Lecture 19

Conditional Independence, Bayesian networks intro

Announcement

Lecture Overview

- Recap lecture 18
- Marginal Independence
- Conditional Independence
- Bayesian Networks Introduction

Probability Distributions

Consider the case where possible worlds are simply assignments to one random variable.

Definition (probability distribution)

A probability distribution P on a random variable X is a function dom(X) [0,1] such that

x P(X=x)

Example: X represents a female adult's hight in Canada with domain {short, normal, tall} – based on some definition of these terms

short P(hight = short) = 0.2

normal P(hight = normal) = 0.5 tall

 P(hight = tall) = 0.3

Joint Probability Distribution (JPD)

- Joint probability distribution over random variables X_1, …, X_n:
- a probability distribution over the joint random variable $<X_1, \ldots, X_n>$ with domain $dom(X_1) \times \ldots \times dom(X_n)$ (the Cartesian product)
- Think of a joint distribution over nvariables as the table of the corresponding possible worlds
- There is a column (dimension) for each variable, and one for the probability
- Each row corresponds to an assignment

  $X_1 = x_1, \ldots, X_n = x_n$ and its probability $P(X_1 = x_1, \ldots ,X_n = x_n)$
- We can also write $P(X_1 = x_1$ □ … □ $X_n = x_n)$
- The sum of probabilities across the whole table is 1.

{Weather, Temperature} example from before

Recap: Conditioning

Weather	*Temperature*	*μ(w)*
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

- Conditioning: revise beliefs based on new observations

- We need to integrate two sources of knowledge
- Prior probability distribution P(X): all background knowledge
- New evidence e

- Combine the two to form a posterior probability distribution
- The conditional probability $P(h|e)$

Recap: Conditional probability

E.g. $P(T = hot|W = sunny) = \frac{P(T=hot \wedge W=sunny)}{P(W=sunny)}$

Possible world	*Weather*	*Temperature*	*µ(w)*
w_1	sunny	hot	0.10
w_2	sunny	mild	0.20
w_3	sunny	cold	0.10
~~w_4~~	~~cloudy~~	~~hot~~	~~0.05~~
~~w_5~~	~~cloudy~~	~~mild~~	~~0.35~~
~~w_6~~	~~cloudy~~	~~cold~~	~~0.20~~

T	*P(T|W=sunny)*
hot	0.10/0.40=0.25
mild	0.20/0.40=0.50
cold	0.10/0.40=0.25

JPD for P(T|W=sunny)

Definition (conditional probability)

The conditional probability of formula h given evidence e is

$$P(h|e) = \frac{P(h \wedge e)}{P(e)}$$

Recap: Inference by Enumeration

- Great, we can compute arbitrary probabilities now!

- Given
- Prior joint probability distribution (JPD) on set of variables X
- specific values e for the evidence variables E (subset of X)

- We want to compute
- posterior joint distribution of query variables Y (a subset of X) given evidence e

- Step 1: Condition to get distribution P(X|e)

- Step 2: Marginalize to get distribution P(Y|e)

Inference by Enumeration: example

- Given P(W,C,T) as JPD below, and evidence e : “Wind=yes”

Windy W	*Cloudy C*	*Temperature T*	*P(W, C, T)*
yes	no	hot	0.04
yes	no	mild	0.09
yes	no	cold	0.07
yes	yes	hot	0.01
yes	yes	mild	0.10
yes	yes	cold	0.12
~~no~~	~~no~~	~~hot~~	~~0.06~~
~~no~~	~~no~~	~~mild~~	~~0.11~~
~~no~~	~~no~~	~~cold~~	~~0.03~~
~~no~~	~~yes~~	~~hot~~	~~0.04~~
~~no~~	~~yes~~	~~mild~~	~~0.25~~
~~no~~	~~yes~~	~~cold~~	~~0.08~~

Inference by Enumeration: example

- What is the probability that it is cold? I.e., P(T= cold | W=yes)
- Step 1: condition to get distribution P(C, T| W=yes)

Cloudy C	*Temperature T*	*P(C, T| W=yes)*
no	hot	0.04/0.43 □ 0.10
no	mild	0.09/0.43 □ 0.21
no	cold	0.07/0.43 □ 0.16
yes	hot	0.01/0.43 □ 0.02
yes	mild	0.10/0.43 □ 0.23
yes	cold	0.12/0.43 □ 0.28

$$P(C = c \wedge T = t | W = yes) = \frac{P(C=c \wedge T=t \wedge W=yes)}{P(W=yes)}$$

- Given P(W,C,T) as JPD in previous slide, and evidence e : “Wind=yes”
- What is the probability that it is cold? I.e., P(T=cold | W=yes)
- Step 2: marginalize over Cloudy to get distribution P(T | W=yes)

Cloudy C	*Temperature T*	*P(C, T| W=yes)*
sunny	hot	0.10
sunny	mild	0.21
sunny	cold	0.16
cloudy	hot	0.02
cloudy	mild	0.23
cloudy	cold	0.28

Temperature T	*P(T| W=yes)*
hot	0.10+0.02 = 0.12
mild	0.21+0.23 = 0.44
cold	0.16+0.28 = 0.44

- This is a probability distribution: it defines the probability of all the possible values of Temperature (three here), given the observed value for Windy (yes).
- Because this is a probability distribution, the sum of all its values is

P(T=cold | W=yes) is a specific entry of the probability distribution for P(T | W=yes)

Conditional Probability among Random Variables

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

It expresses the conditional probability of each possible value for X given each possible value for Y

$P(X \mid Y)$ = P(Temperature | Weather) = P(Temperature □ Weather) / P(Weather)

Example:

Temperature {hot, cold}; Weather = {sunny, cloudy}

P(Temperature | Weather)

	T = hot	T = cold
W = sunny	P(hot|sunny)	P(cold|sunny)
W = cloudy	P(hot|cloudy)	P(cold|cloudy)

Which of the following is true?

A. The probabilities in each row should sum to 1

B. The probabilities in each column should sum to 1

C. Both of the above

D. None of the above

Conditional Probability among Random Variables

P(X | Y) = P(X , Y) / P(Y)

It expresses the conditional probability of each possible value for X given each possible value for Y

P(X | Y) = P(Temperature | Weather) = P(Temperature Weather) / P(Weathe

Example:

Temperature {hot, cold}; Weather = {sunny, cloudy}

P(Temperature | Weather)

	T = hot	T = cold
W = sunny	P(hot|sunny)	P(cold|sunny)
W = cloudy	P(hot|cloudy)	P(cold|cloudy)

P(T | Weather = sunny)

P(T | Weather = cloudy)

A. The probabilities in each row should

These are two JPDs!

sum to 1

Recap: Inference by Enumeration

- Great, we can compute arbitrary probabilities now!
- Given
- Prior joint probability distribution (JPD) on set of variables X
- specific values e for the evidence variables E (subset of X)
- We want to compute
- posterior joint distribution of query variables Y (a subset of X) given evidence e

- Step 1: Condition to get distribution P(X|e)
- Step 2: Marginalize to get distribution P(Y|e)

Generally applicable, but memory-heavy and slow
We will see a better way to do probabilistic inference

Bayes rule and Chain Rule

Theorem (Bayes theorem, or Bayes rule)

$$P(h|e) = \frac{P(e|h) \times P(h)}{P(e)}$$

$$P(\mathit{fire} \mid \mathit{alarm}) \square$$

Bayes rule and Chain Rule

Theorem (Bayes theorem, or Bayes rule)

$$P(h|e) = \frac{P(e|h) \times P(h)}{P(e)}$$

E.g., $P(fire|\text{alarm}) = \frac{P(alarm|fire) \times P(fire)}{P(alarm)}$

Product Rule

- By definition, we know that :

$$P(f_2 \mid f_1) = \frac{P(f_2 \wedge f_1)}{P(f_1)}$$

- We can rewrite this to

$$P(f_2 \wedge f_1) = P(f_2 \mid f_1) \times P(f_1)$$

Theorem (Product Rule)

$$P(f_n \wedge \cdots \wedge f_{i+1} \wedge f_i \wedge \cdots \wedge f_1) = P(f_n \wedge \cdots \wedge f_{i+1} | f_i \wedge \cdots \wedge f_1) \times P(f_i \wedge \cdots \wedge f_1)$$

- In general

Chain Rule

- We know

$$P(f_2 \wedge f_1) = P(f_2 | f_1) \times P(f_1)$$

- In general:

$$\begin{aligned}
&P(f_n \wedge f_{n-1} \wedge \cdots \wedge f_1) \\
&= P(f_n | f_{n-1} \wedge \cdots \wedge f_1) \times P(f_{n-1} \wedge \cdots \wedge f_1) \\
&= P(f_n | f_{n-1} \wedge \cdots \wedge f_1) \times P(f_{n-1} | f_{n-2} \wedge \cdots \wedge f_1) \\
&\qquad \times P(f_{n-2} \wedge \cdots \wedge f_1) \\
&= \ldots \\
&= \prod_{i=1}^{n} P(f_i | f_{i-1} \wedge \cdots \wedge f_1)
\end{aligned}$$

Theorem: Chain Rule

$$P(f_1, \ldots, f_n) = \prod_{i=1}^{n} P(f_i \mid f_{i-1}, \ldots, f_1)$$

Chain Rule example

$$P(f_1, \dots, f_n) = \prod_{i=1}^{n} P(fi | f_{i-1}, \dots, f_1)$$

P(A,B,C,D)

= P(D|A,B,C) × P(A,B,C) =

= P(D|A,B,C) × P(C|A,B) × P(A,B)

= P(D|A,B,C) × P(C|B,A) × P(B|A) × P(A)

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

Chain Rule

- Allows representing a Join Probability Distribution (JPD) as the product of conditional probability distributions

Theorem: Chain Rule

$$P(f_1, \ldots, f_n) = \prod_{i=1}^{n} P(f_i \mid f_{i-1}, \ldots, f_1)$$

E.g. $P(A,B,C,D) = P(A) \times P(B|A) \times P(C|A,B) \times P(D|A,B,C)$

Why does the chain rule help us?

We will see how, under specific circumstances (variables independence), this rule helps gain compactness

- We can represent the JPD as a product of marginal distributions
- We can simplify some terms when the variables involved are marginally independent or conditionally independent

Lecture Overview

- Recap lecture 18
- Marginal Independence
- Conditional Independence
- Bayesian Networks Introduction

Marginal Independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y, written $X \perp\!\!\!\perp Y$, if for all $x \in \mathrm{dom}(X)$, $y_j \in \mathrm{dom}(Y)$ and $y_k \in \mathrm{dom}(Y)$, the following equation holds:

$$
\begin{aligned}
& P(X = x | Y = y_j) \\
= \; & P(X = x | Y = y_k) \\
= \; & P(X = x)
\end{aligned}
$$

- Intuitively: if $X \perp\!\!\!\perp Y$, then
  - learning that Y=y does not change your belief in X
  - and this is true for all values y that Y could take

- For example, weather is marginally independent of the result of a coin toss

Examples for marginal independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$
\begin{aligned}
& P(X = xi|Y = y_j) \\
= {} & P(X = xi|Y = y_k) \\
= {} & P(X = xi)
\end{aligned}
$$

Weather W	*Temperature T*	*P(W,T)*
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05

Examples for marginal independence

- Is Temperature marginally independent of Weather (see previous example)?

A. yes

B. no

C. It depends of the value of T

Weather W	*Temperature T*	*P(W,T)*
sunny	hot	0.10
sunny	mild	0.20
cloudy	mild	0.35
cloudy	cold	0.20

sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

- Is

D. It depends of the value of W

Temperature marginally independent of Weather (see previous example

Examples for marginal independence

T	P(T
hot	0.1
mild	0.5
cold	0.3

T	P(T|W=sunny)
hot	0.25
mild	0.50
cold	0.25

Definition (Marginal independence)

Random variable X is (marginally) independent of rando variable Y if, for all $x_i \in$ dom(X), $y_j \in$ dom(Y) and $y_k \in$ do the following equation holds:

$$P(X = xi | Y = y_j) = P(X = xi | Y = y_k) = P(X = xi)$$

Is Weather marginally independent of temperature?

- No. We saw before that knowing the Temperature changes our belief on the weather
- E.g. P(hot) = 0.15 P(hot|sunny) = 0.25

T	*P(T)*
hot	0.15
mild	0.55
cold	0.30

T	*P(T|W=sunny)*
hot	0.25
mild	0.50
cold	0.25

Examples for marginal independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$P(X = xi|Y = y_j)$$
$$= P(X = xi|Y = y_k)$$
$$= P(X = xi)$$

Is Weather marginally independent of Temperature?

- We could have answered this question even without having the relevant probability distributions.
  - ✓ Meteorological knowledge tells us that the weather influences the temperature, so information on what the weather is like should change one's belief on the temperature

Examples for marginal independence

- If fact, for knowledge representation purposes, the evaluation for independence among variables will generally need to be made without numbers, based on pre-exiting domain knowledge or assumptions

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$P(X = xi|Y = y_j)$$
$$= P(X = xi|Y = y_k)$$
$$= P(X = xi)$$

iclicker.

- Intuitively (without numbers):
  - Boolean random variable “Canucks win the Stanley Cup this season”

Examples for marginal independence

- Numerical random variable “Canucks’ revenue last season” ?
- Are the two marginally independent?

