Philipp Lintl
12152498
philipp.lintl@student.uva.nl

Homework Assignment 4
Machine Learning 2, 19/20

2019-09-30

I worked on it alone, Group E.

1 Problem 1

1. $$p(\boldsymbol{Z}, \boldsymbol{X}) = p(z_1) \cdot \left(\prod_{i=2}^{N} p(z_i | z_{i-1}) \right) \cdot \left(\prod_{j=1}^{N} p(x_j | z_j) \right)$$

2. Representing squares as factors and using the notation from Bishop for the factors between the z variables:
$$f_{\alpha_1}(z_1) = p(z_1), \quad f_{\alpha_i}(z_i, z_{i-1}) = p(z_i|z_{i-1}) \quad \forall i = 2, \ldots, N$$
and correspondingly for the factors between x and z
$$f_{\beta_i}(x_i, z_i) = p(x_i|z_i) \quad \forall i = 1, \ldots, N$$

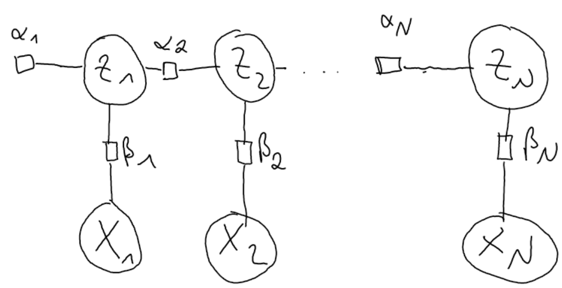

Figure 1: Factor graph for the given Markov chain

3. With notation from above and not including the normalizing factor $\frac{1}{Z}$ (as the potentials are probabilities), we end up with
$$p(\mathbf{Z}, \mathbf{X}) = f_{\alpha_1}(z_1) \cdot \prod_{i=1}^{N} f_{\beta_i}(x_i, z_i) \cdot \prod_{j=2}^{N} f_{\alpha_j}(z_j, z_{j-1})$$

4. At first, we want to express the given term with regards to $\alpha(z_n)$ and $\beta(z_n)$:
$$\begin{aligned}
p(z_n|\mathbf{X}) &= \frac{p(\mathbf{X}|z_n)\, p(z_n)}{p(\mathbf{X})} \\
&= \frac{p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N | \mathbf{z}_n)\, p(\mathbf{z}_n)}{p(\mathbf{X})} \\
(z_n \text{ d-separates } x_{n+1} \text{ and } x_n) \qquad &= \frac{p(\mathbf{x}_1, \ldots \mathbf{x}_n | \mathbf{z}_n)\, p(\mathbf{x}_{n+1}, \ldots \mathbf{x}_N | \mathbf{z}_n)\, p(\mathbf{z}_n)}{p(\mathbf{X})} \\
(\text{Bayes theorem}) \qquad &= \frac{p(\mathbf{x}_1, \ldots \mathbf{x}_n, \mathbf{z}_n)\, p(\mathbf{x}_{n+1}, \ldots \mathbf{x}_N | \mathbf{z}_n)}{p(\mathbf{X})}
\end{aligned}$$

Now, we have the sought form:

$$p(z_n|\mathbf{X}) = \frac{\alpha\left(z_n\right)\beta\left(z_n\right)}{p(\mathbf{X})}$$

Thus, $\alpha(z_n) \doteq p(x_1, \ldots, x_n|z_n)$ and $\beta(z_n) \doteq p(x_{n+1}, \ldots, x_N|z_n)$ We want to express $\alpha(z_n)$ in terms of $\alpha(z_{n-1})$, therefore we get z_{n-1} back in with 'reverse marginalizing' over it:

$$\begin{aligned}
\alpha\left(z_n\right) &= p\left(\mathbf{x}_1, \ldots \mathbf{x}_n, \mathbf{z}_n\right) \\
\text{(reverse marginalizing to get } z_{n-1}\text{)} \qquad &= \sum_{\mathbf{z}_{n-1}} p\left(\mathbf{x}_1, \ldots \mathbf{x}_n, \mathbf{z}_n, \mathbf{z}_{n-1}\right) \\
\text{(Bayes)} \qquad &= \sum_{\mathbf{z}_{n-1}} p\left(\mathbf{x}_1, \ldots \mathbf{x}_n, \mathbf{z}_n|\mathbf{z}_{n-1}\right) p\left(\mathbf{z}_{n-1}\right) \\
\text{(*)} \qquad &= \sum_{\mathbf{z}_{n-1}} p\left(\mathbf{x}_1, \ldots \mathbf{x}_{n-1}|\mathbf{z}_{n-1}\right) p\left(\mathbf{x}_n, \mathbf{z}_n|\mathbf{z}_{n-1}\right) p\left(\mathbf{z}_{n-1}\right) \\
\text{(Bayes)} \qquad &= \sum_{\mathbf{z}_{n-1}} p\left(\mathbf{x}_1, \ldots \mathbf{x}_{n-1}, \mathbf{z}_{n-1}\right) p\left(\mathbf{x}_n, \mathbf{z}_n|\mathbf{z}_{n-1}\right) \\
\text{(plugin } \alpha(...)\text{)} \qquad &= \sum_{\mathbf{z}_{n-1}} \alpha\left(z_{n-1}\right) p\left(\mathbf{x}_{n,\mathbf{Z}_n}|\mathbf{z}_{n-1}\right)
\end{aligned}$$

