Reasoning Under Uncertainty: Variable Elimination

CPSC 322 – Uncertainty 6

Textbook §6.4
Lecture Overview

1 Recap

2 Factors

3 Variable Elimination
alarm and report are independent: false.

alarm and report are independent given leaving: true.

Intuitively, the only way that the alarm affects report is by affecting leaving.
Common ancestors

- \textit{alarm} and \textit{smoke} are independent: false.
- \textit{alarm} and \textit{smoke} are independent given \textit{fire}: true.
- Intuitively, \textit{fire} can explain \textit{alarm} and \textit{smoke}; learning one can affect the other by changing your belief in \textit{fire}.
Common descendants

- $\text{tampering}$ and $\text{fire}$ are independent: true.
- $\text{tampering}$ and $\text{fire}$ are independent given $\text{alarm}$: false.
- Intuitively, $\text{tampering}$ can explain away $\text{fire}$
Belief Network Inference

- **Our goal:** compute probabilities of variables in a belief network
- **Two cases:**
  1. the unconditional (prior) distribution over one or more variables
  2. the posterior distribution over one or more variables, conditioned on one or more observed variables
- **To address both cases,** we only need a computational solution to case 1
- **Our method:** exploiting the structure of the network to efficiently eliminate (sum out) the non-observed, non-query variables one at a time.
# Lecture Overview

1. Recap

2. Factors

3. Variable Elimination
Factors

- A **factor** is a representation of a function from a tuple of random variables into a number.

- We will write factor $f$ on variables $X_1, \ldots, X_j$ as $f(X_1, \ldots, X_j)$.

- A factor denotes a distribution over the given tuple of variables in some (unspecified) context
  - e.g., $P(X_1, X_2)$ is a factor $f(X_1, X_2)$
  - e.g., $P(X_1, X_2, X_3 = v_3)$ is a factor $f(X_1, X_2)$
  - e.g., $P(X_1, X_3 = v_3 | X_2)$ is a factor $f(X_1, X_2)$
Manipulating Factors

- We can make new factors out of an existing factor.
- Our first operation: we can assign some or all of the variables of a factor.
  - $f(X_1 = v_1, X_2, \ldots, X_j)$, where $v_1 \in \text{dom}(X_1)$, is a factor on $X_2, \ldots, X_j$.
  - $f(X_1 = v_1, X_2 = v_2, \ldots, X_j = v_j)$ is a number that is the value of $f$ when each $X_i$ has value $v_i$.
- The former is also written as $f(X_1, X_2, \ldots, X_j)_{X_1 = v_1, \ldots, X_j = v_j}$.
Example factors

\[ r(X, Y, Z) : \]
\[
\begin{array}{ccc|c}
X & Y & Z & \text{val} \\
\hline
\text{t} & \text{t} & \text{t} & 0.1 \\
\text{t} & \text{t} & \text{f} & 0.9 \\
\text{t} & \text{f} & \text{t} & 0.2 \\
\text{t} & \text{f} & \text{f} & 0.8 \\
\text{f} & \text{t} & \text{t} & 0.4 \\
\text{f} & \text{t} & \text{f} & 0.6 \\
\text{f} & \text{f} & \text{t} & 0.3 \\
\text{f} & \text{f} & \text{f} & 0.7 \\
\end{array}
\]

\[ r(X = \text{t}, Y, Z) : \]
\[
\begin{array}{cc|c}
Y & Z & \text{val} \\
\hline
\text{t} & \text{t} & 0.1 \\
\text{t} & \text{f} & 0.9 \\
\text{f} & \text{t} & 0.2 \\
\text{f} & \text{f} & 0.8 \\
\end{array}
\]

\[ r(X = \text{t}, Y, Z = \text{f}) = 0.8 \]

\[ r(X = \text{t}, Y = \text{f}, Z = \text{f}) = 0.8 \]
Summing out variables

Our second operation: we can **sum out** a variable, say $X_1$ with domain $\{v_1, \ldots, v_k\}$, from factor $f(X_1, \ldots, X_j)$, resulting in a factor on $X_2, \ldots, X_j$ defined by:

$$
(\sum_{X_1} f)(X_2, \ldots, X_j) = f(X_1 = v_1, \ldots, X_j) + \cdots + f(X_1 = v_k, \ldots, X_j)
$$
Recap Factors
Variable Elimination

**Summing out a variable example**

<table>
<thead>
<tr>
<th>( A )</th>
<th>( B )</th>
<th>( C )</th>
<th>val</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>t</td>
<td>t</td>
<td>0.03</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>f</td>
<td>0.07</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>t</td>
<td>0.54</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>t</td>
<td>0.06</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>f</td>
<td>0.14</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>t</td>
<td>0.48</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>f</td>
<td>0.32</td>
</tr>
</tbody>
</table>

\( f_3: \)

\[ \sum_B f_3: \]

<table>
<thead>
<tr>
<th>( A )</th>
<th>( C )</th>
<th>val</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>t</td>
<td>0.57</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>0.43</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>0.54</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>0.46</td>
</tr>
</tbody>
</table>
Multiplying factors

