bayes' Rule

L12 Tue Oct 10

How do you model a world?

How to you reason about it?

Simple example: medical test results

- Test report for rare disease is positive, 90% accurate
- What's the probability that you have the disease?
- What if the test is repeated?

- This is the simplest example of reasoning by combining sources of information.

How do we model the problem?

- Which is the correct description of "Test is 90% accurate" ?

$$
\begin{aligned}
P(T = \text{true}) &= 0.9 \\
P(T = \text{true} | D = \text{true}) &= 0.9 \\
P(D = \text{true} | T = \text{true}) &= 0.9
\end{aligned}
$$

- What do we want to know?

$$
\begin{gathered}
P(T = \text{true}) \\
P(T = \text{true} | D = \text{true}) \\
P(D = \text{true} | T = \text{true})
\end{gathered}
$$

- More compact notation:

$$
\begin{aligned}
P(T = \text{true} | D = \text{true}) &\rightarrow P(T|D) \\
P(T = \text{false} | D = \text{false}) &\rightarrow P(\bar{T}|\bar{D})
\end{aligned}
$$

Evaluating the posterior probability through Bayesian inference

- We want P(D|T) = “The probability of the having the disease given a positive test”
- Use Bayes rule to relate it to what we know: P(T|D)

likelihood *prior*

$$\text{posterior} \quad P(D|T) = \frac{P(T|D)P(D)}{P(T)}$$

normalizing constant

- What’s the prior P(D)?
- Disease is rare, so let’s assume

$$P(D) = 0.001$$

- What about P(T)?
- What’s the interpretation of that?

Evaluating the normalizing constant

$$\underbrace{P(D|T)}_{\text{posterior}} = \frac{\overbrace{P(T|D)}^{\text{likelihood}}\overbrace{P(D)}^{\text{prior}}}{\underbrace{P(T)}_{\text{normalizing constant}}}$$

- P(T) is the marginal probability of P(T,D) = P(T|D) P(D)
- So, compute with summation

$$P(T) = \sum_{\text{all values of D}} P(T|D)P(D)$$

- For true or false propositions:

$$P(T) = P(T|D)P(D) + P(T|\bar{D})P(\bar{D})$$

What are these?

Refining our model of the test

- We also have to consider the negative case to incorporate all information:

$$\begin{aligned} P(T|D) &= 0.9 \\ P(T|\bar{D}) &= ? \end{aligned}$$

- What should it be?
  - It can be any value between 0 and 1. It does not depend on P(T|D).

- What about this:

$$P(T|D) = 0.9 \quad \Rightarrow \quad P(\bar{T}|D) = 0.1$$

- This is because P(T|D) must be a valid probability distribution

$$\sum_{T=\text{True},T=\text{False}} P(T|D) = 1$$

False positives and false negatives

- What is the expression for the probability of false positives?
  - The probability that the test is true given the disease is false:

$$P(T = \text{True}|D = \text{False})$$

- What is the expression for the probability of false negatives?
  - The probability that the test is false given the disease is true.

$$P(T = \text{False}|D = \text{True})$$

- What would you call $P(T = \text{True}|D = \text{True})$ and $P(T = \text{False}|D = \text{False})$?
- The probability of a true positive and a true negative.
- This is closer to what is meant when we say the test is “90% accurate”.

Plugging in the numbers

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})}$$

$$P(D|T) = \frac{0.9 \times 0.001}{0.9 \times 0.001 + 0.1 \times 0.999} = 0.0089$$

Note: here we assume the false positive rate and the false negative rate are the same.

Same problem different situation

- Suppose we have a test to determine if you won the lottery.
- It's 90% accurate.
- What is P($ = true | T = true) then?

Playing around with the numbers

$$P(D|T) \quad = \quad \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})}$$

- What if the test were 100% reliable?

$$P(D|T) = \frac{1.0 \times 0.001}{1.0 \times 0.001 + 0.0 \times 0.999} = 1.0$$

- What if the test was the same, but disease wasn't so rare?

$$P(D|T) = \frac{0.9 \times 0.1}{0.9 \times 0.1 + 0.1 \times 0.9} = 0.5$$

Repeating the test

- We can relax, P(D|T) = 0.0089, right?
- Just to be sure the doctor recommends repeating the test.
- How do we represent this?

$$P(D|T_1, T_2)$$

- Again, we apply Bayes' rule

$$P(D|T_1, T_2) = \frac{P(T_1, T_2|D)P(D)}{P(T_1, T_2)}$$

- How do we model $P(T_1,T_2|D)$?

