EECS 491
Probabilistic Graphical Models

Probabilistic Reasoning

Probabilistic Reasoning

- book reading: Barber Ch. 1
- brief review basic probability
- expressing reasoning using probabilities
- reasoning using Bayes' rule

Motivating examples: Barber Example 1.2

- Doctors find that people with Kreuzfeld-Jacob disease (KJ) almost invariably ate hamburgers.
  - What is the probability that a hamburger eater will have Kreuzfeld-Jacob disease?
  - How do we express this relationship?
  - How do we compute the probabilities?
- What are the variables?
  - person has Kreuzfeld-Jacob's disease (or not) = K = T|F
  - person eats hamburgers (or not) = H = T|F
  - note the variables are binary; they represent a state of the world
- What are the assumptions?
  - eating hamburgers *causes* Kreuzfeld-Jacob's disease
- What is expression for the proposition we want to evaluate?
  - $P(K = T \mid H = T)$

Motivating examples: Barber Example 1.3

- Inspector Clouseau arrives at the scene of a crime.
  - The victim lies dead in the room alongside the possible murder weapon, a knife.
  - The butler and maid are the inspector's main suspects.
- What are the variables?
  - suspects: butler = B; maid = M
  - knife = K (or murder weapon = W ?)
- What do the variables represent?
  - B = true: butler is the murder
  - M = true: maid is the murder
  - K = true: knife was the murder weapon
- More generally we could write:
  - W = knife | rope | candlestick | etc.
  - S (for suspect) = butler | maid | Prof. Plum | etc.

Barber Example 1.3 (continued)

- What do we want to evaluate?
- In principle, could want to know everything: P(B, M, K)
  - This is the full joint probability: probabilities of all states of the world, e.g.
  - P(B = T, M = F, K=T)
  - "The probability that the Butler is the murderer, the Maid is not, and the knife was the murder weapon"
- Also might want: P(B|K)
  - What does this represent?
  - "The probability that the butler was the murderer assuming that ("given that") the knife was the murder weapon.
  - What about the maid?
- How do we specify the model?
  - Can't just enter probabilities for all entries in the table for P(B, M, K)
  - Need to specify the joint in terms of our real-world knowledge, and *then* derive these using the rules of probability

Axioms of probability

- Axioms (Kolmogorov):

  0 ≤ P(A) ≤ 1

  P(true) = 1

  P(false) = 0

  P(A or B) = P(A) + P(B) − P(A and B)

- Corollaries:
  - A single random variable must sum to 1:

$$\sum_{i=1}^{n} P(D = d_i) = 1$$

  - The joint probability of a set of variables must also sum to 1.
  - If A and B are mutually exclusive:

  P(A or B) = P(A) + P(B)

Rules of probability

- conditional probability

$$Pr(A|B) = \frac{Pr(A \,\mathrm{and}\, B)}{Pr(B)}, \qquad Pr(B) > 0$$

- corollary (Bayes' rule)

$$\begin{aligned} Pr(B|A)Pr(A) &= Pr(A \,\mathrm{and}\, B) = Pr(A|B)Pr(B) \\ \Rightarrow Pr(B|A) &= \frac{Pr(A|B)Pr(B)}{Pr(A)} \end{aligned}$$

Basic concepts

Making rational decisions when faced with uncertainty:

- *Probability*

  the precise representation of knowledge and uncertainty

- *Probability theory*

  how to optimally update your knowledge based on new information

- *Decision theory: probability theory + utility theory*

  how to use this information to achieve maximum expected utility

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*

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

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

likelihood *prior*

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

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?

- What about $P(\bar{D})$?

Plugging in the numbers

- Our complete expression is

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

- Plugging in the numbers we get:

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

- Does this make intuitive sense?

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*.

$$P(T_1, T_2|D) = P(T1|D)P(T_2|D)$$

- This also implies:

$$P(T_1, T_2) = P(T1)P(T_2)$$

- 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

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

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

- Plugging in the numbers gives the same answer:

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

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

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

Can use an identical approach to solve the Hamburgers problem

Example 1.2 (Hamburgers)**.** Consider the following fictitious scientific information: Doctors find that people with Kreuzfeld-Jacob disease (KJ) almost invariably ate hamburgers, thus $p(Hamburger\ Eater|KJ\ ) = 0.9$. The probability of an individual having KJ is currently rather low, about one in 100,000.