(*): $\{\mathbf{x_1}, \ldots, \mathbf{x_{n-1}}\}$ is d-separated from $\{\mathbf{x_n}, \mathbf{z_n}\}$ by $\mathbf{z_{n-1}}$.

Similarly for $\beta(z_n)$, we want to express it in terms of z_{n+1}:

$$\begin{aligned}
\beta\left(z_n\right) &= p\left(\mathbf{x}_{n+1}, \ldots \mathbf{x}_N|\mathbf{z}_n\right) \\
&= \sum_{\mathbf{z}_{n+1}} p\left(\mathbf{x}_{n+1}, \ldots \mathbf{x}_N, \mathbf{z}_{n+1}|\mathbf{z}_n\right) \\
\text{(Bayes)} \qquad &= \sum_{\mathbf{z}_{n+1}} \frac{p\left(\mathbf{x}_{n+1}, \ldots \mathbf{x}_N, \mathbf{z}_{n+1}, \mathbf{z}_n\right)}{p\left(\mathbf{z}_n\right)} \\
\text{(**)} \qquad &= \sum_{\mathbf{z}_{n+1}} \frac{p\left(\mathbf{x}_{n+1}, \mathbf{z}_n|\mathbf{z}_{n+1}\right) p\left(\mathbf{x}_{n+2}, \ldots \mathbf{x}_N|\mathbf{z}_{n+1}\right) p\left(\mathbf{z}_{n+1}\right)}{p\left(\mathbf{z}_n\right)} \\
\text{(Bayes)} \qquad &= \sum_{\mathbf{z}_{n+1}} \frac{p\left(\mathbf{x}_{n+1}, \mathbf{z}_n, \mathbf{z}_{n+1}\right) p\left(\mathbf{x}_{n+2}, \ldots \mathbf{x}_N|\mathbf{z}_{n+1}\right)}{p\left(\mathbf{z}_n\right)} \\
\text{(Bayes)} \qquad &= \sum_{\mathbf{z}_{n+1}} p\left(\mathbf{x}_{n+2}, \ldots \mathbf{x}_N|\mathbf{z}_{n+1}\right) p\left(\mathbf{x}_{n+1}, \mathbf{z}_{n+1}|\mathbf{z}_n\right) \\
\text{(plugin } \beta(...)\text{)} &= \sum_{\mathbf{z}_{n+1}} \beta\left(z_{n+1}\right) p\left(\mathbf{x}_{n+1}, \mathbf{z}_{n+1}|\mathbf{z}_n\right)
\end{aligned}$$

(**): This time, given $\mathbf{z_{n+1}}$, $\{\mathbf{x_{n+2}}, \ldots, \mathbf{x_N}\}$ is d-separated from $\{\mathbf{x_{n+1}}, \mathbf{z_n}\}$ and apply Bayes theorem.

2 Problem 2

At first, we define the respective factor graph as a preparation for the sum-product algorithm.

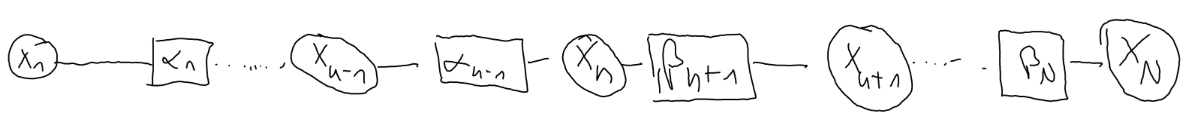

Figure 2: The respective Factor graph to the given chain of nodes model

The first step is a forward pass from the left to the root x_n on the right.

$$
\begin{aligned}
\mu_{x_1 \to \alpha_1}(x_1) &= 1 \\
\mu_{\alpha_1 \to x_2}(x_2) &= \sum_{x_1} f_{\alpha_1}(x_1, x_2)\, \mu_{x_1 \to \alpha_1}(x_1) = \sum_{x_1} f_{\alpha_1}(x_1, x_2) \\
\mu_{x_2 \to \alpha_2}(x_2) &= \mu_{\alpha_1 \to x_2}(x_2) \\
&\vdots \\
\mu_{\alpha_{n-1} \to x_n}(x_n) &= \sum_{x_{n-1}} f_{\alpha_{n-1}}(x_{n-1}, x_n)\, \mu_{x_{n-1} \to \alpha_{n-1}}(x_{n-1})
\end{aligned}
$$

