

Uncertainty and Belief Networks

Introduction to Artificial Intelligence

CS 151

Lecture 1 continued

Ok, Lecture 2!

This lecture

- Conditional Independence
- Bayesian (Belief) Networks: Syntax and semantics

Reading

– Chapter 14.1-14.2

Propositions and Random Variables

- Letting **A** refer to a proposition which may either be true or false, **P(A)** denotes the probability that **A** is true – also denoted **P(A=TRUE)**.
- **A** is called a binary or propositional random variable.
- We also have multi-valued random variables
 - e.g., **Weather** is one of **<sunny,rain,cloudy,snow>**

Priors, Distribution, Joint

Prior or **unconditional probabilities** of propositions

e.g., $P(\text{Cavity}) = P(\text{Cavity}=\text{TRUE}) = 0.1$

$P(\text{Weather}=\text{sunny}) = 0.72$

correspond to belief prior to arrival of any (new) evidence

Probability distribution gives probabilities of all possible values of the random variable.

Weather is one of $\langle \text{sunny}, \text{rain}, \text{cloudy}, \text{snow} \rangle$

$P(\text{Weather}) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$

(normalized, i.e., sums to 1)

Joint probability distribution

Joint probability distribution for a set of variables gives values for each possible assignment to all the variables

$P(\text{Toothache}, \text{Cavity})$ is a 2 by 2 matrix.

	Toothache=true	Toothache = false
Cavity=true	0.04	0.06
Cavity=false	0.01	0.89

NOTE: Elements in table sum to 1 $\rightarrow$ 3 independent numbers.

Weather=	sunny	rain	cloudy	snow
Cavity=True				
Cavity=False				

Conditional Probabilities

- **Conditional** or posterior probabilities

e.g., $P(\text{Cavity} \mid \text{Toothache}) = 0.8$

What is the probability of having a cavity given that the patient has a toothache?

- Definition of **conditional probability**:

$$P(A|B) = P(A, B) / P(B) \quad \text{if } P(B) \neq 0$$

- **Product rule** gives an alternative formulation:

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

Bayes Rule

From product rule $P(A, B) = P(A|B)P(B) = P(B|A)P(A)$, we can obtain Bayes' rule

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

Bayes Rule: Example

$$P(\textit{Cause} \mid \textit{Effect}) = \frac{P(\textit{Effect} \mid \textit{Cause})P(\textit{Cause})}{P(\textit{Effect})}$$

Let M be meningitis, S be stiff neck

$$P(M)=0.0001$$

$$P(S)= 0.1$$

$$P(S|M) = 0.8$$

$$P(M \mid S) = \frac{P(S \mid M)P(M)}{P(S)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

Note: posterior probability of meningitis still very small!

A generalized Bayes Rule

- More general version conditionalized on some background evidence E

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

Using the full joint distribution

	Toothache=true	Toothache = false
Cavity=true	0.04	0.06
Cavity=false	0.01	0.89

What is the unconditional probability of having a Cavity?

$$\begin{aligned} P(\text{Cavity}) &= P(\text{Cavity} \wedge \text{Toothache}) + P(\text{Cavity} \wedge \sim \text{Toothache}) \\ &= 0.04 + 0.06 = 0.1 \end{aligned}$$

What is the probability of having either a cavity or a Toothache?

$$\begin{aligned} P(\text{Cavity} \vee \text{Toothache}) &= P(\text{Cavity}, \text{Toothache}) + P(\text{Cavity}, \sim \text{Toothache}) + P(\sim \text{Cavity}, \text{Toothache}) \\ &= 0.04 + 0.06 + 0.01 = 0.11 \end{aligned}$$

Using the full joint distribution

	Toothache=true	Toothache = false
Cavity=true	0.04	0.06
Cavity=false	0.01	0.89

What is the probability of having a cavity given that you already have a toothache?

$$P(\text{Cavity} \mid \text{Toothache}) = \frac{P(\text{Cavity} \wedge \text{Toothache})}{P(\text{Toothache})} = \frac{0.04}{0.04 + 0.01} = 0.8$$

Normalization

Suppose we wish to compute a posterior distribution over random variable A given $B=b$, and suppose A has possible values $a_1 \dots a_m$

