Last Time: Factor Zoo

Doan luk at meh! How shud ai noz?

Prepare for

Otter annhialation!
Last Time: VE

$P(R)$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$+r$</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>$-r$</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

$P(T|R)$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$+r$</td>
<td>$+t$</td>
<td>0.8</td>
</tr>
<tr>
<td>$+r$</td>
<td>$-t$</td>
<td>0.2</td>
</tr>
<tr>
<td>$-r$</td>
<td>$+t$</td>
<td>0.1</td>
</tr>
<tr>
<td>$-r$</td>
<td>$-t$</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Join R

$P(R,T)$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$+r$</td>
<td>$+t$</td>
<td>0.08</td>
</tr>
<tr>
<td>$+r$</td>
<td>$-t$</td>
<td>0.02</td>
</tr>
<tr>
<td>$-r$</td>
<td>$+t$</td>
<td>0.09</td>
</tr>
<tr>
<td>$-r$</td>
<td>$-t$</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Sum out R

$P(T)$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$+t$</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>$-t$</td>
<td>0.83</td>
<td></td>
</tr>
</tbody>
</table>

$P(L|T)$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$+t$</td>
<td>$+l$</td>
<td>0.3</td>
</tr>
<tr>
<td>$+t$</td>
<td>$-l$</td>
<td>0.7</td>
</tr>
<tr>
<td>$-t$</td>
<td>$+l$</td>
<td>0.1</td>
</tr>
<tr>
<td>$-t$</td>
<td>$-l$</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Join R

$P(R,T)$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$+t$</td>
<td>$+l$</td>
<td>0.3</td>
</tr>
<tr>
<td>$+t$</td>
<td>$-l$</td>
<td>0.7</td>
</tr>
<tr>
<td>$-t$</td>
<td>$+l$</td>
<td>0.1</td>
</tr>
<tr>
<td>$-t$</td>
<td>$-l$</td>
<td>0.9</td>
</tr>
</tbody>
</table>

$P(L|T)$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$+t$</td>
<td>$+l$</td>
<td>0.3</td>
</tr>
<tr>
<td>$+t$</td>
<td>$-l$</td>
<td>0.7</td>
</tr>
<tr>
<td>$-t$</td>
<td>$+l$</td>
<td>0.1</td>
</tr>
<tr>
<td>$-t$</td>
<td>$-l$</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Last Time: VE

\[
P(T) = \begin{align*}
+ t & \quad 0.17 \\
- t & \quad 0.83
\end{align*}
\]

\[
P(L|T) = \begin{align*}
+ t & \quad + l & \quad 0.3 \\
+ t & \quad - l & \quad 0.7 \\
- t & \quad + l & \quad 0.1 \\
- t & \quad - l & \quad 0.9
\end{align*}
\]

Join T

\[
P(T, L) = \begin{align*}
+ t & \quad + l & \quad 0.051 \\
+ t & \quad - l & \quad 0.119 \\
- t & \quad + l & \quad 0.083 \\
- t & \quad - l & \quad 0.747
\end{align*}
\]

Sum out T

\[
P(L) = \begin{align*}
+ l & \quad 0.134 \\
- l & \quad 0.886
\end{align*}
\]
Evidence

- If evidence, start with factors that select that evidence
  - No evidence uses these initial factors:

| $P(R)$ | $P(T|R)$ | $P(L|T)$ |
|--------|----------|-----------|
| $+r$   | 0.1      | $+t$      | 0.8       | $+t$ | 0.3  |
| $-r$   | 0.9      | $-t$      | 0.2       | $+t$ | 0.7  |

- Computing $P(L|+r)$ the initial factors become:

| $P(+r)$ | $P(T|+r)$ | $P(L|T)$ |
|---------|-----------|-----------|
| $+r$    | 0.1      | $+t$      | 0.8       | $+t$ | 0.3  |
| $-r$    | $+t$     | 0.2       | $+t$ | 0.7  |

- We eliminate all vars other than query + evidence
Evidence II

• Result will be a selected joint of query and evidence
  – E.g. for $P(L \mid +r)$, we’d end up with:

\[
\begin{array}{c|c|c}
+ r & + l & 0.026 \\
+ r & - l & 0.074 \\
\end{array}
\]

Normalize

\[
\begin{array}{c|c}
+ l & 0.26 \\
- l & 0.74 \\
\end{array}
\]

• To get our answer, just normalize this!

