Machine Learning II Exercise 2

Matthias Tangemann and Paul Schmidt-Barbo

June 26, 2018

1 Bayesian Networks

E1

The path $(X_1, X_4, X_2, X_6, X_7, X_5)$ is not blocked: In nodes X_4 and X_7 the edges of the path meet head-to-head, but both nodes are in the set of observed variables C. In the remaining intermediate nodes X_2 and X_6, the edges meet tail-to-tail and head-to-tail, respectively. But neither of the nodes is in C.

E2

If $C_1 := \{x_2, x_3\}$ are given, both paths from x_1 to x_4 are blocked: In the intermediate nodes x_2 and x_3 of the paths (x_1, x_2, x_4) and (x_1, x_3, x_4), the edges meet head-to-tail and both nodes are in C_1. So $x_1 \perp x_4 | (x_2, x_3)$.

If $C_2 := \{x_1, x_4\}$ are given, the path (x_2, x_4, x_3) is not blocked, because in the only intermediate node x_4, the edges meet head-to-head, but $x_4 \in C_2$. So $x_2 \perp x_3 | (x_1, x_4)$ does not hold.

E3

Let $C_2 = \{x_1, x_4\}$ be given. Then both paths from x_2 to x_3, e.g. (x_2, x_1, x_3) and (x_2, x_4, x_3), are blocked: In both intermediate nodes x_1 and x_4 the edges meet tail-to-tail, and both nodes are in C_2. So C_1. So $x_1 \perp x_4 | (x_2, x_3)$.

If $C_1 = \{x_2, x_3\}$ are given, the path (x_1, x_2, x_4) is not blocked, because in the only intermediate node x_2, the edges meet head-to-head, but $x_2 \in C_1$. So $x_1 \perp x_4 | (x_2, x_3)$ does not hold.

2 Hidden Markov Models

E4

The given Bayesian network represent the following factorization of the combined probability:

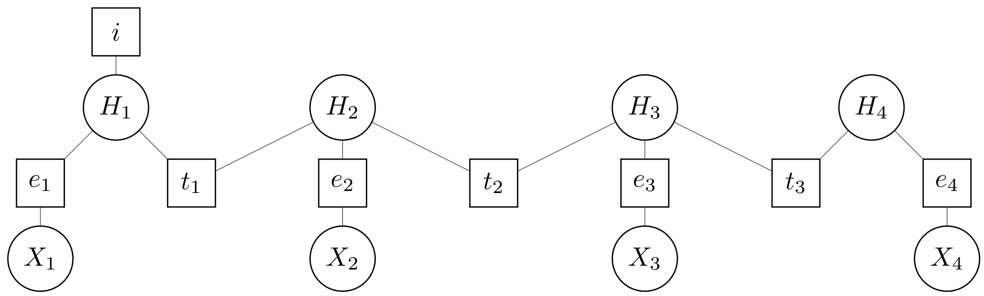

Figure 1: The factor graph corresponding to the hidden markov model

$$P(H_1, H_2, H_3, H_4, X_1, X_2, X_3, X_4) =$$
$$P(H_1)P(X_1|H_1)P(H_2|H_1)P(X_2|H_2)P(H_3|H_2)P(X_3|H_3)P(H_4|H_3)P(X_4|H_4)$$

So using the factors $i(H_1) := P(H_1)$, $e_i(X_i, H_i) := P(X_i|H_i)$ and $t_i(H_{i+1}, H_i) := P(H_{i+1}|H_i)$, the network can be transformed into the factor graph shown in figure 1.

The sum-product algorithm can be applied, because the graph is singly connected: Between each pair of nodes, there is exactly one path.