Modeling repeated tests

$$P(D|T_1, T_2) = \frac{P(T_1, T_2|D)P(D)}{P(T_1, T_2)}$$

- Easiest is to assume the tests are *independent*.

$$P(T_1, T_2|D) = P(T_1|D)P(T_2|D)$$

- This also implies:

$$P(T_1, T_2) = P(T_1)P(T_2)$$

- Plugging these in, we have

$$P(D|T_1, T_2) = \frac{P(T_1|D)P(T_2|D)P(D)}{P(T_1)P(T_2)}$$

Evaluating the normalizing constant again

- Expanding as before we have

$$P(D|T_1, T_2) = \frac{P(T_1|D)P(T_2|D)P(D)}{\sum_{D=\{t,f\}} P(T_1|D)P(T_2|D)P(D)}$$

- Plugging in the numbers gives us

$$P(D|T) = \frac{0.9 \times 0.9 \times 0.001}{0.9 \times 0.9 \times 0.001 + 0.1 \times 0.1 \times 0.999} = 0.075$$

- Another way to think about this:
  - What's the chance of 1 false positive from the test?
  - What's the chance of 2 false positives?

- The chance of 2 false positives is still 10x more likely than the a prior probability of having the disease.

Simpler: Combining information the Bayesian way

- Let's look at the equation again:

$$P(D|T_1, T_2) = \frac{P(T_1|D)P(T_2|D)P(D)}{P(T_1)P(T_2)}$$

- If we rearrange slightly:

$$P(D|T_1, T_2) = \frac{P(T_2|D)P(T_1|D)P(D)}{P(T_2)P(T_1)}$$

We've seen this before!

- It's the posterior for the first test, which we just computed

$$P(D|T_1) = \frac{P(T_1|D)P(D)}{P(T_1)}$$

The old posterior is the new prior

- We can just plugin the value of the old posterior
- It plays exactly the same role as our old prior

$$P(D|T_1, T_2) = \frac{P(T_2|D)P(T_1|D)P(D)}{P(T_2)P(T_1)}$$

$$P(D|T_1, T2) = \frac{P(T_2|D) \times 0.0089}{P(T_2)}$$

This is how Bayesian reasoning combines old information with new information to update our belief states.

- Plugging in the numbers gives the same answer:

$$P(D|T) \quad = \quad \frac{P(T|D)P'(D)}{P(T|D)P'(D) + P(T|\bar{D})P'(\bar{D})}$$

$$P(D|T) = \frac{0.9 \times 0.0089}{0.9 \times 0.0089 + 0.1 \times 0.9911} = 0.075$$

Generalizing sequential inference

- What if we were still worried and wanted to do more tests?

$$p(D|T_{1:n}) = ?$$

Can write P() or p() or Pr()

- Reasoning is exactly the same

$$\begin{aligned}
P(D|T_{1:n}) &= \frac{P(D, T_{1:n})}{P(T_{1:n})} && \text{conditional probability} \\
&= \frac{P(T_{1:n}|D)P(D)}{P(T_{1:n})} && \text{Bayes' rule} \\
&= \frac{P(T_{1:n}|D)P(D)}{\sum_{D=t,f} P(D, T_{1:n})} && \text{marginalization} \\
&= \frac{P(T_{1:n}|D)P(D)}{\sum_{D=t,f} P(T_{1:n}|D)P(T_{1:n})} && \text{conditional probability} \\
&= \frac{\prod_i P(T_i|D)P(D)}{\sum_{D=t,f} \prod_i P(T_i|D)P(D)} && \text{independence assumption}
\end{aligned}$$

Another problem (Pearl, 1988)

- Suppose you are witness to a night-time hit and run accident involving a taxi in Athens.
- All the taxis in Athens are either blue or green.
- You swear under oath that the taxi was blue.
- But, under dim lighting conditions, discrimination between blue and green is only 75% reliable.
- It is possible to calculate the most likely color for the taxi?

Specification of problems

- Not all problems need be specified by the joint probability
- In the test-disease problem, we used the conditional probabilities:

$$P(D|T) \quad = \quad \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})}$$

$$P(D|T) = \frac{0.9 \times 0.001}{0.9 \times 0.001 + 0.1 \times 0.999} = 0.0089$$