

22c:145 Artificial Intelligence

Bayesian Networks

- Reading: Ch 14. Russell & Norvig

1

Review of Probability Theory

- **Random Variables**

- The probability that a random variable X has value val is written as $P(X=val)$

- P : domain $\subseteq [0, 1]$

– Sums to 1 over the domain:

- » $P(\text{Raining} = \text{true}) = P(\text{Raining}) = 0.2$
- » $P(\text{Raining} = \text{false}) = P(\neg \text{Raining}) = 0.8$

- **Joint distribution:**

- $P(X_1, X_2, \dots, X_n)$

	Toothache	\neg Toothache
Cavity	0.04	0.06
\neg Cavity	0.01	0.89

- Probability assignment to all combinations of values of random variables and provide complete information about the probabilities of its random variables.

- A JPD table for n random variables, each ranging over k distinct values, has k^n entries!

2

Review of Probability Theory

- **Conditioning**

- $P(A) = P(A | B) P(B) + P(A | \neg B) P(\neg B)$
 $= P(A \wedge B) + P(A \wedge \neg B)$

- A and B are **independent** iff

- $P(A \wedge B) = P(A) P(B)$
- $P(A | B) = P(A)$
- $P(B | A) = P(B)$

- A and B are **conditionally independent** given C iff

- $P(A | B, C) = P(A | C)$
- $P(B | A, C) = P(B | C)$
- $P(A \wedge B | C) = P(A | C) P(B | C)$

- **Bayes' Rule**

- $P(A | B) = P(B | A) P(A) / P(B)$
- $P(A | B, C) = P(B | A, C) P(A | C) / P(B | C)$

3

Bayesian Networks

- To do probabilistic reasoning, you need to know the joint probability distribution
- But, in a domain with N propositional variables, one needs 2^N numbers to specify the joint probability distribution
- We want to exploit independences in the domain
- Two components: structure and numerical parameters

4

Bayesian (Belief) Networks

- Set of random variables, each has a finite set of values
- Set of directed arcs between them forming acyclic graph, representing causal relation
- Every node A , with parents B_1, \dots, B_n , has $P(A \mid B_1, \dots, B_n)$ specified

5

Key Advantage

- The conditional independencies (missing arrows) mean that we can store and compute the joint probability distribution more efficiently

How to design a Belief Network?

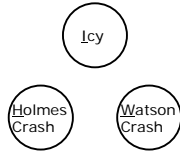
- Explore the causal relations

6

Icy Roads

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

"Causal" Component

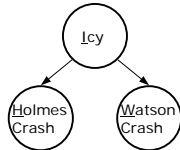


7

Icy Roads

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

"Causal" Component

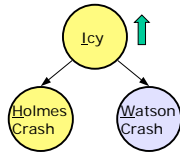


8

Icy Roads

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

"Causal" Component

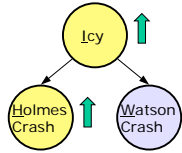


9

Icy Roads

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

"Causal" Component



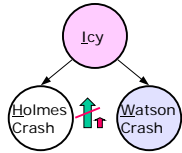
H and W are dependent,

10

Icy Roads

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

"Causal" Component

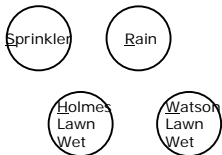


H and W are dependent, but conditionally independent given I

11

Holmes and Watson in IA

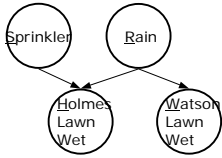
Holmes and Watson have moved to IA. He wakes up to find his lawn wet. He wonders if it has rained or if he left his sprinkler on. He looks at his neighbor Watson's lawn and he sees it is wet too. So, he concludes it must have rained.



12

Holmes and Watson in IA

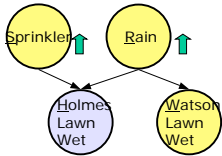
Holmes and Watson have moved to IA. He wakes up to find his lawn wet. He wonders if it has rained or if he left his sprinkler on. He looks at his neighbor Watson's lawn and he sees it is wet too. So, he concludes it must have rained.



