Tufts UNIVERSITY

PROBABILITY THEORY 2

ARTIFICIAL INTELLIGENCE | COMP 131

TODAY ON AI

- Conditional queries
- Joint queries
- Inference with Bayes' rule
- Questions?

Probabilistic inference computes a desired probability from other known probabilities.

- We generally compute conditional probabilities that represent some **belief** given the evidence:

$$P(on\_time \mid no\_accidents) = 0.90$$

- Probabilities is **updated** with new evidence:

$$P(on\_time \mid no\_accidents, \mathbf{5}AM) = 0.95$$
$$P(on\_time \mid no\_accidents, 5AM, \boldsymbol{raining}) = 0.80$$

SECTION 01

Conditional queries

Given the joint distribution, we have a **conditional query**: $P(Q_1, \dots, Q_l | e_1, \dots, e_k)$

- **Evidence variables**: $E_1 = e_1, \dots, E_k = e_k$
- **Query variable(s)**: $Q_1, \dots, Q_l$
- **Hidden variables**: $H_1, \dots, H_r$

All the variables of the model $X_1, \dots, X_n$

- Step 1: Select the entries in the joint distribution consistent with the evidence

- Step 2: Sum out H to get joint of query and evidence

$$P(Q_1, \dots, Q_l, e_1, \dots, e_k) = \sum_{h_1,\dots,h_r} P(Q_1, \dots, Q_l, h_1, \dots, h_r, e_1, \dots, e_k)$$

- Step 3: Normalize

$$Z = \sum_{Q_1,\dots,Q_l} P(Q_1, \dots, Q_l, e_1, \dots, e_k) \qquad P(Q_1, \dots, Q_l | e_1, \dots, e_k) = \frac{1}{\mathrm{Z}} \mathrm{P}(\mathrm{Q}_1, \dots, \mathrm{Q}_l, e_1, \dots, e_k)$$

PROBLEM

Worst-case time complexity $O(d^n)$

Space complexity $O(d^n)$ to store the joint distribution

$P(Season, Temperature, Weather)$

S	**T**	**W**	**$P(S,T,W)$**
$summer$	hot	sun	0.30
$summer$	hot	$rain$	0.05
$summer$	$cold$	sun	0.10
$summer$	$cold$	$rain$	0.05
$winter$	hot	sun	0.10
$winter$	hot	$rain$	0.05
$winter$	$cold$	sun	0.15
$winter$	$cold$	$rain$	0.20

$P(W)?$

$Q = W$
$H = S, T$
$E = \emptyset$

W	**P**
sun	0.65 / 1.00
$rain$	0.35 / 1.00

W	**P**
sun	0.65
$rain$	0.35

$P(W \mid winter)?$

$Q = W$
$H = T$
$E = S$

W	**P**
sun	0.25 / 0.50
$rain$	0.25 / 0.50

W	**P**
sun	0.50
$rain$	0.50

$P(W \mid winter, hot)?$

$Q = W$
$H = \emptyset$
$E = S, T$

W	**P**
sun	0.10 / 0.15
$rain$	0.05 / 0.15

W	**P**
sun	0.67
$rain$	0.33

Conditional queries

EXAMPLE

SECTION 02

Joint queries

Given a conditional distribution and the marginal distributions, we have a **joint query**: $P(Q_1, \dots, Q_l, e_1, \dots, e_k)$

- **Evidence variables**: $E_1 = e_1, \dots, E_k = e_k$
- **Query variable(s)**: $Q_1, \dots, Q_l$
- **Hidden variables**: $H_1, \dots, H_r$

All the variables of the model $X_1, \dots, X_n$

- Step 1: Select the entries in the conditional distribution consistent with the evidence

- Step 2: Multiply them for the corresponding marginal distribution

$$P(x|y) = \frac{P(x,y)}{P(y)} \qquad P(x,y) = P(x|y)P(y)$$

- $P(Dryness, Weather)$?