We can apply Bayes' rule for each value of A :

$$P(A=a_1|B=b) = P(B=b|A=a_1)P(A=a_1)/P(B=b)$$

...

$$P(A=a_m | B=b) = P(B=b|A=a_m)P(A=a_m)/P(B=b)$$

Adding these up, and noting that $\sum_i P(A=a_i | B=b) = 1$:

$$P(B=b) = \sum_i P(B=b|A=a_i)P(A=a_i)$$

This is the normalization factor denoted $\alpha = 1/P(B=b)$:

$$P(A | B=b) = \alpha P(B=b | A)P(A)$$

Typically, we compute an unnormalized distribution, and then normalize at end

e.g., suppose $P(B=b | A)P(A) = \langle 0.4, 0.2, 0.2 \rangle$

then $P(A|B=b) = \alpha \langle 0.4, 0.2, 0.2 \rangle$

$$= \langle 0.4, 0.2, 0.2 \rangle / (0.4+0.2+0.2) = \langle 0.5, 0.25, 0.25 \rangle$$

Marginalization

Given a condition distribution $P(X | Z)$, we can create the unconditional distribution $P(X)$ by *marginalization*:

$$P(X) = \sum_z P(X, Z=z)$$

In general, given a joint distribution over a set of variables, the distribution over any subset (called a **marginal distribution** for historical reasons) can be calculated by summing out the other variables.

Conditioning

A variant of marginalization using conditional probabilities:

$$P(X) = \sum_z P(X \mid Z=z) P(Z=z)$$

Or, more generally:

$$P(X|Y) = \sum_z P(X \mid Y, Z=z) P(Z=z \mid Y)$$

Intuition: it is often easier to assess more *specific circumstances*, e.g.,

$$\begin{aligned} &P(\text{RunOver} \mid \text{Cross}) \\ &= P(\text{RunOver} \mid \text{Cross}, \text{Light}=\text{green})P(\text{Light}=\text{green} \mid \text{Cross}) \\ &\quad + P(\text{RunOver} \mid \text{Cross}, \text{Light}=\text{yellow})P(\text{Light}=\text{yellow} \mid \text{Cross}) \\ &\quad + P(\text{RunOver} \mid \text{Cross}, \text{Light}=\text{red})P(\text{Light}=\text{red} \mid \text{Cross}) \end{aligned}$$

Absolute Independence

- Two random variables A and B are (absolutely) **independent** iff
 $P(A, B) = P(A)P(B)$

- Using product rule for A & B independent, we can show:

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

$$\text{Therefore } P(A | B) = P(A)$$

- If n Boolean variables are independent, the full joint is:

$$P(X_1, \dots, X_n) = \prod_i P(X_i)$$

Full joint is generally specified by $2^n - 1$ numbers, but when independent only n numbers are needed.

- Absolute independence is a very strong requirement, seldom met!!

Conditional Independence

- Some evidence may be irrelevant, allowing simplification, e.g.,

$$P(\text{Cavity} \mid \text{Toothache}, \text{CubsWin}) = P(\text{Cavity} \mid \text{Toothache}) = 0.8$$

- This property is known as **Conditional Independence** and can be expressed as:

$$P(X \mid Y, Z) = P(X \mid Z)$$

which says that X and Y independent given Z.

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

1. $P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity})$

i.e., Catch is conditionally independent of Toothache given Cavity

The same independence holds if I haven't got a cavity:

2. $P(\text{Catch} \mid \text{Toothache}, \sim\text{Cavity}) = P(\text{Catch} \mid \sim\text{Cavity})$

Conditional independence contd.