- Our third operation: factors can be multiplied together.
- The **product** of factor $f_1(X, Y)$ and $f_2(Y, Z)$, where $Y$ are the variables in common, is the factor $(f_1 \times f_2)(X, Y, Z)$ defined by:

\[ (f_1 \times f_2)(X, Y, Z) = f_1(X, Y)f_2(Y, Z). \]

- Note: it’s defined on all $X, Y, Z$ *triples*, obtained by multiplying together the appropriate pair of entries from $f_1$ and $f_2$. 
## Multiplying factors example

### Recap

Factors Variable Elimination

### Multiplying factors example

#### $f_1$:

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>val</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>t</td>
<td>0.1</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>0.9</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>0.2</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>0.8</td>
</tr>
</tbody>
</table>

#### $f_2$:

<table>
<thead>
<tr>
<th>$B$</th>
<th>$C$</th>
<th>val</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>t</td>
<td>0.3</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>0.7</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>0.6</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>0.4</td>
</tr>
</tbody>
</table>

#### $f_1 \times f_2$:

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>val</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>t</td>
<td>t</td>
<td>0.03</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>f</td>
<td>0.07</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>t</td>
<td>0.54</td>
</tr>
<tr>
<td>t</td>
<td>f</td>
<td>f</td>
<td>0.36</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>t</td>
<td>0.06</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>f</td>
<td>0.14</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>t</td>
<td>0.48</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>f</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Lecture Overview

1. Recap
2. Factors
3. Variable Elimination
Recap

Factors

Variable Elimination

Probability of a conjunction

- Suppose the variables of the belief network are $X_1, \ldots, X_n$.
- What we want to compute: the factor
  \[ P(X_q, X_{o_1} = v_1, \ldots, X_{o_j} = v_j) \]
- We can compute $P(X_q, X_{o_1} = v_1, \ldots, X_{o_j} = v_j)$ by summing out the variables
  \[ X_{s_1}, \ldots, X_{s_k} = \{X_1, \ldots, X_n\} \setminus \{X_q, X_{o_1}, \ldots, X_{o_j}\}. \]
- We sum out these variables one at a time
  - the order in which we do this is called our elimination ordering.

\[
P(X_q, X_{o_1} = v_1, \ldots, X_{o_j} = v_j) = \sum_{X_{s_k}} \cdots \sum_{X_{s_1}} P(X_1, \ldots, X_n) X_{o_1} = v_1, \ldots, X_{o_j} = v_j.
\]
Recap

Factors

Variable Elimination

Probability of a conjunction

- What we know: the factors $P(X_i | pX_i)$.
- Using the chain rule and the definition of a belief network, we can write $P(X_1, \ldots, X_n)$ as $\prod_{i=1}^{n} P(X_i | pX_i)$. Thus:

$$P(X_q, X_{o1} = v_1, \ldots, X_{oj} = v_j)$$

$$= \sum_{X_{s_k}} \cdots \sum_{X_{s_1}} P(X_1, \ldots, X_n)_{X_{o1} = v_1, \ldots, X_{oj} = v_j}.$$ 

$$= \sum_{X_{s_k}} \cdots \sum_{X_{s_1}} \prod_{i=1}^{n} P(X_i | pX_i)_{X_{o1} = v_1, \ldots, X_{oj} = v_j}.$$
Computing sums of products

Computation in belief networks thus reduces to computing the sums of products.

- It takes 14 multiplications or additions to evaluate the expression $ab + ac + ad + aeh + afh + agh$. How can this expression be evaluated more efficiently?
Computing sums of products

Computation in belief networks thus reduces to computing the sums of products.

- It takes 14 multiplications or additions to evaluate the expression $ab + ac + ad + aeh + afh + agh$. How can this expression be evaluated more efficiently?
  - factor out the $a$ and then the $h$ giving $a(b + c + d + h(e + f + g))$
  - this takes only 7 multiplications or additions
Computing sums of products

Computation in belief networks thus reduces to computing the sums of products.

- It takes 14 multiplications or additions to evaluate the expression $ab + ac + ad + aeh + afh + agh$. How can this expression be evaluated more efficiently?
  - factor out the $a$ and then the $h$ giving $a(b + c + d + h(e + f + g))$
  - this takes only 7 multiplications or additions

- How can we compute $\sum_{X_{s1}} \prod_{i=1}^{n} P(X_i|pX_i)$ efficiently?
Computing sums of products

Computation in belief networks thus reduces to computing the sums of products.

- It takes 14 multiplications or additions to evaluate the expression \( ab + ac + ad + aeh + afh + agh \). How can this expression be evaluated more efficiently?
  - factor out the \( a \) and then the \( h \) giving \( a(b + c + d + h(e + f + g)) \)
  - this takes only 7 multiplications or additions

- How can we compute \( \sum_{X_{s_1}} \prod_{i=1}^{n} P(X_i|pX_i) \) efficiently?
  - Factor out those terms that don’t involve \( X_{s_1} \):

\[
\left( \prod_{i|X_{s_1} \not\in \{X_i\} \cup pX_i} P(X_i|pX_i) \right) \left( \sum_{X_{s_1}} \prod_{i|X_{s_1} \in \{X_i\} \cup pX_i} P(X_i|pX_i) \right)
\]

(terms that do not involve \( X_{s_i} \))

(terms that involve \( X_{s_i} \))