Bayesian Networks

Knowledge Representation

- Andreas Sauter
- Dec. 2023
- (Content adapted from Erman Acar)

Overview

1. Foundations
   Degrees of Belief, Belief Dynamics, Independence, Bayes Theorem, Marginalization
2. Bayesian Networks
   Graphs and their Independencies, Bayesian Networks, d-Separation
3. Tools for Inference
   Factors, Variable Elimination, Elimination Order, Interaction Graphs, Graph pruning
4. Exact Inference in Bayesian Networks
   Posterior Marginal, Maximum – A-Posteriori, Most Probable Explanation

Lecture 4: Exact Inference in Bayesian Networks

Lecture Overview

Posterior Marginals

Joint Marginal, Normalization, Computation

More About Factors

Maximizing-Out, Extended Factors

Most Likely Instantiations

Maximum A-Posteriori, Most Probable Explanation, Computation

Computing Posterior Marginals

Definition

We already saw how we can compute the prior marginal $\Pr(Q)$ with $Q \subseteq V$ via variable elimination.

Now we want to compute the distribution over Q given some evidence e via variable elimination

This is called the posterior marginal $Pr(Q|e)$.

E.g.: $\Pr(Alarm = true \mid Earthquake = True)$

Note that the prior marginal is just a special case of the posterior marginal where the evidence is empty.

Department of Computer Science

Approach

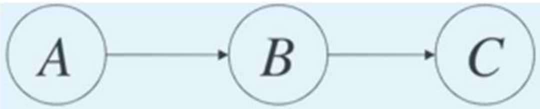
Recall that $\Pr(A|B) = \frac{\Pr(A \wedge B)}{\Pr(B)}$, hence we can compute the posterior marginal through normalizing the joint marginal.

So for query variables Q and evidence e we compute $\Pr(\mathrm{Q}|\mathrm{e}) = \frac{\Pr(Q \wedge e)}{\Pr(e)}$

The overall approach is then as follows:

- Reduce all factors w.r.t. e
- Compute the joint marginal $\Pr(Q \wedge e)$
- Sum-out Q to obtain $\Pr(e)$
- Compute $\Pr(Q|e) = \frac{\Pr(Q \wedge e)}{\Pr(e)}$

Example



G4

A	B	$\Theta_{B\|A}$
true	true	.9
true	false	.1
false	true	.2
false	false	.8

B	C	$\Theta_{C\|B}$
true	true	.3
true	false	.7
false	true	.5
false	false	.5

A	Θ_A
true	.6
false	.4

$$Q = \{C\}, e = \{A = true\}$$

$$\Pr(Q \wedge e) = \Sigma_{A,B}(\Theta_A \Theta_{B|A} \Theta_{C|B})^e$$

$$= \Sigma_{A,B} \Theta_A^e \Theta_{B|A}^e \Theta_{C|B}^e$$

First, reduce all factors w.r.t. $A = true$:

A	$\Theta_A^{\mathbf{e}}$
true	.6

A	B	$\Theta_{B\|A}^{\mathbf{e}}$
true	true	.9
true	false	.1

B	C	$\Theta_{C\|B}^{\mathbf{e}}$
true	true	.3
true	false	.7
false	true	.5
false	false	.5

Then we pick an ordering, say, first A then B

resulting in $\Pr(Q \wedge e) = \Sigma_B \Theta_{C|B}^e \Sigma_A \Theta_A^e \Theta_{B|A}^e$

Example – Continued

So given

A	Θ^e_A
true	.6

A	B	$\Theta^e_{B\|A}$
true	true	.9
true	false	.1

B	C	$\Theta^e_{C\|B}$
true	true	.3
true	false	.7
false	true	.5
false	false	.5

We can compute $\Pr(Q \wedge e) = \Sigma_B \Theta^e_{C|B} \Sigma_A \Theta^e_A \Theta^e_{B|A}$ via variable elimination.

A	B	$\Theta^e_A \Theta^e_{B\|A}$
true	true	.54
true	false	.06

B	$\sum_A \Theta^e_A \Theta^e_{B\|A}$
true	.54
false	.06

B	C	$\Theta^e_{C\|B} \sum_A \Theta^e_A \Theta^e_{B\|A}$
true	true	.162
true	false	.378
false	true	.030
false	false	.030

C	$\sum_B \Theta^e_{C\|B} \sum_A \Theta^e_A \Theta^e_{B\|A}$
true	.192
false	.408

Now, we can sum-out C to obtain $\Pr(e)$, for normalizing to obtain $\Pr(Q \mid e)$:

$$\Pr(\mathrm{a}) = 0.6, \qquad \Pr(c|a) = \frac{\Pr(c \wedge a)}{\Pr(a)} = \frac{0.192}{0.6} = 0.32, \qquad \Pr(\neg c|a) = \frac{\Pr(\neg c \wedge a)}{\Pr(a)} = \frac{0.408}{0.6} = 0.68$$

