Bayesian Networks 2016–2017

Solutions to Tutorial III – Probabilistic inference

Note: To compute the probability distribution $P(V)$ you need to compute the probabilities for all the values of the variable V. For example, if V is binary then $P(v)$ and $P(\neg v)$ must be computed. Of course, computing one of them is sufficient to derive the other as they sum up to 1.

Answers to Exercise 1

a.
$$P^*(v_2) = \alpha \cdot \pi(v_2) \cdot \lambda(v_2) \ \ (data\ fusion\ lemma) \tag{1}$$

$$\lambda(v_2) = 1 \ \ (initialisation)$$

$$\pi(v_2) = P(v_2 \mid v_1) \cdot \pi_{V_2}^{V_1}(v_1) + P(v_2 \mid \neg v_1) \cdot \pi_{V_2}^{V_1}(\neg v_1). \tag{2}$$

Since V_1 does not have parents then $\pi_{V_2}^{V_1}(v_1) = P(v_1)$ and $\pi_{V_2}^{V_1}(\neg v_1) = P(\neg v_1)$. Hence, substituting with the numbers in Equations 1 and 2, respectively gives $P^*(v_2) = \alpha \cdot 0.66$. Analogous computation for $P^*(\neg v_2)$ is made and $P^*(\neg v_2) = \alpha \cdot 0.34$. Hence, $\alpha = 1$.

b. Since we have evidence that $V_1 = \textit{true}$ then $P^*(v_1) = 1$ and $P^*(\neg v_1) = 0$.

For V_2, we have as in 1a that $\lambda(v_2) = 1$ and:

$$\begin{aligned}
P^*(v_2) &= \alpha \cdot \lambda(v_2) \cdot \pi(v_2) \\
&= \alpha \cdot 1 \cdot \left(P(v_2 \mid v_1) \cdot \pi_{V_2}^{V_1}(v_1) + P(v_2 \mid \neg v_1) \cdot \pi_{V_2}^{V_1}(\neg v_1) \right) \\
&= \alpha \cdot (P(v_2 \mid v_1) \cdot P^*(v_1) + P(v_2 \mid \neg v_1) \cdot P^*(\neg v_1)) \\
&= \alpha \cdot P(v_2 \mid v_1) = \alpha \cdot 0.3.
\end{aligned}$$

Analogously, $P^*(\neg v_2) = \alpha \cdot P(\neg v_2 \mid v_1) = \alpha \cdot 0.7$ and $\alpha = 1$.

Finally for V_3, we have $\lambda(v_3) = 1$ and

$$\begin{aligned}
P^*(v_3) &= \alpha \cdot \lambda(v_3) \cdot \pi(v_3) \\
&= \alpha \cdot 1 \cdot \left(P(v_3 \mid v_2) \cdot \pi_{V_3}^{V_2}(v_2) + P(v_3 \mid \neg v_2) \cdot \pi_{V_3}^{V_2}(\neg v_2) \right) \\
&= \alpha \cdot (P(v_3 \mid v_2) \cdot P^*(v_2) + P(v_3 \mid \neg v_2) \cdot P^*(\neg v_2)) \\
&= \alpha \cdot (0.7 \cdot 0.3 + 0.1 \cdot 0.7) = \alpha \cdot 0.28
\end{aligned}$$

Analogously, $P^*(\neg v_3) = \alpha \cdot 0.72$

c. Since we have that $V_1 = \textit{true}$ and $V_3 = \textit{false}$ then $P^*(v_1) = 1$ and $P^*(\neg v_1) = 0$, $P^*(v_3) = 0$ and $P^*(\neg v_3) = 1$.

   For V_2 we compute

$$
\begin{aligned}
P^*(v_2) &= \alpha \cdot \lambda(v_2) \cdot \pi(v_2) \\
&= \alpha \cdot \lambda_{V_3}^{V_2}(v_2) \cdot \textstyle\sum_{V_1} P(v_2 \mid V_1)\pi_{V_2}^{V_1}(V_1)
\end{aligned}
$$

   Now, $\lambda_{V_3}^{V_2}(v_2) = \beta \sum_{V_3} \lambda(V_3) P(V_3 \mid v_2) \cdot 1 = \beta P(\neg v_3 \mid v_2)$ as $\lambda(v_3) = 0$ and $\lambda(\neg v_3) = 1$ due to the evidence $V_3 = \textit{false}$; in addition, we do not have other parents than V_2. (So in the formula $\mathrm{pa}(V_3) = V_2$, whereas the summation Σ excludes V_2 and becomes empty; as we substitute the value true for V_2 (a parameter of $\lambda_{V_3}^{V_2}(V_2)$), we get v_2.) The β is simply absorbed into α. Thus,

$$P^*(v_2) = \alpha \cdot P(\neg v_3 \mid v_2) \cdot P(v_2 \mid v_1) = \alpha \cdot 0.09$$

