Graphical Models - Inference -

Variable Elimination


Outline

- Introduction
- Reminder: Probability theory
- Basics of Bayesian Networks
- Modeling Bayesian networks
- Inference
- Excourse: Markov Networks
- Learning Bayesian networks
- Relational Models

Advanced I WS 06/07

Elimination in Chains

- Forward pass

\[ P(e) = \sum_d \sum_c \sum_b \sum_a P(a, b, c, d, e) \]

- Using definition of probability, we have

By chain decomposition, we get

\[ P(e) = \sum_d \sum_c \sum_b \sum_a P(a, b, c, d, e) \]
\[ P(e) = \sum_d \sum_c \sum_b \sum_a P(a) P(b|a) P(c|b) P(d|c) P(e|d) \]
Elimination in Chains

• Rearranging terms ...

\[ P(e) = \sum_d \sum_c \sum_b \sum_a P(a)P(b|a)P(c|b)P(d|c)P(e|d) \]

\[ P(e) = \sum_d \sum_c \sum_b \sum_a P(c|b)P(d|c)P(e|d) \sum_a P(a)P(b|a) \]

Elimination in Chains

• Now we can perform innermost summation

\[ P(e) = \sum_d \sum_c \sum_b P(c|b)P(d|c)P(e|d) \sum_a P(a)P(b|a) \]

\[ P(e) = \sum_d \sum_c \sum_b P(c|b)P(d|c)P(e|d)P(b) \]

Elimination in Chains with Evidence

• Similar due to Bayes’ Rule

\[ P(a|e) = \frac{P(a,e)}{P(e)} \]

• We write the query in explicit form

\[ P(a,e) = \sum_d \sum_c \sum_b P(a,b,c,d,e) \]

\[ P(a,e) = \sum_d \sum_c \sum_b P(a)P(b|a)P(c|b)P(d|c)P(e|d) \]
Variable Elimination

- Write query in the form
  \[ P(x_n, e) = \sum_{x_1} \sum_{x_2} \ldots \sum_{x_{n-1}} \prod_{i=1}^{n} P(x_i | \text{pa}(x_i)) \]
  Eliminate irrelevant variables  
  Semantic Bayesian Network

- Iteratively
  - Move all irrelevant terms outside of innermost sum
  - Perform innermost sum, getting a new term
  - Insert the new term into the product

Asia - A More Complex Example

Visit to Asia
Smoking
Tuberculosis
Lung Cancer
Abnormality in Chest
Bronchitis
Dyspnoea
X-Ray

We want to compute \( P(d) \)

Need to eliminate: \( v, s, x, t, l, a, b \)

Initial factors
\[ P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t, l)P(x|a), P(d|a, b) \]

Eliminate: \( v \)
Compute: 
\[ f_v(t) = \sum_v P(v)P(t|v) \]
\[ \Rightarrow f_v(t)P(s)P(l|s)P(b|s)P(a|t, l)P(x|a), P(d|a, b) \]

but in general the result of eliminating a variable is not necessarily a probability term!
We want to compute $P(d)$
Need to eliminate: $s, x, t, l, a, b$

Initial factors
$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t, l)P(x|a), P(d|a, b)$$
Rightarrow
$$f_v(t)P(l|s)P(b|s)P(a|t, l)P(x|a), P(d|a, b)$$

Eliminate: $s$
Compute: $f_s(b, l) = \sum_s P(s)P(l|s)P(b|s)$
Rightarrow
$$f_v(t)f_s(b, l)P(a|t, l)P(x|a), P(d|a, b)$$

Summing on $s$ results in a factor with two arguments $f_s(b, l)$. In general, result of elimination may be a function of several variables.

Note that $f_x(a) = 1$ for all values of $a$.
Bayesian Networks

Advanced I

WS 06/07

Initial factors

\[ P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t, l)P(x|a), P(d|a, b) \]

\[ \Rightarrow f_v(t)P(s)P(l|s)P(b|s)P(a|t, l)P(x|a), P(d|a, b) \]

\[ \Rightarrow f_v(t)f_s(b, l)P(a|t, l)P(x|a), P(d|a, b) \]

\[ \Rightarrow f_v(t)f_s(b, l)f_x(a)P(a|t, l)P(d|a, b) \]

\[ \Rightarrow f_s(b, l)f_x(a)f_t(a, l)P(d|a, b) \]

\[ \Rightarrow f_x(a)f_t(b, a)P(d|a, b) \]

Eliminate: \( a, b \)

Compute:

\[ f_a(b, d) = \sum_a f_x(a)f_t(b, a)P(d|a, b) \]

\[ f_b(d) = \sum_b f_a(b, d) \]

Variable Elimination (VE)

- We now understand VE as a sequence of rewriting operations
- Actual computation is done in elimination step
- Exactly the same procedure applies to Markov networks
- Computation depends on order of elimination

Dealing with Evidence

- How do we deal with evidence?
- Suppose get evidence \( V = t, S = f, D = t \)
- We want to compute \( P(L, V = t, S = f, D = t) \)

Dealing with Evidence

- We start by writing the factors:

\[ P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t, l)P(x|a), P(d|a, b) \]
We start by writing the factors:

\[ P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a), P(d|a,b) \]

Since we know that \( V = t \), we don’t need to eliminate \( V \)

Instead, we can replace the factors \( P(V) \) and \( P(T|V) \) with

\[ f_{P(v)} = P(v = t) \quad f_{P(t|v)}(t) = P(t|v = t) \]

This "selects" the appropriate parts of the original factors given the evidence

Note that \( f_{P(v)} \) is a constant, and thus does not appear in elimination of other variables.

Initial factors, after setting evidence

\[ f_{P(s)}f_{P(a)}f_{P(t|v)}(t)f_{P(l|a)}(l)f_{P(b|a)}(b)P(a|t,l)P(x|a)f_{P(d|a,b)}(a,b) \]

Eliminating \( x \), we get

\[ f_{P(v)}f_{P(s)}f_{P(t|v)}(t)f_{P(l|a)}(l)f_{P(b|a)}(b)P(a|t,l)f_{x}(a)f_{P(d|a,b)}(a,b) \]
Dealing with Evidence

- Initial factors, after setting evidence
  \[ f_{P(L)} f_{P(a)} f_{P(t|V)} f_{P(l|a)} f_{P(b)} P(a|t, l) P(x|a) f_{P(d|a,b)} (a, b) \]
- Eliminating \( x, t, a, \) we get
  \[ f_{P(V)} f_{P(a)} f_{a}(b, l) f_{P(l|a)} f_{P(b|a)} (b) \]
- Finally, when eliminating \( b, \) we get

Given evidence \( V = t, S = f, D = t, \) compute \( P(L, V = t, S = f, D = t) \)
Outline

- Introduction
- Reminder: Probability theory
- Basics of Bayesian Networks
- Modeling Bayesian networks
- Inference (VE, Junction Tree)
- Excourse: Markov Networks
- Learning Bayesian networks
- Relational Models