Tufts

Lecture 11:
Probability Theory, II

Artificial Intelligence (CS 131)

1

Review: Basic Axioms and Joint Probability

1. For any event a, $0 \leq P(a) \leq 1$.
2. $P(True) = 1$ and $P(False) = 0$.
3. $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$.

- For random variables, $\{X_1, X_2, \ldots, X_n\}$, the joint distribution gives the probability of each possible combination of outcomes

	StudentXGrade = A	StudentXGrade = B
StudentYGrade = A	0.72	0.08
StudentYGrade = B	0.18	0.02

2

Review: Marginal Probability

	StudentXGrade = A	StudentXGrade = B
StudentYGrade = A	0.72	0.08
StudentYGrade = B	0.18	0.02

- Given the joint distribution we can get the probability of any outcome by marginalizing: summing all values where the outcome we want is true
- For example, probability student X gets an A:

$$P(StudentXGrade = A) = (0.72 + 0.18) = 0.9$$

- Probability that *either* X or Y gets an A:

$$P(StudentXGrade = A \vee StudentYGrade = A) = (0.72 + 0.18 + 0.008) = 0.98$$

3

Marginalizing and Conditioning

- For any variable X, the probability of one outcome, x, can be found by marginalizing over all the possible event combinations for the other variables that we have:

$$P(X = x) = \sum_{\mathbf{y} \in \mathbf{Y}} P(X = x, \mathbf{y}) = \sum_{y_1 \in Y_1} \sum_{y_2 \in Y_2} \cdots \sum_{y_n \in Y_n} P(X = x, Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$$

Note: comma notation here means AND (∧)

- By product rule, this is equivalent to conditioning on Y:

$$P(X = x) = \sum_{\mathbf{y} \in \mathbf{Y}} P(X = x \,|\, \mathbf{y}) P(\mathbf{y}) = \sum_{y_1 \in Y_1} \sum_{y_2 \in Y_2} \cdots \sum_{y_n \in Y_n} P(X = x \,|\, Y_1 = y_1, \ldots, Y_n = y_n) P(Y_1 = y_1, \ldots, Y_n = y_n)$$

4

Additional Ideas: Chain Rule

- We already have the product rule:

$$P(a \wedge b) = P(a \mid b)P(b)$$

  - Remember that b here can be *any* proposition, including a *combination* of different events

- Thus, using commas to mean AND ($\wedge$), we can apply the product rule over and over, giving us the chain rule:

$$\begin{aligned} P(a, b, c, d) &= P(a \mid b, c, d)P(b, c, d) \\ &= P(a \mid b, c, d)P(b \mid c, d)P(c, d) \\ &= P(a \mid b, c, d)P(b \mid c, d)P(c \mid d)P(d) \end{aligned}$$

Additional Ideas: Conditionalized Product Rule

- We can prove a form of product rule for cases where we have conjunction of two things (a, b) and some *further evidence*, e
- By the basic definition of conditional probability:

$$P(a, b \mid \mathbf{e}) = \frac{P(a, b, \mathbf{e})}{P(\mathbf{e})}$$

- Which gives us, by the chain rule, a conditionalized product rule:

$$\begin{aligned} P(a, b \mid \mathbf{e}) &= \frac{P(a \mid b, \mathbf{e})\, P(b, \mathbf{e})}{P(\mathbf{e})} \\ &= P(a \mid b, \mathbf{e})\, \frac{P(b, \mathbf{e})}{P(\mathbf{e})} \\ &= P(a \mid b, \mathbf{e})\, P(b \mid \mathbf{e}) \end{aligned}$$

A Question to Ponder

- HIV is rare (1 million people of 300 million in US, or 0.33% have it), but it does occur
- Suppose an HIV test is *98% accurate*, namely:
  - If you have HIV, says **YES** with probability 0.98

$$P(PosTest \mid HaveHIV) = 0.98$$

  - If you do not, says **NO** with probability 0.98

$$P(\neg PosTest \mid \neg HaveHIV) = 0.98$$
$$P(PosTest \mid \neg HaveHIV) = 0.02$$

- Now, suppose that you *do* test positive: what is the probability that you actually have HIV?

$$P(HaveHIV \mid PosTest) = ?$$

The Test-Positivity Question

- If HIV test is 98% accurate, and 0.33% of people have it, what is the chance that you have it, if you test positive?
- We want to calculate the chance:

$$P(HaveHIV \mid PosTest) = ?$$

- But this is *not the same* as:

$$P(PosTest \mid HaveHIV) = 0.98$$

- Believing that these are the same leads to some of the worst errors seemingly rational human beings can make!