Analogously, $P^*(\neg v_2) = \alpha \cdot P(\neg v_3 \mid \neg v_2) \cdot P(\neg v_2 \mid v_1) = \alpha \cdot 0.63$

Answers to Exercise 2

a. For V_4, we have that $\lambda(v_4) = 1$ and:

$$
\begin{aligned}
P^*(v_4) &= \alpha \cdot \lambda(v_4) \cdot \pi(v_4) \\
&= \alpha \cdot 1 \cdot \Big(P(v_4 \mid v_3) \cdot \pi_{V_4}^{V_3}(v_3) + P(v_4 \mid \neg v_3) \cdot \pi_{V_4}^{V_3}(\neg v_3) \Big)
\end{aligned} \tag{3}
$$

   Then we compute:

$$
\begin{aligned}
\pi_{V_4}^{V_3}(v_3) = \pi(v_3) &= P(v_3 \mid v_1, v_2)\pi_{V_3}^{V_1}(v_1)\pi_{V_3}^{V_2}(v_2) \\
&\quad + P(v_3 \mid \neg v_1, v_2)\pi_{V_3}^{V_1}(\neg v_1)\pi_{V_3}^{V_2}(v_2) \\
&\quad + P(v_3 \mid v_1, \neg v_2)\pi_{V_3}^{V_1}(v_1)\pi_{V_3}^{V_2}(\neg v_2) \\
&\quad + P(v_3 \mid \neg v_1, \neg v_2)\pi_{V_3}^{V_1}(\neg v1)\pi_{V_3}^{V_2}(\neg v_2) \\
&= P(v_3 \mid v_1, v_2)P(v_1)P(v_2) \\
&\quad + P(v_3 \mid \neg v_1, v_2)P(\neg v_1)P(v_2) \\
&\quad + P(v_3 \mid v_1, \neg v_2)P(v_1)P(\neg v_2) \\
&\quad + P(v_3 \mid \neg v_1, \neg v_2)P(\neg v_1)P(\neg v_2) \\
&= 0.244
\end{aligned}
$$

   Analogously we compute $\pi_{V_4}^{V_3}(\neg v_3) = 0.76$. Substituting in Equation 3 gives us:

$$P^*(v_4) = \alpha \cdot (0.3 \cdot 0.244 + 0.5 \cdot 0.76) = \alpha \cdot 0.4532$$

   Following the same computational procedure you can compute $P^*(\neg v_4)$.

b. For $P^*(v_4)$ and $P^*(\neg v_4)$, it now holds that $\pi_{V_4}^{V_3}(v_3) = 1$ and $\pi_{V_4}^{V_3}(\neg v_3) = 0$. So $P^*(v_4) = P(v_4 \mid v_3)$.

For V_1 we have that $\pi(V_1) = P(V_1)$ and $\lambda(V_1) = \lambda^{V_1}_{V3}(V_1)$. It holds that:

$$\lambda^{V_1}_{V_3}(V_1) = \sum_{V_3} \lambda(V_3) \sum_{V_2} P(V_3 \mid V_2, V_1) \pi^{V_2}_{V_3}(V_2)$$

Since $\lambda(v_3) = 1, \lambda(\neg v_3) = 0$ and $\pi^{V_2}_{V_3}(v_2) = 1, \pi^{V_2}_{V_3}(\neg v_2) = 0$, we get $\lambda^{V_1}_{V_3}(v_1) = P(v_3 \mid v_2, v_1)$ and $\lambda^{V_1}_{V_3}(\neg v_1) = P(v_3 \mid v_2, \neg v_1)$. From this it follows: $P^*(v_1) = \alpha \cdot P(v_1) \cdot P(v_3 \mid v_2, v_1) = 0.04\alpha \approx 0.18$ and $P^*(\neg v_1) = \alpha \cdot P(\neg v_1) \cdot P(v_3 \mid v_2, \neg v_1) = 0.18\alpha \approx 0.82$.

Answer to Exercise 3

a. V_1 and V_2 (not V_3 as this would create a dependence between V_1 and V_2).

b. $P^*(v_2) = 0.2$ and $P^*(\neg v_2) = 0.8$. We could also remove $V_1 \to V_3$; however, this is not necessary for the algorithm.

c. After deleting arc $V_1 \to V_2$, we obtain a probability distribution that factorises as:

$$P'(V) = P'(V_3 \mid V_1, V_2) P'(V_2) P'(V_1)$$

If V_1 is assumed to be true for V_2, it holds that $P'(V_2) = P(V_2 \mid v_1)$. Suppose we compute $P^*(v_3)$ in this instantiated network:

$$\begin{aligned} P^*(v_3) &= \sum_{V_1,V_2} P'(v_3 \mid V_1, V_2) P'(V_2) P'(V_1) \\ &= \sum_{V_1,V_2} P(v_3 \mid V_1, V_2) P(V_2 \mid v_1) P(V_1) \end{aligned}$$