E5

The marginal distributions $P(X_i)$ can be computed using the sum-product algorithm. The variable node $P(X_1)$ is arbitrarily chosen as root. In the resulting tree, the factor node i and variable nodes X_2, X_3, X_4 are leaves. Starting from this leaves, messages are iteratively passed along the tree, until the root is reached:

$$
\begin{aligned}
\mu_{i \to H_1}(H_1) &= i(H_1) \\
\mu_{X_2 \to e_2}(X_2) &= 1 \\
\mu_{e_2 \to H_2}(H_2) &= \sum_{X_2} e_2(X_2, H_2) \\
\mu_{X_3 \to e_3}(X_3) &= 1 \\
\mu_{e_3 \to H_3}(H_3) &= \sum_{X_3} e_3(X_3, H_3) \\
\mu_{X_4 \to e_4}(X_4) &= 1 \\
\mu_{e_4 \to H_4}(X_4) &= \sum_{X_4} e_4(X_4, H_4) \\
\mu_{H_4 \to t_3}(H_4) &= \mu_{e_4 \to H_4}(X_4) \\
\mu_{t_3 \to H_3}(H_3) &= \sum_{H_4} t_3(H_4, H_3)\mu_{H_4 \to t_3}(H_4) \\
\mu_{H_3 \to t_2}(H_3) &= \mu_{e_3 \to H_3}(H_3)\mu_{t_3 \to H_3}(H_3) \\
\mu_{t_2 \to H_2}(H_2) &= \sum_{H_3} t_2(H_3, H_2)\mu_{H_3 \to t_2}(H_3) \\
\mu_{H_2 \to t_1}(H_2) &= \mu_{e_2 \to H_2}(H_2)\mu_{t_2 \to H_2}(H_2) \\
\mu_{t_1 \to H_1}(H_1) &= \sum_{H_2} t_1(H_2, H_1)\mu_{H_2 \to t_1}(H_2) \\
\mu_{H_1 \to e_1}(H_1) &= i(H_1)\mu_{t_1 \to H_1}(H_1) \\
\mu_{e_1 \to X_1}(X_1) &= \sum_{H_1} e_1(X_1, H_1)\mu_{H_1 \to e_1}(H_1)
\end{aligned}
$$

The remaining messages can now be calculated by iteratively traversing the tree from the root to the leaves:

$$
\begin{aligned}
\mu_{X_1\to e_1}(X_1) &= 1\\
\mu_{e_1\to H_1}(H_1) &= \sum_{X_1} e_1(X_1,H_1)\\
\mu_{H_1\to i}(H_1) &= \mu_{e_1\to H_1}(H_1)\mu_{t_1\to H_1}(H_1)\\
\mu_{H_1\to t_1}(H_1) &= \mu_{i\to H_1}(H_1)\mu_{e_1\to H_1}(H_1)\\
\mu_{t_1\to H_2}(H_2) &= \sum_{H_1} t_1(H_2,H_1)\mu_{H_1\to t_1}(H_1)\\
\mu_{H_2\to e_2}(H_2) &= \mu_{t_1\to H_2}(H_2)\mu_{t_2\to H_2}(H_2)\\
\mu_{e_2\to X_2}(X_2) &= \sum_{H_2} e_2(X_2,H_2)\mu_{H_2\to e_2}(H_2)\\
\mu_{H_2\to t_2}(H_2) &= \mu_{t_1\to H_2}(H_2)\mu_{e_2\to H_2}(H_2)\\
\mu_{t_2\to H_3}(H_3) &= \sum_{H_2} t_2(H_3,H_2)\mu_{H_2\to t_2}(H_2)\\
\mu_{H_3\to e_3}(H_3) &= \mu_{t_2\to H_3}(H_3)\mu_{t_3\to H_3}(H_3)\\
\mu_{e_3\to X_3}(X_3) &= \sum_{H_3} e_3(X_3,H_3)\mu_{H_3\to e_3}(H_3)\\
\mu_{H_3\to t_3}(H_3) &= \mu_{t_2\to H_3}(H_3)\mu_{e_3\to H_3}(H_3)\\
\mu_{t_3\to H_4}(H_4) &= \sum_{H_3} t_3(H_4,H_3)\mu_{H_3\to t_3}(H_3)\\
\mu_{H_4\to e_4}(H_4) &= \mu_{t_3\to H_4}(H_4)\\
\mu_{e_4\to X_4}(X_4) &= \sum_{H_4} e_4(X_4,H_4)\mu_{H_4\to e_4}(H_4)
\end{aligned}
$$

Now one message has passed in each direction across each link, and only values available at the respective step were used for the computation. The marginals can now be evaluated as $P(X_i) = \mu_{e_i\to X_i}(X_i)$. As the initial condition and the transition probabilites are symmetric with respect to values of the state variables, and due to the symmetric emission probabilities it can be easiliy seen that the marginal distributions evalutate to $P(X_i = T) = P(X_i = F) = 0.5$ for $i = 1, 2, 3, 4$.