$P(Weather)$

W	P
sun	0.8
$rain$	0.2

$\forall w \in W: P(w)$

$P(Dryness|Weather)$

W	D	$P(D\|W)$
sun	wet	0.10
sun	dry	0.90
$rain$	wet	0.70
$rain$	dry	0.30

$\forall d \in D: P(d, w) = P(d|w)P(w)$

$P(Dryness, Weather)$

W	D	$P(D, W)$
sun	wet	0.08
sun	dry	0.72
$rain$	wet	0.14
$rain$	dry	0.06

The **chain rule** allows you to write any joint distribution as an incremental product of conditional distributions:

$$P(x_1, x_2) = P(x_1)P(x_2|x_1)$$

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

$$P(x_1, \dots, x_n) = P(x_1)\prod_{i=2}^{n} P(x_i|x_1, \dots, x_{i-1})$$

$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i|x_1, \dots, x_{i-1})$$

SECTION 03

Inference with Bayes' rule

In order to measure the performance of a classification test, we can use the two metrics **sensitivity** and **specificity**:

- **Sensitivity** measures the percentage of actual positives that are correctly identified by the test.
- **Specificity** measures the percentage of actual negatives that are correctly identified by the test.

The Bayes' rule can be used to calculate **cause-effect** probabilities:

$$P(cause|effect) = \frac{P(effect|cause)}{P(effect)} P(cause)$$

An example that many doctors get wrong:

1. **100 out of 10,000** women at age forty who have a routine screening have breast cancer
2. **80 of every 100** women with breast cancer will be positive to mammography (sensitivity)
3. **950 out of 9,900** women without breast cancer will also be positive (inverse of specificity)

EFFECT CAUSE

How many women with **positive mammography** will have **breast cancer**? $\boldsymbol{P(BreastCancer|Positive)}$?

$P(BreastCancer) = \mathbf{0.01}$ $P(Positive|BreastCancer) = \mathbf{0.80}$ $P(Positive|\neg BreastCancer) = \mathbf{0.096}$

$$P(Positive) = P(Positive \mid BreastCancer)P(BreastCancer) + P(Positive \mid \neg BreastCancer)\, P(\neg BreastCancer)$$

$$P(Positive) = 0.80 \times 0.01 + 0.096 \times 0.99 = 0.008 + 0.095 = \mathbf{0.103}$$

$$P(BreastCancer|Positive) = \frac{P(Positive|BreastCancer)}{P(Positive)} P(BreastCancer)$$

$$\boldsymbol{P(BreastCancer|Positive)} = \mathbf{0.078}$$

- If we know $P(effect|cause)$ for **every cause**, we can avoid having to know $P(effect)$:

$$P(cause \mid effect) = \frac{P(effect \mid cause)\, P(cause)}{P(effect)} = \frac{P(effect \mid cause)\, P(cause)}{\sum_{\forall c \in Causes} P(effect \mid c)\, P(c)}$$

- But sometimes it's harder to find out $P(effect \mid cause)$ for all causes independently than it is simply to find out $P(effect)$.

- Note that Bayes' rule relies on the fact the effect must have arisen because of **one** of the hypothesised causes. We **cannot** reason directly about causes we have not imagined.

- Suppose we have several pieces of evidence we want to combine:

I have toothache and the dental probe catches on my tooth

- How do we do this?

$$P(cavity|toothache, catch) = \alpha\, P(toothache, catch|cavity)P(cavity)$$

- As we have more effects our causal model becomes very complicated (for N binary effects there will be 2^N different combinations of evidence that we need to model given a cause):

$$P(toothache, catch|cavity)\ ,\quad P(toothache, \neg catch|cavity)\quad \cdots$$

- In many practical applications there are not a few evidence variables but hundreds. Therefore, 2^N can be very big.

- **Conditional independence** helps as toothache and catch **are not independent**. However, they **are independent** given the presence or absence of a cavity.

- In other words, we can use the knowledge that cavities cause toothache and they cause the catch, but the catch and the toothache do not cause each other. Rather, they have a single common cause.

PRACTICE

Exercises from the textbook:

any exercise of chapter 13

READINGS

Chapters 13

QUESTIONS ?

ARTIFICIAL INTELLIGENCE

COMP 131

FABRIZIO SANTINI

VERSION 3.1