1. Assuming eating lots of hamburgers is rather widespread, say $p(Hamburger\ Eater) = 0.5$, what is the probability that a hamburger eater will have Kreuzfeld-Jacob disease?

   This may be computed as

   $$p(KJ\ |Hamburger\ Eater) = \frac{p(Hamburger\ Eater, KJ\ )}{p(Hamburger\ Eater)} = \frac{p(Hamburger\ Eater|KJ\ )p(KJ\ )}{p(Hamburger\ Eater)} \tag{1.2.1}$$

   $$= \frac{\frac{9}{10} \times \frac{1}{100000}}{\frac{1}{2}} = 1.8 \times 10^{-5} \tag{1.2.2}$$

2. If the fraction of people eating hamburgers was rather small, $p(Hamburger\ Eater) = 0.001$, what is the probability that a regular hamburger eater will have Kreuzfeld-Jacob disease? Repeating the above calculation, this is given by

   $$\frac{\frac{9}{10} \times \frac{1}{100000}}{\frac{1}{1000}} \approx 1/100 \tag{1.2.3}$$

   This is much higher than in scenario (1) since here we can be more sure that eating hamburgers is related to the illness.

Return to Inspector Clouseau

- Variables: B | M = Butler | Maid was murder, K = Knife was used
- Need to convert Inspector's real-world knowledge into probabilities.
- He might assign these *a priori* probabilities:
  - p(B = murder) = 0.6 p(M = murderer) = 0.2
- What about other knowledge? Barber writes:

$$
\begin{aligned}
&p(\text{knife used}|B = \text{not murderer}, && M = \text{not murderer}) && = 0.3 \\
&p(\text{knife used}|B = \text{not murderer}, && M = \text{murderer}) && = 0.2 \\
&p(\text{knife used}|B = \text{murderer}, && M = \text{not murderer}) && = 0.6 \\
&p(\text{knife used}|B = \text{murderer}, && M = \text{murderer}) && = 0.1
\end{aligned}
$$

- These are arbitrary, but can imagine other cases:
  - p(rope used | G = strong gardener is murderer) = 0.4 // possible, uses that rope
  - p(rope used | M = one-armed butler is murderer) = 0.01 // not really plausible
- We can calculation of other probabilities starting with the full joint probability.

Return to Inspector Clouseau

- Suppose we know (or have very good evidence) that the knife is the murder weapon.
- Then we want to know p(B | K) and p(M | K) (using the shorthand notation)
- To compute these apply the rules of probability.
  - Using the definition of conditional probability and marginalization:

$$p(B\,|\,K) = \sum_m p(B, m\,|\,K)$$

  - Here the sum is over *m* = Maid was murder and Maid was not murder.
  - This is summing values of Maid over the joint conditional probability.
- We don't know p(B, M | K) — rewrite until in terms of known distributions:

known

$$p(B\,|\,K) = \underbrace{\sum_m p(B, m\,|\,K)}_{\text{marginalization}} = \underbrace{\sum_m \frac{p(B, m, K)}{p(K)}}_{\text{conditional prob.}} = \underbrace{\frac{\sum_m \boxed{p(K\,|\,B, m)}\, p(B, m)}{\sum_{m,b} P(K\,|\,b, m)\, \boxed{p(b, m)}}}_{\text{marginalization + factorization}}$$

What is?

- Assume butler and maid aren't colluding, i.e. they're acting independently:

$$p(B, M) = P(B)P(M)$$

- Now we have all the information necessary to evaluate the propositions.

$$p(B = \text{murderer}|\text{knife used}) = \frac{\frac{6}{10}\left(\frac{2}{10}\times\frac{1}{10}+\frac{8}{10}\times\frac{6}{10}\right)}{\frac{6}{10}\left(\frac{2}{10}\times\frac{1}{10}+\frac{8}{10}\times\frac{6}{10}\right)+\frac{4}{10}\left(\frac{2}{10}\times\frac{2}{10}+\frac{8}{10}\times\frac{3}{10}\right)} = \frac{300}{412} \approx 0.73$$