E6

If some variables are observed, the sum-product algorithm can be adapted to compute conditional probabilites: When summing over an observed variable $V = v$, only the term including the observed value is considered, all others are ignored. Then computing the marginals yields the conditional probabilites up to a normalization factor. So starting from the leaves i, X_1, X_2, X_3, X_4 the messages are updated to reflect the observed variables. Not that in this case this equals the forward algorithm for hidden markov models.

$$
\begin{aligned}
\mu_{H_1\to t_1}(H_1|S) &= \mu_{i\to H_1}(H_1|S)\mu_{e_1\to H_1}(H_1|S)\\
&= P(H_1)P(X_1 = T|H_1)\\
&= \begin{cases} 4.0\cdot 10^{-1} & \text{if } H_1 = T\\ 1.0\cdot 10^{-1} & \text{if } H_1 = F\end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\mu_{H_2\to t_2}(H_2|S) &= \mu_{e_2\to H_2}(H_2|S)\sum_{H_1} t_1(H_2, H_1)\mu_{H_1\to t_1}(H_1|S)\\
&= P(X_2 = F|H_2)\sum_{H_1} P(H_2|H_1)\mu_{H_1\to t_1}(H_1|S)\\
&= \begin{cases} 0.2\cdot(0.7\cdot 4.0\cdot 10^{-1} + 0.3\cdot 1.0\cdot 10^{-1}) = 6.2\cdot 10^{-2} & \text{if } H_2 = T\\ 0.8\cdot(0.3\cdot 4.0\cdot 10^{-1} + 0.7\cdot 1.0\cdot 10^{-1}) = 1.52\cdot 10^{-1} & \text{if } H_2 = F\end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\mu_{H_3\to t_3}(H_3|S) &= \mu_{e_3\to H_3}(H_3|S)\sum_{H_2} t_2(H_3, H_2)\mu_{H_2\to t_2}(H_2|S)\\
&= P(X_3 = F|H_3)\sum_{H_2} P(H_3|H_2)\mu_{H_2\to t_2}(H_2|S)\\
&= \begin{cases} 0.2\cdot(0.7\cdot 6.2\cdot 10^{-2} + 0.3\cdot 1.52\cdot 10^{-1}) = 1.78\cdot 10^{-2} & \text{if } H_3 = T\\ 0.8\cdot(0.3\cdot 6.2\cdot 10^{-2} + 0.7\cdot 1.52\cdot 10^{-1}) = 1.0\cdot 10^{-1} & \text{if } H_3 = F\end{cases}
\end{aligned}
$$

Now the unnormalized probabilites for H_4 can be computed by multiplying the neighbouring factors:

$$
\begin{aligned}
\tilde{p}(H_4|S) &= \mu_{e_4\to H_4}(H_4|S)\sum_{H_3} t_3(H_4, H_3)\mu_{H_3\to t_3}(H_3|S)\\
&= P(X_4 = T|H_4)\sum_{H_3} P(H_4|H_3)\mu_{H_3\to t_3}(H_3|S)\\
&= \begin{cases} 0.8\cdot(0.7\cdot 1.78\cdot 10^{-2} + 0.3\cdot 1.0\cdot 10^{-1}) = 6.03\cdot 10^{-2} & \text{if } H_4 = T\\ 0.2\cdot(0.3\cdot 1.78\cdot 10^{-2} + 0.7\cdot 1.0\cdot 10^{-1}) = 7.8\cdot 10^{-3} & \text{if } H_4 = F\end{cases}
\end{aligned}
$$

Normalizing then gives $P(H_4 = T|S) \approx 0.885$ and $P(H_4 = F|S) \approx 0.115$.

