

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:

$$P(\textit{on_time} \mid \textit{no_accidents}) = 0.90$$

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

$$P(\textit{on_time} \mid \textit{no_accidents}, \textit{5AM}) = 0.95$$

$$P(\textit{on_time} \mid \textit{no_accidents}, \textit{5AM}, \textit{raining}) = 0.80$$

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

PROBLEM

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

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

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

- **Step 3**: Normalize

$$Z = \sum_{Q_1, \dots, Q_l} P(Q_1, \dots, Q_l, e_1, \dots, e_k) \quad P(Q_1, \dots, Q_l | e_1, \dots, e_k) = \frac{1}{Z} P(Q_1, \dots, Q_l, e_1, \dots, e_k)$$

$P(\text{Season, Temperature, Weather})$

<i>S</i>	<i>T</i>	<i>W</i>	<i>P(S,T,W)</i>
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

$P(W)?$

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

<i>W</i>	<i>P</i>
sun	0.65 / 1.00
rain	0.35 / 1.00

<i>W</i>	<i>P</i>
sun	0.65
rain	0.35

$P(W \mid \text{winter})?$

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

<i>W</i>	<i>P</i>
sun	0.25 / 0.50
rain	0.25 / 0.50

<i>W</i>	<i>P</i>
sun	0.50
rain	0.50

$P(W \mid \text{winter, hot})?$

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

<i>W</i>	<i>P</i>
sun	0.10 / 0.15
rain	0.05 / 0.15

<i>W</i>	<i>P</i>
sun	0.67
rain	0.33

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

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

PROBLEM

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

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

■ $P(\text{Dryness}, \text{Weather})?$

$P(\text{Weather})$		
W	P	
sun	0.8	↓ $\forall w \in W: P(w)$
rain	0.2	

$P(\text{Dryness} \text{Weather})$			
W	D	$P(D W)$	
sun	wet	0.10	→ $\forall d \in D: P(d, w) = P(d w)P(w)$
sun	dry	0.90	
rain	wet	0.70	
rain	dry	0.30	

$P(\text{Dryness}, \text{Weather})$		
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:

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

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

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

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

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:

$$P(\text{cause} \mid \text{observation}) = \frac{P(\text{observation} \mid \text{cause})}{P(\text{observation})} P(\text{cause})$$

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)

OBSERVATION

CAUSE

How many women with **positive mammography** will have **breast cancer**?

$P(\text{BreastCancer} \mid \text{Positive})?$

100 out of 10,000 women at age forty who have a routine screening have breast cancer

$$P(\text{BreastCancer}) = 0.01$$

80 of every 100 women with breast cancer will be positive to mammography (sensitivity)

$$P(\text{Positive} \mid \text{BreastCancer}) = 0.80$$

950 out of 9,900 women without breast cancer will also be positive (inverse of specificity)

$$P(\text{Positive} \mid \neg \text{BreastCancer}) = 0.096$$

$P(\text{BreastCancer} \mid \text{Positive})?$

$$P(\text{BreastCancer}) = \mathbf{0.01} \quad P(\text{Positive} \mid \text{BreastCancer}) = \mathbf{0.80} \quad P(\text{Positive} \mid \neg \text{BreastCancer}) = \mathbf{0.096}$$

$$P(\text{Positive}) = P(\text{Positive} \mid \text{BreastCancer})P(\text{BreastCancer}) + P(\text{Positive} \mid \neg \text{BreastCancer}) P(\neg \text{BreastCancer})$$

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

$$P(\text{BreastCancer} \mid \text{Positive}) = \frac{P(\text{Positive} \mid \text{BreastCancer})}{P(\text{Positive})} P(\text{BreastCancer})$$

$$P(\text{BreastCancer} \mid \text{Positive}) = \mathbf{0.078}$$

- 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?

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

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

$$P(\text{toothache}, \text{catch}|\text{cavity}) , \quad P(\text{toothache}, \neg \text{catch}|\text{cavity}) \quad \dots$$

- In many practical applications there can be hundreds evidence variables. Therefore, 2^N can be very big in the simple case of binary domains.
- **Conditional independence** helps as toothache and catch **are not independent**. However, they **are independent** given the presence or absence of a cavity.
- We can use the knowledge of the variables to simplify the relationships:

Cavities cause toothache, and they cause the catch, but the catch and the toothache do not cause each other

PRACTICE

Exercises from the textbook:
any exercise of chapter 13

QUESTIONS ?

ARTIFICIAL INTELLIGENCE

COMP 131

FABRIZIO SANTINI