Equivalent statements to

$$P(\text{Catch} \mid \text{Toothache}) = P(\text{Catch}) \quad (*)$$

$$1.a \quad P(\text{Toothache} \mid \text{Catch}) = P(\text{Toothache}) \quad (1a)$$

$$\begin{aligned} P(\text{Toothache} \mid \text{Catch}) &= P(\text{Catch} \mid \text{Toothache}) P(\text{Toothache}) / P(\text{Catch}) \\ &= P(\text{Catch}) P(\text{Toothache}) / P(\text{Catch}) \quad (\text{from } *) \\ &= P(\text{Toothache}) \end{aligned}$$

$$1.b \quad P(\text{Toothache}, \text{Catch}) = P(\text{Toothache})P(\text{Catch})$$

$$\begin{aligned} P(\text{Toothache}, \text{Catch}) &= P(\text{Toothache} \mid \text{Catch}) P(\text{Catch}) \quad (\text{product rule}) \\ &= P(\text{Toothache}) P(\text{Catch}) \quad (\text{from } 1a) \end{aligned}$$

Conditional independence contd.

Equivalent statements to

$$P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity}) \quad (*)$$

$$1.a \quad P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \quad (1a)$$

$$\begin{aligned} P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) \\ &= P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) P(\text{Toothache} \mid \text{Cavity}) / P(\text{Catch} \mid \text{Cavity}) \\ &= P(\text{Catch} \mid \text{Cavity}) P(\text{Toothache} \mid \text{Cavity}) / P(\text{Catch} \mid \text{Cavity}) \quad (\text{from } *) \\ &= P(\text{Toothache} \mid \text{Cavity}) \end{aligned}$$

$$1.b \quad P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity})$$

(Toothache and catch are conditionally independent, given a cavity)

$$\begin{aligned} P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) \\ &= P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) P(\text{Catch} \mid \text{Cavity}) \quad (\text{product rule}) \\ &= P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity}) \quad (\text{from } 1a) \end{aligned}$$

Using Conditional Independence

Suppose the dentist wants to know whether there is a cavity or not, given that the patient has a toothache and the dentist's probe catches on his tooth:

$$\begin{aligned} P(\text{Cavity} | \text{Toothache}, \text{Catch}) &= (\text{using full joint}) \\ &= \alpha < 0.108, 0.016 > \cong < 0.871, 0.129 > \end{aligned}$$

Try Bayes' Rule:

$$= \alpha P(\text{Toothache}, \text{Catch} | \text{Cavity}) P(\text{Cavity})$$

But we would still not scale up if we need to know the left hand term for many variables. However, while **Toothache** and **Catch** are not independent, they *are* independent *given* **Cavity**:

$$= \alpha P(\text{Toothache} | \text{Cavity}) P(\text{Catch} | \text{Cavity}) P(\text{Cavity}) \text{ (from 1.b)}$$

Using Conditional Independence

Full joint distribution can now be written as

$$\begin{aligned} &P(\text{Toothache}, \text{Catch}, \text{Cavity}) \\ &= P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) P(\text{Cavity}) \\ &= P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity}) P(\text{Cavity}) \quad (\text{from 1.b}) \end{aligned}$$

Specified by: $2 + 2 + 1 = 5$ independent numbers

Compared to 7 for general joint

or 3 for unconditionally independent.

Belief (Bayes) networks

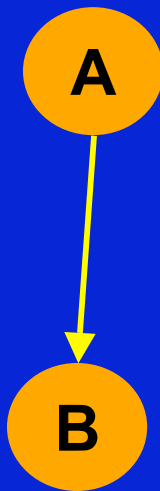
A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions.

Syntax:

1. a set of nodes, one per random variable
2. links mean parent “directly influences” child
3. a directed, acyclic graph
4. a conditional distribution for each node given its parents:
 $P(X_i \mid \text{Parents}(X_i))$

In the simplest case, conditional distribution represented as a conditional probability table (CPT)

A two node net & Conditional probability table



- Node A is independent of Node B, so it is described by an unconditional probability $P(A)$
- $P(\neg A)$ is given by $1 - P(A)$
- Node B is conditionally dependent on A. It is described by four numbers, $P(B | A)$, $P(B | \neg A)$, $P(\neg B | A)$ and $P(\neg B | \neg A)$.
- This can be expressed as 2 by 2 conditional probability table (CPT).
- But $P(\neg B | A) = 1 - P(B | A)$ and $P(\neg B | \neg A) = 1 - P(B | \neg A)$.
- Therefore, only two independent numbers in CPT

Example

I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls

Network topology reflects "causal" knowledge:

