CS 331: Artificial Intelligence
Fundamentals of Probability III

Thanks to Andrew Moore for some course material

1

Full Joint Probability Distributions

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

The probabilities in the last column sum to 1

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

2

Marginalization

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

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

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

or

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

3

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 $\mathbf{P}(X \mid \boldsymbol{e})$

$$\mathbf{P}(X \mid \boldsymbol{e}) = \alpha \mathbf{P}(X, \boldsymbol{e}) = \alpha \sum_{\boldsymbol{y}} \mathbf{P}(X, \boldsymbol{e}, \boldsymbol{y})$$

Summation is over all possible combinations of the 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

Bayes' Rule

The product rule can be written in two ways:

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

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

You can combine the equations above to get:

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

6

Bayes' Rule

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

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

Note that these are distributions

Sometimes, you can treat $\boldsymbol{P}$(B) as a normalization constant α

$$\boldsymbol{P}(A \mid B) = \alpha \boldsymbol{P}(B \mid A)\boldsymbol{P}(A)$$

More General Forms of Bayes Rule

If A takes 2 values:

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

If A takes n_A values:

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

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(effect | cause) rather than P(cause | effect)

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

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* | *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* | *m*) = 0.5
P(*m*) = 0.00002
P(*s*) = 0.05

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

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

How is Bayes Rule Used

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

$$\underbrace{\boldsymbol{P}(h \mid D)}_{\text{Posterior probability}} = \frac{\overbrace{\boldsymbol{P}(D \mid h)}^{\text{Likelihood of the data}}\overbrace{\boldsymbol{P}(h)}^{\text{Prior probability}}}{\boldsymbol{P}(D)}$$

h = hypothesis
D = data

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* | *Card=red, Candy=1*)
= α ***P***(*Card=red, Candy=1* | *Coin*) ***P***(*Coin*)

In order to calculate ***P***(*Card=red*, *Candy=1* | *Coin*), you need a table of 6 probability values. With N Boolean evidence variables, you need 2^N probability values.

Independence

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

$$\boldsymbol{P}(X \mid Y) = \boldsymbol{P}(X) \quad \text{or}$$

$$\boldsymbol{P}(Y \mid X) = \boldsymbol{P}(Y) \quad \text{or}$$

$$\boldsymbol{P}(X, Y) = \boldsymbol{P}(X)\boldsymbol{P}(Y)$$

Why is independence useful?

$$P(Card, Candy) = \underbrace{P(Card)}_{\text{This table has 2 values}}\underbrace{P(Candy)}_{\text{This table has 3 values}}$$

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

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?

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!

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

Conditional Independence

General form:

$$\boldsymbol{P}(A, B \mid C) = \boldsymbol{P}(A \mid C)\boldsymbol{P}(B \mid C)$$

Or equivalently:

$$\boldsymbol{P}(A \mid B, C) = \boldsymbol{P}(A \mid C) \quad \text{and}$$

$$\boldsymbol{P}(B \mid A, C) = \boldsymbol{P}(B \mid C)$$

How to think about conditional independence:

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

Conditional Independence

$$\overbrace{\boldsymbol{P}(Coin, Card, Candy)}^{\text{11 independent values in table (have to sum to 1)}}$$

$= \boldsymbol{P}(Candy \mid Coin, Card)\ \boldsymbol{P}(Coin, Card)$

$$= \underbrace{\boldsymbol{P}(Candy \mid Card)}_{\text{4 independent values in table}}\ \underbrace{\boldsymbol{P}(Card \mid Coin)}_{\text{2 independent values in table}}\ \underbrace{\boldsymbol{P}(Coin)}_{\text{1 independent value in table}}$$

Conditional independence permits probabilistic systems to scale up!

Candy Example

Coin	P(Coin)
tails	0.5
heads	0.5

Coin	Card	P(Card | Coin)
tails	black	0.6
tails	red	0.4
heads	black	0.3
heads	red	0.7

Card	Candy	P(Candy | Card)
black	1	0.5
black	2	0.2
black	3	0.3
red	1	0.1
red	2	0.3
red	3	0.6

$P(Coin = heads, Card = red, Candy = 3) =$

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

$0.5 \times 0.7 \times 0.6 = 0.21$

Practice

Coin	P(Coin)
tails	0.5
heads	0.5

Coin	Card	P(Card | Coin)
tails	black	0.6
tails	red	0.4
heads	black	0.3
heads	red	0.7

Card	Candy	P(Candy | Card)
black	1	0.5
black	2	0.2
black	3	0.3
red	1	0.1
red	2	0.3
red	3	0.6

Compute $P(Coin = tails \mid Card = red)$

What You Should Know

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