E7

The probability distribution of $P(H_1|S)$ can be calculated analogue to the previous exercise:

$$
\begin{aligned}
\mu_{H_4\to t_3}(H_4|S) &= \mu_{e_4\to H_4}(X_4|S)\\
&= P(X_4 = T, H_4)\\
&= \begin{cases} 8.0\cdot 10^{-1} & \text{if } H_4 = T\\ 2.0\cdot 10^{-1} & \text{if } H_4 = F\end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\mu_{H_3\to t_2}(H_3|S) &= \mu_{e_3\to H_3}(H_3|S)\mu_{t_3\to H_3}(H_3|S)\\
&= P(X_3 = F, H_3)\sum_{H_4} P(H_4|H_3)\mu_{H_4\to t_3}(H_4|S)\\
&= \begin{cases} 0.2\cdot(0.7\cdot 8.0\cdot 10^{-1} + 0.3\cdot 2.0\cdot 10^{-1}) = 1.24\cdot 10^{-1} & \text{if } H_3 = T\\ 0.8\cdot(0.3\cdot 8.0\cdot 10^{-1} + 0.7\cdot 2.0\cdot 10^{-1}) = 3.04\cdot 10^{-1} & \text{if } H_3 = F\end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\mu_{H_2\to t_1}(H_2|S) &= \mu_{e_2\to H_2}(H_2|S)\mu_{t_2\to H_2}(H_2|S)\\
&= P(X_2 = F, H_2)\sum_{H_3} P(H_3|H_2)\mu_{H_3\to t_2}(H_3|S)\\
&= \begin{cases} 0.2\cdot(0.7\cdot 1.24\cdot 10^{-1} + 0.3\cdot 3.04\cdot 10^{-1}) = 3.56\cdot 10^{-2} & \text{if } H_2 = T\\ 0.8\cdot(0.3\cdot 1.24\cdot 10^{-1} + 0.7\cdot 3.04\cdot 10^{-1}) = 2.0\cdot 10^{-1} & \text{if } H_2 = F\end{cases}
\end{aligned}
$$

Now the unnormalized probabilites for H_1 can be computed by multiplying the neighbouring factors:

$$
\begin{aligned}
\tilde{p}(H_1|S) &= \mu_{i\to H_1}(H_1|S)\mu_{e_1\to H_1}(H_1|S)\sum_{H_2} t_1(H_2, H_1)\mu_{H_2\to t_1}(H_2|S)\\
&= P(H_1)P(X_1 = T|H_1)\sum_{H_2} P(H_2|H_1)\mu_{H_2\to t_1}(H_2|S)\\
&= \begin{cases} 0.5\cdot 0.8\cdot(0.7\cdot 3.56\cdot 10^{-2} + 0.3\cdot 2.0\cdot 10^{-1}) = 3.4\cdot 10^{-2} & \text{if } H_2 = T\\ 0.5\cdot 0.2\cdot(0.3\cdot 3.56\cdot 10^{-2} + 0.7\cdot 2.0\cdot 10^{-1}) = 1.51\cdot 10^{-2} & \text{if } H_2 = F\end{cases}
\end{aligned}
$$

Normalizing then gives $P(H_1 = T|S) \approx 0.692$ and $P(H_4 = F|S) \approx 0.308$.

E8

The most likely values of H_1 and H_4 can be directly determined from the conditional distributions calculated in the previous exercises. The intermediate results obtained can be used, to calculate the unnormalized probabilites of H_2 and H_3:

$$
\begin{aligned}
\tilde{p}(H_2|S) &= \mu_{t_1\to H_2}(H_2|S)\mu_{e_2\to H_2}(H_2|S)\mu_{t_2\to H_2}(H_2|S)\\
&= P(X_2 = F|H_2)\Big[\sum_{H_1} P(H_2|H_1)\mu_{H_1\to t_1}(H_1)\Big]\Big[\sum_{H_3} P(H_3|H_2)\mu_{H_3\to t_2}(H_3)\Big]\\
\Rightarrow \tilde{p}(H_2 = T|S) &= 0.2\cdot(0.7\cdot 4.0\cdot 10^{-1} + 0.3\cdot 1.0\cdot 10^{-1})(0.7\cdot 1.24\cdot 10^{-1} + 0.3\cdot 3.04\cdot 10^{-1})\\
&= 1.10\cdot 10^{-2}\\
\Rightarrow \tilde{p}(H_2 = F|S) &= 0.8\cdot(0.3\cdot 4.0\cdot 10^{-1} + 0.7\cdot 1.0\cdot 10^{-1})(0.3\cdot 1.24\cdot 10^{-1} + 0.7\cdot 3.04\cdot 10^{-1})\\
&= 3.8\cdot 10^{-2}
\end{aligned}
$$

$$
\begin{aligned}
\tilde{p}(H_3|S) &= \mu_{t_2\to H_3}(H_3|S)\mu_{e_3\to H_3}(H_3|S)\mu_{t_3\to H_3}(H_3|S)\\
&= P(X_3=F|H_3)\big[\sum_{H_2} P(H_3|H_2)\mu_{H_2\to t_2}(H_2)\big]\big[\sum_{H_4} P(H_4|H_3)\mu_{H_4\to t_3}(H_4)\big]\\
\Rightarrow \tilde{p}(H_3=T|S) &= 0.2\cdot(0.7\cdot 6.2\cdot 10^{-2}+0.3\cdot 1.52\cdot 10^{-1})(0.7\cdot 8.0\cdot 10^{-1}+0.3\cdot 2.0\cdot 10^{-1})\\
&= 1.10\cdot 10^{-2}\\
\Rightarrow \tilde{p}(H_3=F|S) &= 0.8\cdot(0.3\cdot 6.2\cdot 10^{-2}+0.7\cdot 1.52\cdot 10^{-1})(0.3\cdot 8.0\cdot 10^{-1}+0.7\cdot 2.0\cdot 10^{-1})\\
&= 3.8\cdot 10^{-2}
\end{aligned}
$$

So the most probabable sequence of the hidden variables is $(H_1=T, H_2=F, H_3=F, H_4=T)$.

E9

Since the first three elements of the observed sequence S' match that of the sequence regarded in the previous exercises S, many of the intermediate results can be reused. It is

$$
\begin{aligned}
\tilde{p}(X_4|S') &= \mu_{e_4\to X_4}\\
&= \sum_{H_4} P(X_4|H_4)\mu_{H_4\to e_4}(H_4|S')\\
&= \sum_{H_4} P(X_4|H_4)\mu_{t_3\to H_4}(H_4|S')\\
&= \sum_{H_4} P(X_4|H_4)\sum_{H_3} P(H_4|H_3)\mu_{H_3\to t_3}(H_3|S')\\
\Rightarrow \tilde{p}(X_4=T|S') &= 0.8\cdot 0.7\cdot 1.78\cdot 10^{-2}+0.8\cdot 0.3\cdot 1.0\cdot 10^{-1}+0.2\cdot 0.3\cdot 1.78\cdot 10^{-2}+0.2\cdot 0.8\cdot 1.0\cdot 10^{-1}\\
&= 4.90\cdot 10^{-2}\\
\Rightarrow \tilde{p}(X_4=T|S') &= 0.2\cdot 0.7\cdot 1.78\cdot 10^{-2}+0.2\cdot 0.3\cdot 1.0\cdot 10^{-1}+0.8\cdot 0.3\cdot 1.78\cdot 10^{-2}+0.8\cdot 0.8\cdot 1.0\cdot 10^{-1}\\
&= 6,88\cdot 10^{-2}
\end{aligned}
$$

Normalizing then gives $P(X_4=T|S')\approx 0.416$ and $P(X_4=F|S)\approx 0.584$.

E10

The independence of variables can be obtained by applying d-Independence on the original Bayesian network:

If no variable is observed, e.g. $C_1=\emptyset$, then X_1 and X_4 are not independent: In the Bayesian network, the path $(X_1, H_1, H_2, H_3, H_4, X_4)$ is not blocked, because in all intermediate nodes H_i the edges meet tail-to-tail oder head-to-tail, but none of them is in C_1.

If $C_2=H_3$ is observed, the same path is blocked: The edges meet head-to-tail in H_3, and $H_3\in C_2$. So in this case, X_1 and X_4 are independent.