13

Holmes and Watson in IA

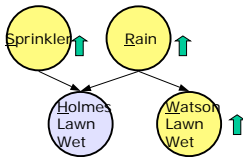
Holmes and Watson have moved to IA. He wakes up to find his lawn wet. He wonders if it has rained or if he left his sprinkler on. He looks at his neighbor Watson's lawn and he sees it is wet too. So, he concludes it must have rained.



14

Holmes and Watson in IA

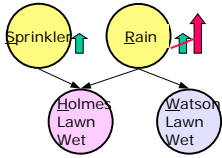
Holmes and Watson have moved to IA. He wakes up to find his lawn wet. He wonders if it has rained or if he left his sprinkler on. He looks at his neighbor Watson's lawn and he sees it is wet too. So, he concludes it must have rained.



15

Holmes and Watson in IA

Holmes and Watson have moved to IA. He wakes up to find his lawn wet. He wonders if it has rained or if he left his sprinkler on. He looks at his neighbor Watson's lawn and he sees it is wet too. So, he concludes it must have rained.

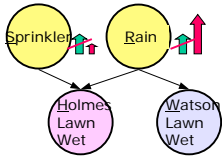


Given W, P(R) goes up

16

Holmes and Watson in IA

Holmes and Watson have moved to IA. He wakes up to find his lawn wet. He wonders if it has rained or if he left his sprinkler on. He looks at his neighbor Watson's lawn and he sees it is wet too. So, he concludes it must have rained.



Given W, P(R) goes up and P(S) goes down – “explaining away”

17

Inference in Bayesian Networks

Query Types

Given a Bayesian network, what questions might we want to ask?

- Conditional probability query: $P(x | e)$
- Maximum a posteriori probability:
What value of x maximizes $P(x|e)$?

General question: What's the whole probability distribution over variable X given evidence e , $P(X | e)$?

18

Using the joint distribution

To answer any query involving a conjunction of variables, sum over the variables not involved in the query.

$$\begin{aligned} \Pr(d) &= \sum_{ABC} \Pr(a,b,c,d) \\ &= \sum_{a \in \text{dom}(A)} \sum_{b \in \text{dom}(B)} \sum_{c \in \text{dom}(C)} \Pr(A=a \wedge B=b \wedge C=c) \end{aligned}$$

19

Using the joint distribution

To answer any query involving a conjunction of variables, sum over the variables not involved in the query.

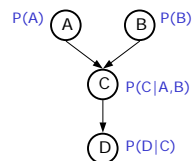
$$\begin{aligned} \Pr(d) &= \sum_{ABC} \Pr(a,b,c,d) \\ &= \sum_{a \in \text{dom}(A)} \sum_{b \in \text{dom}(B)} \sum_{c \in \text{dom}(C)} \Pr(A=a \wedge B=b \wedge C=c) \end{aligned}$$

$$\Pr(d | b) = \frac{\Pr(b,d)}{\Pr(b)} = \frac{\sum_{AC} \Pr(a,b,c,d)}{\sum_{ACD} \Pr(a,b,c,d)}$$

20

Chain Rule

- Variables: V_1, \dots, V_n
- Values: v_1, \dots, v_n
- $P(V_1=v_1, V_2=v_2, \dots, V_n=v_n) = \prod_i P(V_i=v_i | \text{parents}(V_i))$

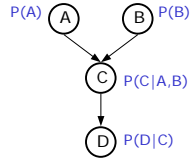


21

Chain Rule

- Variables: V_1, \dots, V_n
- Values: v_1, \dots, v_n
- $P(V_1=v_1, V_2=v_2, \dots, V_n=v_n) = \prod_i P(V_i=v_i \mid \text{parents}(V_i))$

$$P(ABCD) = P(A=\text{true}, B=\text{true}, C=\text{true}, D=\text{true})$$



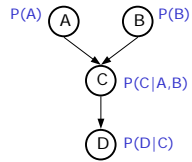
22

Chain Rule

- Variables: V_1, \dots, V_n
- Values: v_1, \dots, v_n
- $P(V_1=v_1, V_2=v_2, \dots, V_n=v_n) = \prod_i P(V_i=v_i \mid \text{parents}(V_i))$

$$P(ABCD) = P(A=\text{true}, B=\text{true}, C=\text{true}, D=\text{true})$$

$$P(ABCD) = P(D|ABC)P(ABC)$$



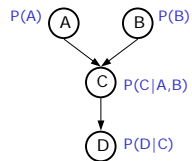
23

Chain Rule

- Variables: V_1, \dots, V_n
- Values: v_1, \dots, v_n
- $P(V_1=v_1, V_2=v_2, \dots, V_n=v_n) = \prod_i P(V_i=v_i \mid \text{parents}(V_i))$

$$P(ABCD) = P(A=\text{true}, B=\text{true}, C=\text{true}, D=\text{true})$$

$$P(ABCD) = P(D|ABC)P(ABC) = P(D|C) P(ABC) =$$



A independent from D given C
B independent from D given C

24

Chain Rule

- Variables: V_1, \dots, V_n
- Values: v_1, \dots, v_n
- $P(V_1=v_1, V_2=v_2, \dots, V_n=v_n) = \prod_i P(V_i=v_i \mid \text{parents}(V_i))$

$P(ABCD) = P(A=\text{true}, B=\text{true}, C=\text{true}, D=\text{true})$

$P(ABCD) =$
 $P(D|ABC)P(ABC) =$
 $P(D|C) P(ABC) =$
 $P(D|C) P(C|AB) P(AB) =$

A independent from D given C
 B independent from D given C

25

Chain Rule

- Variables: V_1, \dots, V_n
- Values: v_1, \dots, v_n
- $P(V_1=v_1, V_2=v_2, \dots, V_n=v_n) = \prod_i P(V_i=v_i \mid \text{parents}(V_i))$

$P(ABCD) = P(A=\text{true}, B=\text{true}, C=\text{true}, D=\text{true})$

$P(ABCD) =$
 $P(D|ABC)P(ABC) =$
 $P(D|C) P(ABC) =$
 $P(D|C) P(C|AB) P(AB) =$
 $P(D|C) P(C|AB) P(A)P(B) =$

A independent from D given C
 B independent from D given C
 A independent from B

26

Chain Rule

- Variables: V_1, \dots, V_n
- Values: v_1, \dots, v_n
- $P(V_1=v_1, V_2=v_2, \dots, V_n=v_n) = \prod_i P(V_i=v_i \mid \text{parents}(V_i))$

$P(ABCD) = P(A=\text{true}, B=\text{true}, C=\text{true}, D=\text{true})$

$P(ABCD) =$
 $P(D|ABC)P(ABC) =$
 $P(D|C) P(ABC) =$
 $P(D|C) P(C|AB) P(AB) =$
 $P(D|C) P(C|AB) P(A)P(B) =$

A independent from D given C
 B independent from D given C
 A independent from B

27

Icy Roads with Numbers

t = true
f = false

P(I=t)	P(I=f)
0.7	0.3

The right-hand column in these tables is redundant, since we know the entries in each row must add to 1.
NB: the columns need NOT add to 1.

28

Icy Roads with Numbers

t = true
f = false

P(I=t)	P(I=f)
0.7	0.3

	P(W=t I)	P(W=f I)
I=t	0.8	0.2
I=f	0.1	0.9

The right-hand column in these tables is redundant, since we know the entries in each row must add to 1.
Note: the columns need NOT add to 1.

29

Icy Roads with Numbers

t = true
f = false

P(I=t)	P(I=f)
0.7	0.3

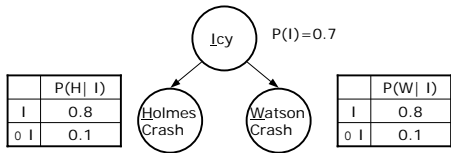
	P(H=t I)	P(H=f I)
I=t	0.8	0.2
I=f	0.1	0.9

	P(W=t I)	P(W=f I)
I=t	0.8	0.2
I=f	0.1	0.9

The right-hand column in these tables is redundant, since we know the entries in each row must add to 1.
Note: the columns need NOT add to 1.

30

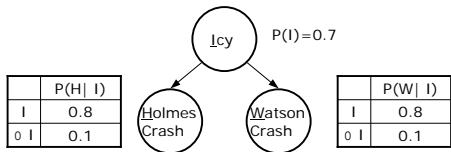
Probability that Watson Crashes



$P(W) =$

31

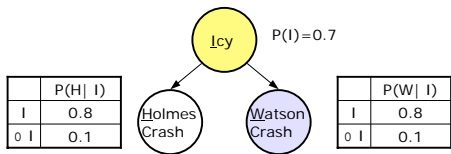
Probability that Watson Crashes



$$\begin{aligned}
 P(W) &= P(W| I) P(I) + P(W|o I) P(o I) \\
 &= 0.8 \cdot 0.7 + 0.1 \cdot 0.3 \\
 &= 0.56 + 0.03 \\
 &= 0.59
 \end{aligned}$$

32

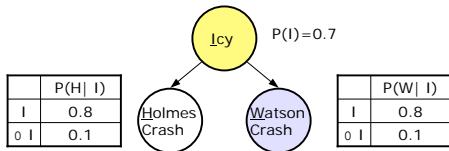
Probability of Icy given Watson



$P(I | W) =$

33

Probability of Icy given Watson

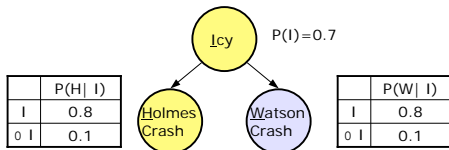


$$\begin{aligned}
 P(I | W) &= P(W | I) P(I) / P(W) \\
 &= 0.8 * 0.7 / 0.59 \\
 &= 0.95
 \end{aligned}$$

We started with $P(I) = 0.7$; knowing that Watson crashed raised the probability to 0.95

34

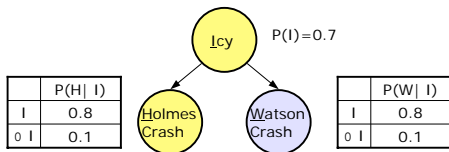
Probability of Holmes given Watson



$$P(H|W) =$$

35

Probability of Holmes given Watson

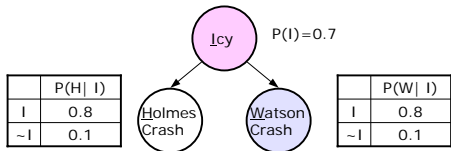


$$\begin{aligned}
 P(H|W) &= P(H, I | W) + P(H, -I | W) \\
 &= P(H|W, I)P(I|W) + P(H|W, o I) P(o I | W) \\
 &= P(H|I)P(I|W) + P(H|o I) P(o I | W) \\
 &= 0.8 * 0.95 + 0.1 * 0.05 \\
 &= 0.765
 \end{aligned}$$

We started with $P(H) = 0.59$; knowing that Watson crashed raised the probability to 0.765

36

Prob of Holmes given Icy and Watson

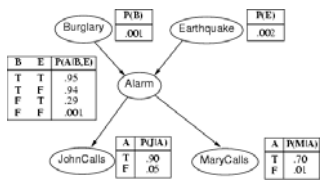


$$P(H|W, \sim I) = P(H \sim I) = 0.1$$

H and W are independent given I, so H and W are conditionally independent given I

37

Exercises



- $P(J, M, A, B, E) = ?$
- $P(M, A, B) = ?$
- $P(\sim M, A, B) = ?$
- $P(A, B) = ?$
- $P(M, B) = ?$
- $P(A | J) = ?$

38