More About Factors

Maximising-Out – Introduction

Before we saw how to eliminate a variable by summing it out:

B	C	D	f_1
true	true	true	.95
true	true	false	.05
true	false	true	.9
true	false	false	.1
false	true	true	.8
false	true	false	.2
false	false	true	0
false	false	false	1

Sum-out D

B	C	$\sum_D f_1$
true	true	1
true	false	1
false	true	1
false	false	1

For some applications, we want to only keep the instantiation with the maximum probability:

B	C	D	f_1
true	true	true	.95
true	true	false	.05
true	false	true	.9
true	false	false	.1
false	true	true	.8
false	true	false	.2
false	false	true	0
false	false	false	1

max-out D

B	C	$\max_D f_1$
true	true	.95
true	false	.9
false	true	.8
false	false	1

Maximizing-Out – Formalization

Given a factor f over variables $\boldsymbol{X}$ maximizing-out $X \in \boldsymbol{X}$ results in a new factor $(\max_X f)(\boldsymbol{Y}) \stackrel{\text{def}}{=} \max_X f(X, \boldsymbol{Y})$ over variables $\mathbf{Y} = \boldsymbol{X} \backslash \mathrm{X}$.

For an instantiation y of $\boldsymbol{Y}$, $(\max_X f)(y) \stackrel{\text{def}}{=} \max(f(x, y), f(\neg x, y))$

Maximizing out is commutative:

$$\max_X \max_Y f = \max_Y \max_X f$$

If f_1, f_2 are factors and if variable X only appears in factor f_2 then:

$$\max_X f_1 f_2 = f_1 \max_X f_2$$

Extended Factors

When we maximize-out on normal factors, we lose the information on which instantiation maximized the probability.

With extended factors, we can keep this information which is necessary for certain queries.

For every maxed-out instantiation, the extended factor assigns a probability and the instantiation which led to the maximization:

B	C	D	f_1
true	true	true	.95
true	true	false	.05
true	false	true	.9
true	false	false	.1
false	true	true	.8
false	true	false	.2
false	false	true	0
false	false	false	1

max-out D

B	C	$\max_D f_1$
true	true	.95, D=true
true	false	.9, D=true
false	true	.8, D=true
false	false	1, D=false

Most Likely Instantiations

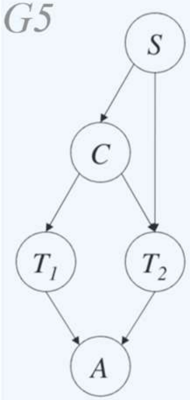
Introduction

Example: S=sex, C=has condition, T_1, T_2= Test results 1 and 2, A= tests agree.

We want to investigate 4 groups: Female\male that have the condition or not.

Query: Given we only know that the the tests agree for a person (A=true), which group does this person belong to?

G5



S	θ_s
male	.55
female	.45

S	C	$\theta_{c\|s}$
male	yes	.05
male	no	.95
female	yes	.01
female	no	.99

C	T_1	$\theta_{t_1\|c}$
yes	+ve	.80
yes	−ve	.20
no	+ve	.20
no	−ve	.80

S	C	T_2	$\theta_{t_2\|c,s}$
male	yes	+ve	.80
male	yes	−ve	.20
male	no	+ve	.20
male	no	−ve	.80
female	yes	+ve	.95
female	yes	−ve	.05
female	no	+ve	.05
female	no	−ve	.95

T_1	T_2	A	$\theta_{a\|t_1,t_2}$
+ve	+ve	yes	1
+ve	+ve	no	0
+ve	−ve	yes	0
+ve	−ve	no	1
−ve	+ve	yes	0
−ve	+ve	no	1
−ve	−ve	yes	1
−ve	−ve	no	0

In other words: what is the most likely instantiation of S and C given A=true?

Maximum A-Posteriori Query

Given a BN over variables V query variables $Q \subseteq V$, and an evidence e the maximum a-posteriori query (MAP) gives us the instantiation $argmax_Q \Pr(Q, e)$ with a probability of $\max_Q \Pr(Q, e)$.

We can do this by first computing $Pr(Q, e)$ with variable elimination as before and then maximizing-out Q using extended factors.

In our example with $Q = \{S, C\}, e = \{A = true\}$, we get a MAP with S=male, C=no with a probability of 0.49.

Most Probable Explanation

Often we are interested in the MAP of all variables except the evidence.

We call a MAP a most probable explanation (MPE), if $Q = V \setminus \mathrm{E}$, where E is the set of variables of the evidence.

We can compute it by maximizing-out all variables with extended factors.

In general, MPE is easier to compute that MAP. Hence, MPE is sometimes used to approximate MAP, but they might disagree.

Example 1 – MPE

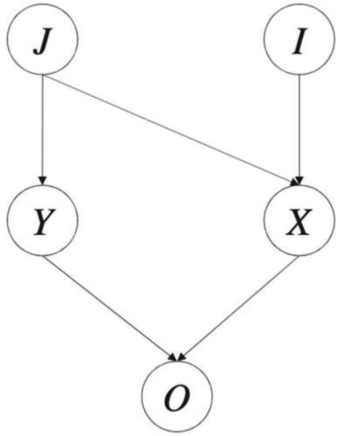
I	Θ_I
true	.5
false	.5

J	Θ_J
true	.5
false	.5

J	Y	$\Theta_{Y\|J}$
true	true	.01
true	false	.99
false	true	.99
false	false	.01

I	J	X	$\Theta_{X\|IJ}$
true	true	true	.95
true	true	false	.05
true	false	true	.05
true	false	false	.95
false	true	true	.05
false	true	false	.95
false	false	true	.05
false	false	false	.95

X	Y	O	$\Theta_{O\|XY}$
true	true	true	.98
true	true	false	.02
true	false	true	.98
true	false	false	.02
false	true	true	.98
false	true	false	.02
false	false	true	.02
false	false	false	.98



What is the MPE given J=true, O=false.

So we have to compute

$$MPE(e) = \max_{V}(\Theta^e_I \Theta^e_J \Theta^e_{Y|J} \Theta^e_{X|I,J} \Theta^e_{O|X,Y})$$

We start by doing network pruning, resulting in:

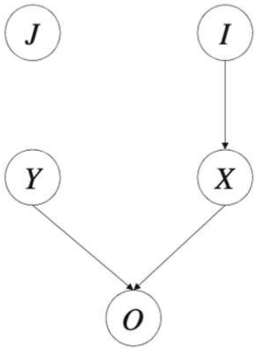
I	Θ^e_I	
true	.5	⊤
false	.5	⊤

J	Θ^e_J	
true	.5	⊤

Y	Θ^e_Y	
true	.01	⊤
false	.99	⊤

I	X	$\Theta^e_{X\|I}$	
true	true	.95	⊤
true	false	.05	⊤
false	true	.05	⊤
false	false	.95	⊤

X	Y	O	$\Theta^e_{O\|XY}$	
true	true	false	.02	⊤
true	false	false	.02	⊤
false	true	false	.02	⊤
false	false	false	.98	⊤

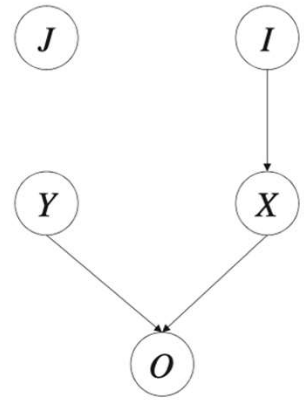


Example 1 - Continued

I	Θ_I^e	
true	.5	$\top$
false	.5	$\top$

J	Θ_J^e	
true	.5	$\top$

Y	Θ_Y^e	
true	.01	$\top$
false	.99	$\top$



Assuming elimination order J,I,X,Y,O:

I	X	$\Theta_{X\|I}^e$	
true	true	.95	$\top$
true	false	.05	$\top$
false	true	.05	$\top$
false	false	.95	$\top$

X	Y	O	$\Theta_{O\|XY}^e$	
true	true	false	.02	$\top$
true	false	false	.02	$\top$
false	true	false	.02	$\top$
false	false	false	.98	$\top$

$$\max_V(\Theta_I^e\Theta_J^e\Theta_Y^e\Theta_{X|I}^e\Theta_{O|X,Y}^e) = \max_O\left(\max_Y\left(\max_X\left(\max_I(\Theta_I^e\Theta_{X|I}^e)\Theta_{O|X,Y}\right)\Theta_Y^e\right)\right)\max_J(\Theta_J^e)$$

	$\max_J \Theta_J^e$	
$\top$	.5	$J=$true

I	X	$\Theta_I^e\Theta_{X\|I}^e$	
true	true	.475	$\top$
true	false	.025	$\top$
false	true	.025	$\top$
false	false	.475	$\top$

X	$\max_I \Theta_I^e\Theta_{X\|I}^e$	
true	.475	$I=$true
false	.475	$I=$false

X	Y	O	$\left(\max_I \Theta_I^e\Theta_{X\|I}^e\right)\Theta_{O\|XY}^e$	
true	true	false	.0095	$I=$true
true	false	false	.0095	$I=$true
false	true	false	.0095	$I=$false
false	false	false	.4655	$I=$false

Y	O	$\max_X\left(\max_I \Theta_I^e\Theta_{X\|I}^e\right)\Theta_{O\|XY}^e$	
true	false	.0095	$I=$true, $X=$true
false	false	.4655	$I=$false, $X=$false

Y	O	$\left(\max_X\left(\max_I \Theta_I^e\Theta_{X\|I}^e\right)\Theta_{O\|XY}^e\right)\Theta_Y^e$	
true	false	.000095	$I=$true, $X=$true
false	false	.460845	$I=$false, $X=$false

O	$\max_Y\left(\max_X\left(\max_I \Theta_I^e\Theta_{X\|I}^e\right)\Theta_{O\|XY}^e\right)\Theta_Y^e$	
false	.460845	$I=$false, $X=$false, $Y=$false

	$\max_O\left(\max_Y\left(\max_X\left(\max_I \Theta_I^e\Theta_{X\|I}^e\right)\Theta_{O\|XY}^e\right)\Theta_Y^e\right)$	
$\top$	.460845	$I=$false, $X=$false, $Y=$false, $O=$false

	$\max_O\left(\max_Y\left(\max_X\left(\max_I \Theta_I^e\Theta_{X\|I}^e\right)\Theta_{O\|XY}^e\right)\Theta_Y^e\right)\left(\max_J \Theta_J^e\right)$	
$\top$	.2304225	$J=$true, $I=$false, $X=$false, $Y=$false, $O=$false

Example 2 – Compute MAP

I	Θ_I
true	.5
false	.5

J	Θ_J
true	.5
false	.5

J	Y	$\Theta_{Y\mid J}$
true	true	.01
true	false	.99
false	true	.99
false	false	.01

I	J	X	$\Theta_{X\mid IJ}$
true	true	true	.95
true	true	false	.05
true	false	true	.05
true	false	false	.95
false	true	true	.05
false	true	false	.95
false	false	true	.05
false	false	false	.95

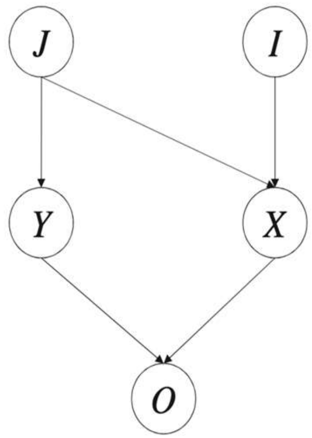
X	Y	O	$\Theta_{O\mid XY}$
true	true	true	.98
true	true	false	.02
true	false	true	.98
true	false	false	.02
false	true	true	.98
false	true	false	.02
false	false	true	.02
false	false	false	.98

We want to know the most likely instantiation of $Q = \{I, J\}$ given $O = true, \pi = O, Y, X, I, J$

We compute $MAP(Q, e) = \max_{Q}\left(\Sigma_{V\setminus Q}\Theta_I^e\Theta_J^e\Theta_{Y|J}^e\Theta_{X|IJ}^e\Theta_{O|X,Y}^e\right)$

$$= \max_{Q} \Pr(Q, e)$$

"First sum-out $V\setminus Q$, then maximize-out Q"



We can compute the marginal $Pr(I, J, e)$ as before resulting in:

I	J	f_1	
true	true	.93248	⊤
true	false	.97088	⊤
false	true	.07712	⊤
false	false	.97088	⊤

I	f_2	
true	.5	⊤
false	.5	⊤

J	f_3	
true	.5	⊤
false	.5	⊤

Example 2 – Continued

I	J	f_1	
true	true	.93248	⊤
true	false	.97088	⊤
false	true	.07712	⊤
false	false	.97088	⊤

I	f_2	
true	.5	⊤
false	.5	⊤

J	f_3	
true	.5	⊤
false	.5	⊤

Now we are left with computing

$$\max_{I,J} \Pr(I,J,e) = \max_{I,J} f_1 f_2 f_3 \;= \max_J(\max_I f_1 f_2\,) f_3$$

I	J	$f_1\ f_2$	
true	true	.466240	⊤
true	false	.485440	⊤
false	true	.038560	⊤
false	false	.485440	⊤

J	$\max_I f_1\ f_2$	
true	.466240	$I=$true
false	.485440	$I=$true

J	$(\max_I f_1\ f_2)\ f_3$	
true	.233120	$I=$true
false	.242720	$I=$true

	$\max_J(\max_I f_1\ f_2)\ f_3$	
⊤	.242720	$I=$true, $J=$false

The MAP of I and J given O=true is: I=true, J=false

Lecture 4 – Summary

- We introduced extended factors and how to maximize them out.

- We saw how to compute posterior marginals.
  “What is the probability of an earthquake when the alarm is ringing.”

- We saw how to compute maximum a-posteriori queries.
  “What is the most likely state of the alarm given someone is calling the police.”

- We saw how to compute most probable explanations.
  “What is the most probable explanation of a ringing alarm.”