This rule holds $\forall n = 2, \ldots, N$, as $\mu_{x_1 \to \alpha_1}(x_1) = 1$
If we approach the root x_n from the right and considering $\mu_{\beta_N \to N}(x_1) = 1$, we can analogously to the pass from the left obtain the recursive form for $\mu_{\beta_{n+1} \to x_n}(x_n)$:

$$
\mu_{\beta_{n+1} \to x_n}(x_n) = \sum_{x_{n+1}} f_{\beta_{n+1}}(x_{n+1}, x_n)\, \mu_{x_{n+1} \to \beta_{n+1}}(x_{n+1})
$$

This rule holds $\forall n = 1, \ldots, N-1$.

Keeping in mind the rules for n and Bishop 8.63, noting the normalization constant as Z, we can formulate:

$$
p(x_n) = \frac{1}{Z} \mu_{\alpha_{n-1} \to x_n}(x_n) \cdot \mu_{\beta_{n+1} \to x_n}(x_n)
$$

From this, we clearly see that for the right notational choice, the message passing algorithm is recovered as a special case:

$$
\begin{aligned}
\mu_\alpha(x_n) &= \mu_{\alpha_{n-1}-1 \to x_n}(x_n) \\
\mu_\beta(x_n) &= \mu_{\beta_{n+1} \to x_n}(x_n) \\
\psi_{n-1,n}(x_{n-1}, x_n) &= f_{\alpha_{n-1}}(x_{n-1}, x_n) \\
\psi_{n+1,n}(x_{n+1}, x_n) &= f_{\beta_{n+1}}(x_{n+1}, x_n)
\end{aligned}
$$

3 Problem 3

Supposing a given/observed x_N, with a value of ξ, the introduction of an indicator function $\mathbb{I}[x_N = \xi]$ will be necessary. This observations will only affect the last belief $\psi_{N-1,N}(\mathbf{x}_{N-1}, \mathbf{x}_N)$.
Then, the belief changes to:

$$
\psi_{N-1,N}(\mathbf{x}_{N-1}, \mathbf{x}_N) = \psi_{N-1,N}(\mathbf{x}_{N-1}, \mathbf{x}_N) \cdot \mathbb{I}(\mathbf{x}_N = \xi)
$$

$\mu_\alpha(x_N)$ and $\mu_\alpha(x_N)$ change, as instead of the sum, only will consist of one term, that fulfills $x_N = \xi$ and respectively, $p(z_n)$ changes, too.

Other than that, in order to obtain the conditional probability $p(\mathbf{x_n}|\mathbf{x_N})$ and in general for incorporating observed values, the passing algorithm stays the same.

4 Problem 4

Show that $p(\mathbf{x}_s) = f_s(\mathbf{x}_s) \prod_{i \in \mathrm{ne}(f_s)} \mu_{x_i \to f_s(x_i)}$ holds.

$$
\begin{aligned}
&(1) \qquad p\left(x_s\right) = \sum_{x \backslash x_s} p(x) \\
&(2) \qquad = \sum_{x \backslash x_s} f_s\left(x_s\right) \prod_{i \in nc(f_s)} \prod_{j \in \mathrm{ne}(x_i) \backslash f_s} F_j\left(x_i, X_j\right) \\
&(3) \qquad = f_s\left(x_s\right) \prod_{i \in \mathrm{ne}(f_s)} \prod_{j \in \mathrm{ne}(x_i) \backslash f_s} \left(\sum_{X_j} F_j\left(x_i, X_j\right) \right) \\
&(4) \qquad = f_s\left(x_s\right) \prod_{i \in \mathrm{ne}(f_s)} \prod_{j \in \mathrm{ne}(x_i) \backslash f_s} \mu_{f_j \rightarrow x_i}\left(x_i\right) \\
&(5) \qquad = f_s\left(x_s\right) \prod_{i \in \mathrm{ne}(f_s)} \mu_{x_i \rightarrow f_s}\left(x_i\right)
\end{aligned}
$$

With:

(1): Per definition, summing the joint distribution over all variables except for $\mathbf{x}_s$ yields the marginal $p(\mathbf{x}_s)$.

(2): $p(\mathbf{x})$ is given as the product over all factors, since we are working with a tree structured factor graph. So, we plug in:

$$
p(\mathbf{x}) = \prod_{\alpha} f_\alpha\left(\mathbf{x}_\alpha\right) = f_s\left(\mathbf{x}_s\right) \left[\prod_{i \in \mathrm{ne}(f_s)} \left(\prod_{j \in \mathrm{ne}(x_i) \backslash f_s} F_j\left(x_i, X_j\right) \right) \right]
$$

, where we used Bishop 8.65.

(3): The sum can be pushed inside, as it does not affect the other terms that are dependent on x_n.

(4): Bishop 8.64

(5): Bishop 8.69