CS 331: Artificial Intelligence Fundamentals of Probability III

Thanks to Andrew Moore for some course material

Full Joint Probability Distributions

The probabilities

Coin	Card	Candy	P(Coin, Card, Candy)
tails	black	1	0.15
tails	black	2	0.06
tails	black	3	0.09
tails	red	1	0.02
tails	red	2	0.06
tails	red	3	0.12
heads	black	1	0.075
heads	black	2	0.03
heads	black	3	0.045
heads	red	1	0.035
heads	red	2	0.105
heads	red	3	0.21

This cell means P(Coin=heads, Card=red, Candy=3) = 0.21

Marginalization

The general marginalization rule for any sets of variables Y and Z:

$$P(Y) = \sum_{\mathbf{z}} P(Y, \mathbf{z})$$

z is over all possible combinations of values of Z (remember **Z** is a set)

$$P(Y) = \sum_{\mathbf{z}} P(Y \mid \mathbf{z}) P(\mathbf{z})$$

Conditional Probabilities

We can also compute conditional probabilities from the joint. Recall:

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

Inference

We will write the query as $P(X \mid e)$

$$P(X | e) = \alpha P(X, e) = \alpha \sum P(X, e, y)$$

Summation is over all possible combinations of values of the unobserved variables Y

X = Query variable (a single variable for now)

E = Set of evidence variables

e = the set of observed values for the evidence variables

Y = Unobserved variables

Independence

We say that variables X and Y are independent if any of the following hold: (note that they are all equivalent)

$$P(X | Y) = P(X)$$
 or

$$P(Y \mid X) = P(Y)$$
 or

$$P(X,Y) = P(X)P(Y)$$

Bayes' Rule

The product rule can be written in two ways:

$$P(A, B) = P(A \mid B)P(B)$$

$$P(A, B) = P(B \mid A)P(A)$$

You can combine the equations above to get:

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

7

Bayes' Rule



More generally, the following is known as Bayes' Rule:

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

Note that these are distributions

Sometimes, you can treat P(B) as a normalization constant α

$$P(A \mid B) = \alpha P(B \mid A) P(A)$$

More General Forms of Bayes Rule

If A takes 2 values:

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

If A takes n_A values:

$$P(A = v_i | B) = \frac{P(B | A = v_i)P(A = v_i)}{\sum_{k=1}^{n_A} P(B | A = v_k)P(A = v_k)}$$

9

When is Bayes Rule Useful?

Sometimes it's easier to get P(X|Y) than P(Y|X).

Information is typically available in the form $P(\text{effect} \mid \text{cause})$ rather than $P(\text{ cause} \mid \text{effect})$

For example, P(symptom | disease) is easy to measure empirically but obtaining P(disease | symptom) is harder

10

Bayes Rule Example

Meningitis causes stiff necks with probability 0.5. The prior probability of having meningitis is 0.00002. The prior probability of having a stiff neck is 0.05. What is the probability of having meningitis given that you have a stiff neck?

Let M =patient has meningitis

Let S =patient has stiff neck

 $P(s \mid m) = 0.5$

P(m) = 0.00002

P(s) = 0.05

$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{(0.5)(0.00002)}{0.05} = 0.0002$$

Bayes Rule Example

Meningitis causes stiff necks with probability 0.5. The prior probability of having meningitis is 0.00002. The prior probability of having a stiff neck is 0.05. What is the probability of having meningitis given that you have a stiff neck?

Let M =patient has meningitis

Let S =patient has stiff neck

 $P(s \mid m) = 0.5$ P(m) = 0.00002

D() 0.05

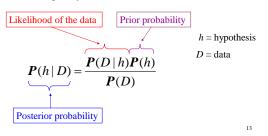
Note: Even though P(s|m) = 0.5, P(m|s) = 0.0002

 $P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{(0.5)(0.00002)}{0.05} = 0.0002$

2

How is Bayes Rule Used

In machine learning, we use Bayes rule in the following way:



Bayes Rule With More Than One Piece of Evidence

Suppose you now have 2 evidence variables Card=red and Candy=1 (note that Coin is uninstantiated below)

 $P(Coin \mid Card = red, Candy = 1)$ = $\alpha P(Card = red, Candy = 1 \mid Coin) P(Coin)$

In order to calculate $P(Card=red, Candy=1 \mid Coin)$, you need a table of 6 probability values.

With N Boolean evidence variables, you need 2^N probability values.

14

Why is independence useful?

P(Card, Candy) = P(Card)P(Candy)This table has 2 values
This table has 3 values

- You now need to store 5 values to calculate *P*(*Coin*, *Card*, *Candy*)
- Without independence, we needed 6

15

Conditional Independence

Suppose I tell you that to select a piece of *Candy*, I first flip a *Coin*. If heads, I select a *Card* from one (stacked) deck; if tails, I select from a different (stacked) deck. The color of the card determines the bag I select the *Candy* from, and each bag has a different mix of the types of *Candy*.

Are Coin and Candy independent?

16

Conditional Independence

Suppose I tell you that to select a piece of *Candy*, I first flip a *Coin*. If heads, I select a *Card* from one deck; if tails, I select from a different deck. The color of the card determines the bag I select the *Candy* from, and each bag has a different mix of the types of *Candy*.

Are Coin and Candy independent? No.

But given Card, they are independent!

 $P(Coin = heads, Candy = 3 | Card) = P(Coin = heads | Card) \times P(Candy = 3 | Card)$

17

Conditional Independence

General form:

P(A,B|C) = P(A|C)P(B|C)

Or equivalently:

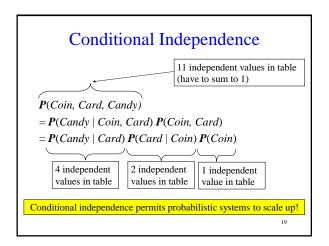
 $P(A \mid B, C) = P(A \mid C)$ and

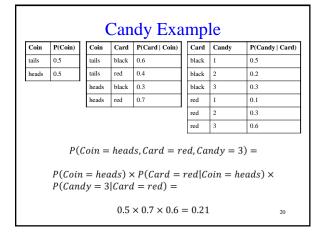
 $P(B \mid A, C) = P(B \mid C)$

How to think about conditional independence:

In $P(A \mid B, C) = P(A \mid C)$: if knowing C tells me everything about A, I don't gain anything by knowing B

18





CW: Practice Card | P(Card | Coin) Card Candy Coin P(Candy | Card) Coin P(Coin) tails 0.5 tails black 0.6 black 0.5 heads 0.5 tails 0.4 black 0.2 red black 0.3 black 0.3 heads heads red 0.7 0.1 red 0.3 red 0.6 Compute $P(Coin = tails \mid Card = red)$ 21

What You Should Know

- How to do inference in joint probability distributions
- How to use Bayes Rule
- Why independence and conditional independence is useful

22