Bayes' Rule

- Derived by Rev. Thomas Bayes (1702–1761), and published after his death
- Allows us to take *known* conditional and prior probabilities and calculate *new* conditionals
- Central to reasoning from evidence and diagnosis in AI systems

Statement of the Rule

- We can derive Bayes' Rule from dual forms of product rule:

$$P(a \wedge b) = P(a \,|\, b)\, P(b)$$
$$P(a \wedge b) = P(b \,|\, a)\, P(a)$$

- Which means, since the left-hand sides are equal:

$$P(b \,|\, a)\, P(a) = P(a \,|\, b)\, P(b)$$

$$P(b \,|\, a) = \frac{P(a \,|\, b)\, P(b)}{P(a)}$$

- If we know how likely a is given b, and how likely a and b are by themselves, we can now calculate how likely b is given a

An Example of Using the Rule

- My cell-phone screen is broken, and I no longer know if it's on silent mode or not at any time (unless it rings)
  - I press buttons randomly to change the mode
  - With 5 possible modes, what is the probability that it is now on silent (assuming uniformity)?
- Now, my wife is supposed to call at 3:00 PM; by 4:00, no call has yet arrived…
  - How does this affect the probability?

An Exact Calculation

- I know that if the phone is on silent, I will *certainly* get no call (probability 1.0)
- I estimate a 1% chance that if the phone were *not* on silent, my wife still wouldn't call when she said she would

$$P(Off \,|\, NoCall) = \frac{P(NoCall \,|\, Off)P(Off)}{P(NoCall)}$$
$$= \frac{P(NoCall \,|\, Off)P(Off)}{(P(NoCall \,|\, Off)P(Off)) + (P(NoCall \,|\, \neg Off)P(\neg Off))}$$
$$= \frac{1.0 \times 0.2}{(1.0 \times 0.2) + (0.01 \times 0.8)} = \frac{0.2}{0.2008} \approx 0.996$$

- There is a *much higher* probability that my phone is actually on silent, conditional on the new evidence I have experienced

Generalizing the Rule

- Bayes' Rule can also be used in situations where we have more complex evidence
- The probability of some outcome a, given *both* that another outcome b has occurred, *and* that we have some other **evidence** e, is given by:

$$P(a \mid b, \mathbf{e}) = \frac{P(b \mid a, \mathbf{e})P(a \mid \mathbf{e})}{P(b \mid \mathbf{e})}$$

13

The HIV Question, Revisited

- If HIV test is 98% accurate, and 0.33% of people have it, what is the chance that you have it, if you test positive?
- We want to calculate the chance:

$$P(HaveHIV \mid PosTest) = ?$$

- But this is *not the same* as:

$$P(PosTest \mid HaveHIV) = 0.98$$

- Believing that these are the same leads to some of the worst errors seemingly rational human beings can make!

14

Chances of Testing Positive

- Given these numbers, we can condition on infection-status to calculate the likelihood of a positive HIV test, given 98% accuracy and a 0.33% infection rate:

$$\begin{aligned} P(PosTest) =& (P(PosTest \mid HaveHIV) \times P(HaveHIV)) + \\ & (P(PosTest \mid \neg HaveHIV) \times P(\neg HaveHIV)) \\ \\ =& (0.98 \times 0.0033) + (0.02 \times 0.9967) \\ =& (0.003234 + 0.019934) = 0.023168 \end{aligned}$$

- Test is positive 2.3% of the time, including both correct diagnoses (true positives) and incorrect ones (false positives)

15

Using Bayes' Theorem

$$\begin{aligned} P(HaveHIV \mid PosTest) &= \frac{P(PosTest \mid HaveHIV)P(HaveHIV)}{P(PosTest)} \\ &= \frac{0.98 \times 0.0033}{(0.98 \times 0.0033) + (0.02 \times 0.9967)} \\ &= \frac{0.003234}{0.003234 + 0.019934} = \frac{0.003234}{0.023168} \approx 0.1396 \end{aligned}$$

- The chance of actually having HIV is only about 14%!