A. yes

B. no

C. It depends of the value of Canucks Win SC

D. It depends of the value of Canucks Revenue

- Intuitively (without numbers):
- Boolean random variable “Canucks win the Stanley Cup this season”

Examples for marginal independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$ and $y_k \in dom(Y)$, the following equation holds:

$$P(X = xi|Y = y_j)$$
$$= P(X = xi|Y = y_k)$$
$$= P(X = xi)$$

- Numerical random variable "Canucks' revenue last season" ?
- Are the two marginally independent?
  ✓No! Without revenue they cannot afford to keep their best players

Exploiting marginal independence

Recall the product rule $p(X=x \wedge Y=y) = p(X=x \mid Y=y) \times p(Y=y)$

If X and Y are marginally independent, $p(X=x \mid Y=y) = p(X=x)$

Thus we have

$$p(X=x \wedge Y=y) = p(X=x) \times p(Y=y)$$

In distribution form $p(X,Y) = p(X) \times p(Y)$

Exploiting marginal independence

- If $X_1, \ldots, X_n$ are marginally independent, then we can represent their JPD as a product of marginal distributions

$$P(X_1, \ldots, Xn) = \prod_{i=1}^{n} P(Xi)$$

- If all of $X_1, \ldots, X_n$ are Boolean, how many entries does the JPD $P(X_1, \ldots, Xn)$ have?

Exploiting marginal independence

- If $X_1, \ldots, X_n$ are marginally independent, then we can represent their JPD as a product of marginal distributions

$$P(X_1, \ldots, Xn) = \prod_{i=1}^{n} P(Xi)$$

- If all of $X_1, \ldots, X_n$ are Boolean, how many entries does the JPD $P(X_1, \ldots, Xn)$ have?
  - One entry for each possible world: 2^n

- How many entries would all the marginal distributions have combined?

A. 2^n B. $2n$ C. $2+n$ D. n^2

Exploiting marginal independence

- If $X_1, \ldots, X_n$ are marginally independent, then we can represent their JPD as a product of marginal distributions

$$P(X_1, \ldots, Xn) = \prod_{i=1}^{n} P(Xi)$$

- If all of $X_1, \ldots, X_n$ are Boolean, how many entries does the JPD $P(X_1, \ldots, Xn)$ have?
  - One entry for each possible world: 2^n

- How many entries would all the marginal distributions have combined?
  - Each of the n tables only has two entries $P(X_1 = \text{true})$ and $P(X_1 = \text{false})$
  - So, in total: 2n. Exponentially fewer than the JPD!
  - Exponentially fewer than the JPD!

Given the binary variables A,B,C,D,

To specify P(A,B,C,D) one needs the JDP below

A	B	C	D	P(A,B,C
T	T	T	T	
T	T	T	F	
T	T	F	T	
T	T	F	F	
T	F	T	T	
T	F	T	F	
T	F	F	T	
T	F	F	F	
F	T	T	T	

To specify P(A)×P(B) ×P(C)×P(D) one needs the JDPs below

A	P(A)
T	
F	

B	P(B)
T	
F	

C	P(C)
T	

F	T	T	F	
F	T	F	T	
F	T	F	F	
F	F	T	T	
F	F	T	F	
F	F	F	T	
F	F	F	F	

F	

D	P(D)
T	
F	

Lecture Overview

- Recap lecture 18
- Marginal Independence

- Conditional Independence
- Bayesian Networks Introduction

Conditional Independence

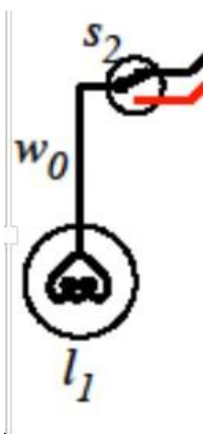
Definition (Conditional independence)

Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$ the following equation holds:

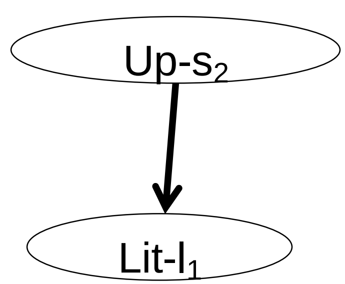
$$P(X = xi|Y = y_j, Z = zm)$$
$$= P(X = xi|Y = y_k, Z = zm)$$
$$= P(X = xi|Z = zm)$$

- Intuitively: if X and Y are conditionally independent given Z, then
- learning that Y=y does not change your belief in X when we already know Z=z
- and this is true for all values y that Y could take and all values z that Z could take

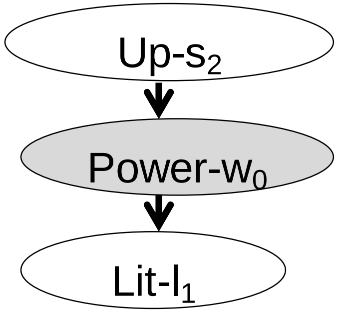
Example for Conditional Independence



- Whether light l_1 is lit (Lit-l_1) and the position of switch s_2 (Up-s_2) are not marginally independent • The position of the switch determines whether there is power in the wire w_0 connected to the light



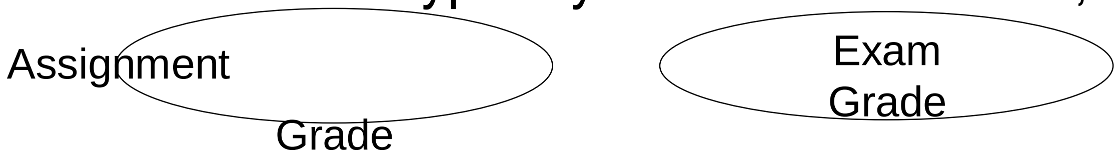
- However, whether light l_1 is lit is conditionally independent from the position of switch s_2 given whether there is power in wire w_0 (Power-w_0)
- Once we know Power-w_0, learning values for Up-s_2 does not change our beliefs about Lit-l_1
- I.e., Lit-l_1 is conditionally independent of Up-s_2 given Power-w_0



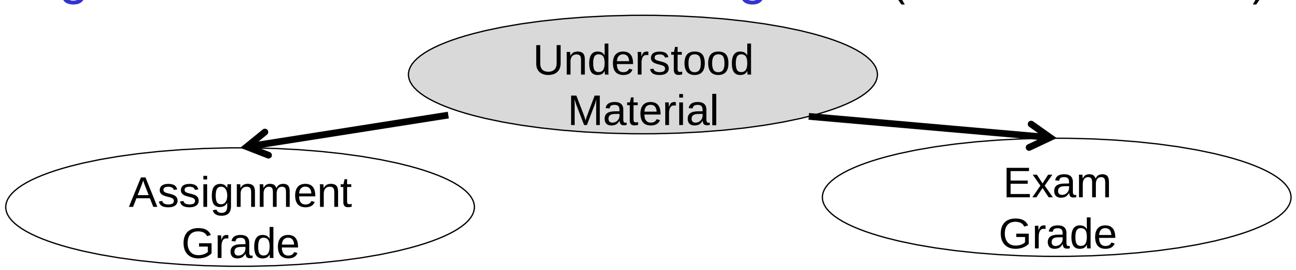
Another example of conditionally but not marginally independent variables

- Exam Grade and Assignment Grade are not marginally independent

- Students who do well on one typically do well on the other, and viceversa



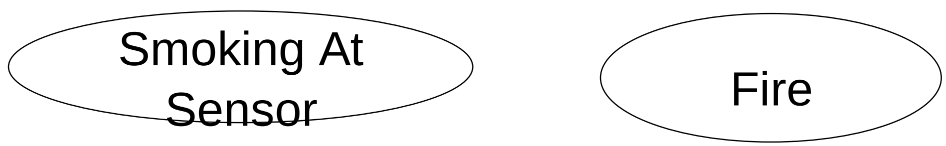
- But, conditional on Understood Material, they are independent
- Variable Understood Material is a common cause of variables Exam Grade and Assignment Grade
- Knowing Understood Material shields any information we could get from Assignment Grade toward Exam grade (and vice versa)



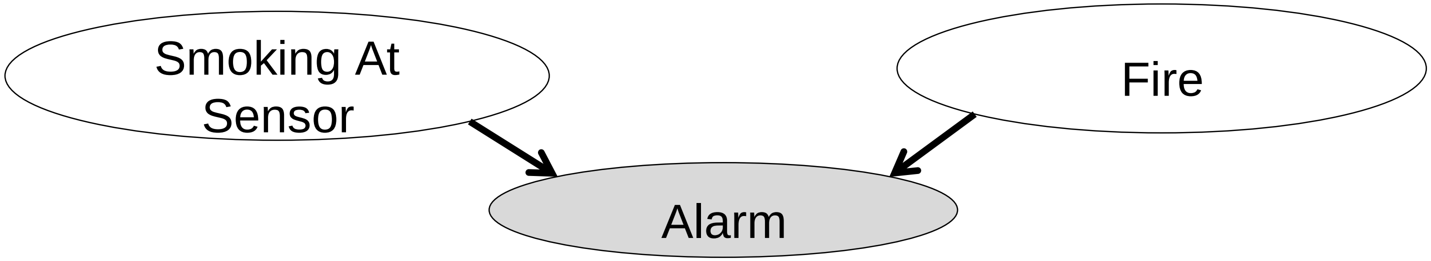
Example: marginally but not conditionally independent

Two variables can be marginally but not conditionally independent

- “Smoking At Sensor” S: resident smokes cigarette next to fire sensor
- “Fire” F: there is a fire somewhere in the building
- “Alarm” A: the fire alarm rings
- S and F are marginally independent
  - ✓ Learning S=true or S=false does not change your belief in F, and viceversa



- But they are not conditionally independent given alarm
  - ✓ They are alternative causes for the alarm ringing – evidence on one of the two causes reduces the belief on the other if the alarm rings
  - ✓ E.g., if the alarm rings and you learn S=true your belief in F decreases

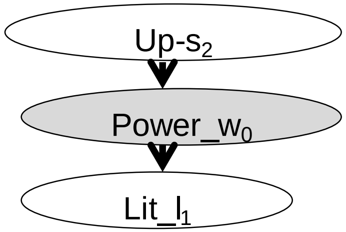


Conditional vs. Marginal Independence

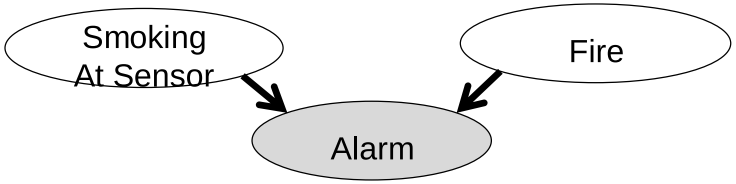
Two variables can be

Conditionally but not marginally independent

- ExamGrade and AssignmentGrade
- ExamGrade and AssignmentGrade given UnderstoodMaterial

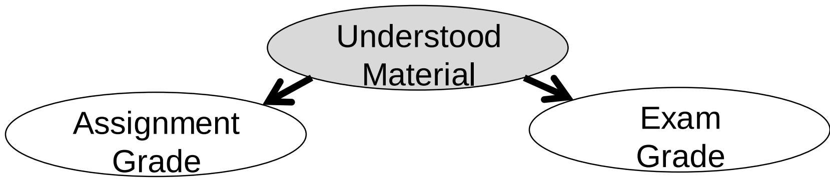


- Lit-l1 and Up-s2
- Lit-l1 and Up-s2 given Power_w_0



Marginally but not conditionally independent

- SmokingAtSensor and Fire
- SmokingAtSensor and Fire given Alarm

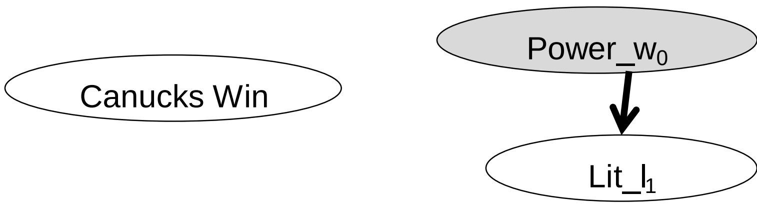


Both marginally and conditionally independent

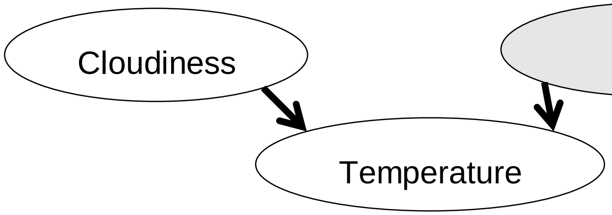
- CanucksWinStanleyCup and Lit_l_1
- CanucksWinStanleyCup and Lit_l_1 given Power_w

Neither marginally nor conditionally independent

- Temperature and Cloudiness
- Temperature and Cloudiness given Wind



0



Exploiting Conditional Independence

- Example 1: Boolean variables A,B,C
- C is conditionally independent of A given B
- We can then rewrite P(C | A,B) as P(   )

Definition (Conditional independence)
Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$ the following equation holds:

$$\begin{aligned} & P(X = xi | Y = y_j, Z = zm) \\ = \; & P(X = xi | Y = y_k, Z = zm) \\ = \; & P(X = xi | Z = zm) \end{aligned}$$

Exploiting Conditional Independence

- Example 1: Boolean variables A,B,C
- C is conditionally independent of A given B
- We can then rewrite P(C | A,B) as P(C|B)

Definition (Conditional independence)
Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$ the following equation holds:

$$\begin{aligned} &P(X = xi | Y = y_j, Z = zm) \\ &= P(X = xi | Y = y_k, Z = zm) \\ &= P(X = xi | Z = zm) \end{aligned}$$

Exploiting Conditional Independence

Example 2: Boolean variables A,B,C,D

- D is conditionally independent of both A and B given C
  - ✓We can rewrite P(D | A,B,C) as P(   )

Definition (Conditional independence)
Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$ the following equation holds:

$$P(X = xi|Y = y_j, Z = zm) \\ = P(X = xi|Y = y_k, Z = zm) \\ = P(X = xi|Z = zm)$$

Exploiting Conditional Independence

Example 2: Boolean variables A,B,C,D

- D is conditionally independent of both A and B given C
  - ✓We can rewrite P(D | A,B,C) as P(D|C)

Definition (Conditional independence)
Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$ the following equation holds:

$$P(X = xi|Y = y_j, Z = zm) \\ = P(X = xi|Y = y_k, Z = zm) \\ = P(X = xi|Z = zm)$$

Exploiting Conditional Independence

- $P(D|C)$ is much simpler to specify than $P(D \mid A,B,C)$!

If A, B, C, D are Boolean variables

P(D | A,B,C) is given by the following table

A	B	C	P(D=T|A,B,C)	P(D=F|A,B,C)
T	T	T		
T	T	F		
T	F	T		
T	F	F		
F	T	T		
F	T	F		
F	F	T		
F	F	F		

How many probability distributions does this table represent?

P(D|C) is given by the following table

C	P(D=T|C)
T	
F	

A. 2 B. 4 C. 8 D. 1

If A, B, C, D are Boolean variables

P(D | A,B,C) is given by the following table

A	B	C	P(D=T|A,B,C)	P(D=F|A,B,C)
T	T	T		
T	T	F		
T	F	T		
T	F	F		
F	T	T		
F	T	F		
F	F	T		
F	F	F		

8 – each row represents the probability distribution for D given the values that A, B and C take in that row

P(D|C) is given by the following table

C	P(D=T|C)	P(D=F|C)
T		
F		

How many probability distributions does this table represent?

If A, B, C, D are Boolean variables

P(D | A,B,C) is given by the following table

A	B	C	P(D=T|A,B,C)	P(D=F|A,B,C)
T	T	T		
T	T	F		
T	F	T		
T	F	F		
F	T	T		
F	T	F		

F	F	T		
F	F	F		

P(D|C) is given by the following table

C	P(D=T|C)	P(D=F|C)
T		
F		

2 – each row represents the probability distribution for D given the value that C takes in that row

Putting It All Together

- Given the JPD P(A,B,C,D), we can apply the chain rule to get

$$P(A,B,C,D) = P(A) \cdot P(B \mid A) \cdot P(C \mid A,B) \cdot P(D \mid A,B,C)$$

- If D is conditionally independent of A and B given C, we can rewrite the above as

$$P(A, B, C, D) \,\square P(A) \square P(B \mid A) \square P(C \mid A, B) \square P(D \mid A)$$

Under independence we gain compactness (fewer/smaller distributions to deal with)

- The chain rule allows us to write the JPD as a product of conditional distributions
- Conditional independence allows us to write them more compactly

Learning Goals For Probability so far

- Define and give examples of random variables, their domains and probability distributions
- Calculate the probability of a proposition f given µ(w) for the set of possible worlds
- Define a joint probability distribution (JPD)

- Marginalize over specific variables to compute distributions over any subset of the variables
- Given a JPD
- Marginalize over specific variables
- Compute distributions over any subset of the variables
- Apply the formula to compute conditional probability P(h|e)
- Use inference by enumeration • to compute joint posterior probability distributions over any subset of variables given evidence
- Derive and use Bayes Rule
- Derive the Chain Rule
- Define and use marginal independence
- Define and use conditional independence

Bayesian (or Belief) Networks

- Bayesian networks and their extensions are Representation & Reasoning systems explicitly defined to exploit independence in probabilistic reasoning

Lecture Overview

- Recap lecture 18
- Marginal Independence
- Conditional Independence

- Bayesian Networks Introduction

FINALLY!

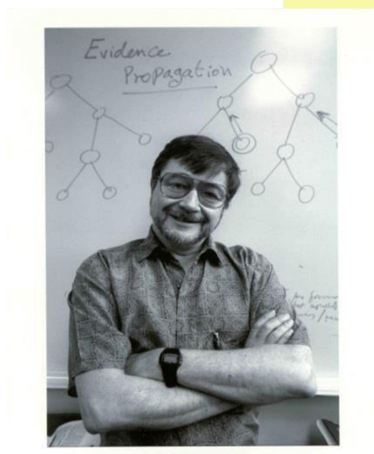
Bayesian Network Motivation

- We want a representation and reasoning system that is based on conditional (and marginal) independence
- Compact yet expressive representation
- Efficient reasoning procedures
- Bayesian (Belief) Networks are such a representation
- Named after Thomas Bayes (ca. 1702 -1761)
- Term coined in 1985 by Judea Pearl (1936 -)

- Their invention changed the primary focus of AI from logic to probability!