Note that this sum contains inconsistent terms such as $P(v_3 \mid \neg v_1, v_2) P(v_2 \mid v_1) P(\neg v_1)$ so it is certainly not equal to $P(v_3 \mid v_1)$. However, if we compute $P'(v_1, v_3)$ we obtain:

$$\begin{aligned} P'(v_1, v_3) &= \sum_{V_2} P'(v_3 \mid v_1, V_2) P'(V_2) P'(v_1) \\ &= \sum_{V_2} P(v_3 \mid v_1, V_2) P(V_2 \mid v_1) P(v_1) \\ &= P(v_1, v_3) \end{aligned}$$

In fact, it holds in general that $P'(V_i, V_e) = P(V_i, V_e)$ where V_i is some variable and V_e is the node that is being instantiated in the network.

d. $P(v_3) = P'(v_1, v_3) + P'(\neg v_1, v_3)$.

Answer to Exercise 4

a. After moralisation, you obtain two cliques: $\{V_1, V_2, V_3\}$ and $\{V_3, V_4\}$.

b. Given these cliques, we have potentials for each clique (φ_{123} and φ_{34}) and one for the separator φ_3. Before message passing it holds that:

$$\begin{aligned}\varphi_{123}(V_1, V_2, V_3) &= P(V_3 \mid V_1, V_2)P(V_1)P(V_2) \\ \varphi_{34}(V_3, V_4) &= P(V_4 \mid V_3) \\ \varphi_3(V_3) &= 1\end{aligned}$$

After message passing, we will obtain the marginals:

$$\begin{aligned}\varphi_{123}(V_1, V_2, V_3) &= P(V_1, V_2, V_3) \\ \varphi_{34}(V_3, V_4) &= P(V_3, V_4) \\ \varphi_3(V_3) &= P(V_3)\end{aligned}$$

We leave it to the reader to verify that both factorisations correspond to the original Bayesian network.

c. We first pass a message from φ_{123} to the separator φ_3:

$$\varphi_3^*(v_3) = \sum_{1,2} P(v_3 \mid V_1, V_2)P(V_1)P(V_2) = 0.244$$

(and the same for $\neg v_3$). Then we can send a messages from φ_3 to φ_{34}, e.g.:

$$\varphi_{34}^*(v_3, v_4) = 0.244/1 \cdot P(v_4 \mid v_3) = 0.07$$

Now we could propagate messages backward and, for example, obtain:

$$\varphi_3^{**}(v_3) = \sum_{V_4} \varphi_{34}^*(v_3, V_4) = 0.244$$

By now we know for sure that $\varphi_3^{**}(v_3) = P(v_3)$, so $P(v_3) = 0.244$.

d. If V_1 is true, then we restrict φ_{123} to φ_{23} so that:

$$\varphi_{23}(V_2, V_3) = \varphi_{123}(v_1, V_2, V_3)$$

Now:

$$\varphi_3^*(v_3) = \sum_{V_2} P(v_3 \mid v_1, V_2)P(V_2)P(v_1) = 0.4 \cdot 0.8 \cdot 0.1 + 0.7 \cdot 0.2 \cdot 0.1 = 0.046$$

For $\neg v_3$, we have:

$$\varphi_3^*(v_3) = \sum_{V_2} P(\neg v_3 \mid v_1, V_2)P(V_2)P(v_1) = 0.6 \cdot 0.8 \cdot 0.1 + 0.3 \cdot 0.2 \cdot 0.1 = 0.054$$

So the normalisation constant is 10. So we obtain $P(v_3 \mid v_1) = 10 \cdot 0.046 = 0.46$.

Answer to Exercise 5

It suffices to show that for any X, Y, Z, if $X \perp\!\!\!\perp_{G^m} Y \mid Z$ then $X \perp\!\!\!\perp_G Y \mid Z$. Take any of such X, Y, Z and assume $X \perp\!\!\!\perp_{G^m} Y \mid Z$. Suppose $X \not\perp\!\!\!\perp_G Y \mid Z$. Then G must contain some path from X to Y that is not separated by Z. Since G^m contains the same path (as it has at a line for each arc in G), and this path is apparently separated by Z, the only possibility is that this path in G contains at least one v-structure $V_1 \rightarrow V_2 \leftarrow V_3$ with $V_2 \in Z$ and all other nodes are not in Z. However, for each of these v-structures there is a line $V_1 - V_3$ in G^m, so then there is a path from X to Y that is not blocked by Z in G^m, which contradicts the assumption that $X \perp\!\!\!\perp_{G^m} Y \mid Z$.

Answer to Exercise 6

See the notes on the junction tree algorithm.