• That’s it!
General Variable Elimination

• Query: \[ P(Q | E_1 = e_1, \ldots E_k = e_k) \]

• Start with initial factors:
  – Local CPTs (but instantiated by evidence)

• While there are still hidden variables (not Q or evidence):
  – Pick a hidden variable H
  – Join all factors mentioning H
  – Eliminate (sum out) H

• Join all remaining factors and normalize
Variable Elimination Bayes Rule

\[
P(B)
\]

\[
\begin{array}{|c|c|}
\hline
B & P \\
\hline
+b & 0.1 \\
\hline
\neg b & 0.9 \\
\hline
\end{array}
\]

\[
P(A|B) \rightarrow P(a|B)
\]

\[
\begin{array}{|c|c|c|}
\hline
B & A & P \\
\hline
+b & +a & 0.8 \\
\hline
+b & \neg a & 0.2 \\
\hline
\neg b & +a & 0.1 \\
\hline
\neg b & a & 0.9 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
A & B & P \\
\hline
+a & +b & 0.08 \\
\hline
+a & \neg b & 0.09 \\
\hline
\end{array}
\]

\[
P(B|a)
\]

\[
\begin{array}{|c|c|c|}
\hline
A & B & P \\
\hline
+a & +b & 8/17 \\
\hline
+a & \neg b & 9/17 \\
\hline
\end{array}
\]
14.15) \( P(B \mid +j, +m) = ? \)

**Enumeration?**

*Sum over what I don’t know*

**VE?**

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>(P(A \mid B, E))</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>+e</td>
<td>0.95</td>
</tr>
<tr>
<td>+b</td>
<td>−e</td>
<td>0.94</td>
</tr>
<tr>
<td>−b</td>
<td>+e</td>
<td>0.29</td>
</tr>
<tr>
<td>−b</td>
<td>−e</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Joint E**

**Eliminate E**

**Join B**

**Normalize**

<table>
<thead>
<tr>
<th>A</th>
<th>(P(J \mid A))</th>
</tr>
</thead>
<tbody>
<tr>
<td>+a</td>
<td>0.9</td>
</tr>
<tr>
<td>−a</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>(P(M \mid A))</th>
</tr>
</thead>
<tbody>
<tr>
<td>+a</td>
<td>0.7</td>
</tr>
<tr>
<td>−a</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Example

\[ P(B|j, m) \propto P(B, j, m) \]

\[
\begin{array}{ccccc}
P(B) & P(E) & P(A|B, E) & P(j|A) & P(m|A) \\
\end{array}
\]

Choose A

\[
\begin{array}{cccc}
P(A|B, E) & P(j|A) & P(m|A) & P(j, m, A|B, E) \\
\end{array}
\]

\[
\sum P(j, m|B, E)
\]

\[
\begin{array}{ccc}
P(B) & P(E) & P(j, m|B, E) \\
\end{array}
\]
Example

\[
P(B) \quad P(E) \quad P(j, m|B, E)
\]

Choose E

\[
P(E) \quad P(j, m, E|B) \quad \sum \quad P(j, m|B)
\]

\[
P(B) \quad P(j, m|B)
\]

Finish with B

\[
P(B) \quad P(j, m, B) \quad \text{Normalize} \quad P(B|j, m)
\]
Variable Elimination

• What you need to know:
  – Should be able to run it on small examples, understand the factor creation / reduction flow
  – Better than enumeration: saves time by marginalizing variables as soon as possible rather than at the end