Thomas Bayes

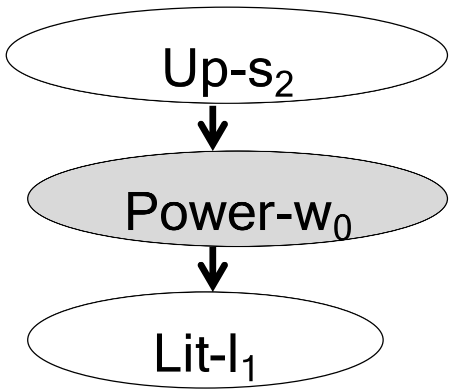
In 2012 Pearl received the very prestigious ACM Turing Award for his contributions to Artificial Intelligence!

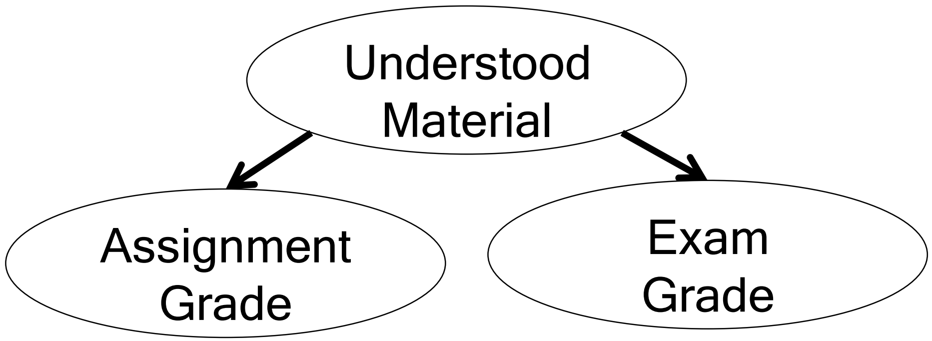


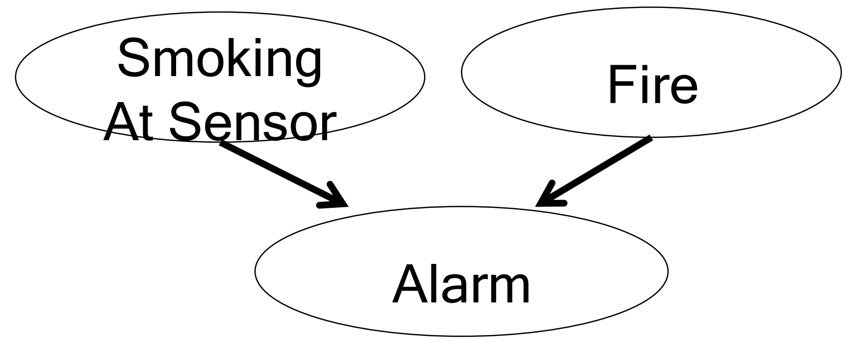
Judea Pearl

Bayesian Networks: Intuition

- A graphical representation for a joint probability distribution
- Nodes are random variables
- Directed edges between nodes reflect dependence
- Some informal examples:







Belief (or Bayesian) networks

Def. A Belief network consists of

- a directed, acyclic graph (DAG) where each node is associated with a random variable X_i
- A domain for each variable X_i
- a set of conditional probability distributions for each node X_i given its parents $Pa(X_i)$ in the graph

$$P\,(X_i \mid Pa(X_i))$$

- The parents $Pa(X_i)$ of a variable X_i are those X_i directly depends on
- A Bayesian network is a compact representation of the JDP for a set of variables $(X_1, \ldots, X_n)$

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i))$$

Bayesian Networks: Definition

Def. A Belief network consists of

- a directed, acyclic graph (DAG) where each node is associated with a random variable X_i
- A domain for each variable X_i
- a set of conditional probability distributions for each node X_i given its parents $Pa(X_i)$ in the graph

$$P\ (X_i \mid Pa(X_i))$$

- Discrete Bayesian networks:
- Domain of each variable is finite
- Conditional probability distribution is a conditional probability table
- We will assume this discrete case

✓But everything we say about independence (marginal & conditional) carries over to the continuous case