16

Probability & Guilt

- In 1999, British mother Sally Clark was convicted of murder, after her two infant sons both died of what was originally thought to be Sudden Infant Death Syndrome (SIDS)
- Her conviction relied upon the expert testimony of famous pediatrician Sir Roy Meadow, who argued:
  1. The chance a random child dies of SIDS is about 1 in 3000
  2. Amongst non-smoking, older parents (like Clark), with at least one wage, this rises to about 1 in 8500
  3. The chance of both of Clark's sons dying of SIDS is therefore about $(1/8500 \times 1/8500)$, or 1 in 73 million
  4. The chance that Sally Clark is innocent is thus 1 in 73 million
- *What should we, the jury, make of this argument?*

A Problem of Numbers

- When we roll two dice, the chance of them both coming up 5 is $(1/6 \text{ x } 1/6) = 1/36$
  - Since dice are independent of one another, the roll of one does not affect the other, so we simply **multiply** probabilities
  - We will look at this in more depth in upcoming lectures
- If *one* infant in a family dies of SIDS, is the chance of a *second* dying exactly the same? Lower? Higher?
  - Calculations in UK show that chances of a randomly selected *pair* of infants dying of SIDS actually 1 in $100{,}000$ (0.0001)
  - If about $500{,}000$ families in UK have two infants in any given year, how many times would this happen?

A Problem of Reasoning

- Even if we accept the 1 in $100{,}000$ value (instead of Meadow's 1 in 73 million), is *that* the probability that Clark is actually innocent?
- ***No***. This is the Prosecutor's Fallacy: the claim that the chance of innocence equals the chance of a rare event
  - Rare events *do happen* (e.g., lotteries)
  - We then have to calculate the conditional probability of guilt, *given* that a tragic event did in fact happen

The Real Chance of Innocence

- Let M = "The children were murdered"
- Let D = "The children died"
- We want $P(M\,|\,D)$, which is, by Bayes' Rule:

$$P(M \mid D) = \frac{P(D \mid M)\, P(M)}{P(D)}$$

- We can get the denominator, $P(D)$, by normalizing:

$$P(M \mid D) = \frac{P(D \mid M)\, P(M)}{P(D \mid M)\, P(M) + P(D \mid \neg M)\, P(\neg M)}$$

- Now, since $P(D\,|\,M) = 1$, and $P(\neg M) = 1 - P(M)$, we have:

$$P(M \mid D) = \frac{P(M)}{P(M) + P(D \mid \neg M)\,(1 - P(M))}$$

The Real Chance of Innocence

- Now, unless we think that the probability of a pair of children being murdered is *very high*, the probability of the mother's innocence is high, too
- An estimate of the number of pairs of children who die of all causes other than murder in the UK in a year is $1 \text{ in } 20{,}000$: $P(D \mid \neg M) = 0.00005$
- Based on this, even if we concede an *extreme over-estimate* to the prosecution, the probability of murder in this case is still low
- Suppose that murder happens with frequency $1/10$ of all other causes combined, so that $1 \text{ in } 200{,}000$ pairs of infants will be murdered in the UK in a year (i.e., $P(M) = 0.000005$); this gives us:

$$P(M \mid D) = \frac{P(M)}{P(M) + P(D \mid \neg M)\,(1 - P(M))}$$

$$= \frac{0.000005}{0.000005 + 0.00005 \times 0.999995}$$

$$\approx 0.091$$

The Real Chance of Innocence

- Given the real probability of murder $P(M \mid D) = 0.091$, the probability of innocence is:

$$P(\neg M \mid D) = 1 - P(M \mid D) = 0.909$$

- Thus, even given *very generous* assumptions in the favor of the prosecution, Sally Clark would have been innocent over 90% of the time
  - Due to fallacious reasoning at her trial, she spent over 3 years in jail for a crime she almost certainly did not commit!