• We will see special cases of VE later
  – On tree-structured graphs, variable elimination is poly. time
  – Implement a tree-structured special case to track invisible ghosts
Approximate Inference

• Simulation has a name: sampling

• Sampling is a hot topic in machine learning, and it’s really simple

• Basic idea:
  – Draw $N$ samples from a sampling distribution $S$
  – Compute an approximate posterior probability
  – Show this converges to the true probability $P$

• Why sample?
  – Learning: get samples from a distribution you don’t know
  – Inference: getting a sample is faster than computing the right answer (e.g., with variable elimination)
Prior Sampling

$P(C)$

<table>
<thead>
<tr>
<th></th>
<th>+c</th>
<th>-c</th>
</tr>
</thead>
<tbody>
<tr>
<td>+c</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>-c</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

$P(S|C)$

<table>
<thead>
<tr>
<th></th>
<th>+c</th>
<th>+s</th>
<th>-s</th>
<th>-c</th>
<th>+s</th>
<th>-s</th>
</tr>
</thead>
<tbody>
<tr>
<td>+c</td>
<td></td>
<td>0.1</td>
<td></td>
<td></td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>+s</td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

$P(R|C)$

<table>
<thead>
<tr>
<th></th>
<th>+c</th>
<th>+r</th>
<th>-r</th>
<th>-c</th>
<th>+r</th>
<th>-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>+c</td>
<td></td>
<td>0.8</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+r</td>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$P(W|S, R)$

<table>
<thead>
<tr>
<th></th>
<th>+s</th>
<th>+r</th>
<th>+w</th>
<th>-w</th>
<th>+s</th>
<th>+r</th>
<th>+w</th>
<th>-w</th>
<th>+s</th>
<th>+r</th>
<th>+w</th>
<th>-w</th>
</tr>
</thead>
<tbody>
<tr>
<td>+s</td>
<td></td>
<td></td>
<td>0.99</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td></td>
<td></td>
<td>0.90</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-s</td>
<td></td>
<td></td>
<td>0.90</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Samples:

+ c, -s, +r, +w
- c, +s, -r, +w
...

14
Prior Sampling

- This process generates samples with probability:

\[ S_{PS}(x_1 \ldots x_n) = \prod_{i=1}^{n} P(x_i|\text{Parents}(X_i)) = P(x_1 \ldots x_n) \]

...i.e. the BN’s joint probability

- Let the number of samples of an event be \( N_{PS}(x_1 \ldots x_n) \)

- Then \( \lim_{N \to \infty} \hat{P}(x_1, \ldots, x_n) = \lim_{N \to \infty} N_{PS}(x_1, \ldots, x_n)/N = S_{PS}(x_1, \ldots, x_n) = P(x_1 \ldots x_n) \)

- i.e., the sampling procedure is consistent
Example

• We’ll get a bunch of samples from the BN:
  - +c, -s, +r, +w
  - +c, +s, +r, +w
  - -c, +s, +r, -w
  - +c, -s, +r, +w
  - -c, -s, -r, +w

• If we want to know P(W)
  - We have counts <+w:4, -w:1>
  - Normalize to get P(W) = <+w:0.8, -w:0.2>
  - This will get closer to the true distribution with more samples
  - Can estimate anything else, too
  - What about P(C| +w)?  P(C| +r, +w)?  P(C| -r, -w)?
  - Fast: can use fewer samples if less time (what’s the drawback?)
Rejection Sampling

• Let’s say we want \( P(C) \)
  – No point keeping all samples around
  – Just tally counts of \( C \) as we go

• Let’s say we want \( P(C| +s) \)
  – Same thing: tally \( C \) outcomes, but ignore (reject) samples which don’t have \( S=+s \)
  – This is called rejection sampling
  – It is also consistent for conditional probabilities (i.e., correct in the limit)