Machine Learning 2 - Homework 4

Sindy Löwe (11594969)

April 29, 2018

Collaborators: Pascal Esser, Andrii Skliar

Problem 1

1.

$$p(\mathbf{Z}, \mathbf{X}) = p(z_1) \cdot \prod_{n=1}^{N} p(x_n|z_n) \cdot \prod_{n=2}^{N} p(z_n|z_{n-1})$$

2. The resulting factor graph can be seen in fig. 1.

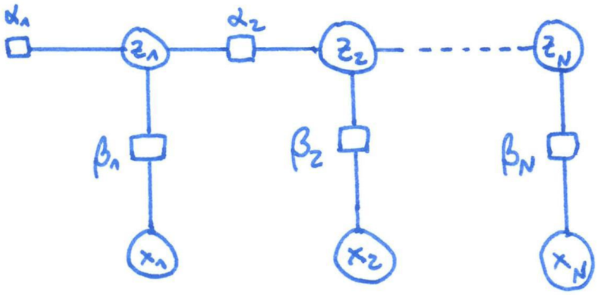

Figure 1: The resulting factor graph for task 1.2

Where $\alpha_1 = p(z_1)$, $\alpha_n = p(z_n|z_{n-1}), \forall n = 2, ..., N$ and $\beta_n = p(x_n|z_n), \forall n = 1, ..., N$.

3.

$$p(\mathbf{Z}, \mathbf{X}) = \alpha_1(z_1) \cdot \prod_{n=1}^{N} \beta_n(x_n|z_n) \cdot \prod_{n=2}^{N} \alpha_n(z_n|z_{n-1})$$

4. Recursive definition for $\alpha(z_n) = p(x_1, ..., x_n, z_n)$:

(a) Start with $n = 1$:

$$\alpha(z_1) = p(x_1, z_1) = p(x_1|z_1)p(z_1)$$

(b) For $n = 2$, we get:

$$\begin{aligned}\alpha(z_2) &= p(x_1, x_2, z_2)\\ &= \sum_{z_1} p(x_1, x_2, z_1, z_2)\\ &= \sum_{z_1} p(x_1, z_1) \cdot p(z_2|z_1) \cdot p(x_2|z_2)\\ &= \sum_{z_1} \alpha(z_1) \cdot p(z_2|z_1) \cdot p(x_2|z_2)\end{aligned}$$

(c) This gives us for general n:

$$\begin{aligned}\alpha(z_n) &= p(x_1, ..., x_n, z_n)\\ &= \sum_{z_{n-1}} p(x_1, ..., x_n, z_{n-1}, z_n)\\ &= \sum_{z_{n-1}} p(x_1, ..., x_{n-1}, z_{n-1}) \cdot p(z_n|z_{n-1}) \cdot p(x_n|z_n)\\ &= \sum_{z_{n-1}} \alpha(z_{n-1}) \cdot p(z_n|z_{n-1}) \cdot p(x_n|z_n)\end{aligned}$$

Where we use the independencies $z_{n-1} \perp\!\!\!\perp x_n|z_n$, $x_1, ..., x_{n-1} \perp\!\!\!\perp x_n|z_n$ and $x_{n-1} \perp\!\!\!\perp z_n|z_{n-1}$ inferred from the graph.

Recursive definition for $\beta_n = p(x_{n+1}, ..., x_N|z_n)$:

(a) Start with $n = N - 1$

$$\begin{aligned}\beta(z_{N-1}) &= p(x_N|z_{N-1})\\ &= \sum_{z_N} p(x_N, z_N|z_{N-1})\\ &= \sum_{z_N} p(x_N|z_N, z_{N-1}) \cdot p(z_N|z_{N-1})\end{aligned}$$

(b) For $n = N - 2$, we get:

$$\begin{aligned}\beta(z_{N-2}) &= p(x_{N-1}, x_N|z_{N-2})\\ &= \sum_{z_{N-1}} p(x_{N-1}, x_N, z_{N-1}|z_{N-2})\\ &= \sum_{z_{N-1}} p(x_{N-1}, z_{N-1}|z_{N-2}) \cdot p(x_N|x_{N-1}, z_{N-1}, z_{N-1})\\ &= \sum_{z_{N-1}} p(x_{N-1}, z_{N-1}|z_{N-2}) \cdot p(x_N|z_{N-1})\\ &= \sum_{z_{N-1}} p(x_{N-1}, z_{N-1}|z_{N-2}) \cdot \beta(z_{N-1})\end{aligned}$$

By using the independencies $x_N \perp\!\!\!\perp z_{N-2}|z_{N-1}$ and $x_N \perp\!\!\!\perp x_{N-1}|z_{N-1}$ inferred from the graph.

(c) This gives us for general n:

$$\begin{aligned}\beta(z_n) &= p(x_{n+1}, ..., x_N|z_n) \\ &= \sum_{z_{n+1}} p(x_{n+1}, ..., x_N, z_{n+1}|z_n) \\ &= \sum_{z_{n+1}} p(x_{n+1}, z_{n+1}|z_n) \cdot p(x_{n+2}, ..., x_N|x_{n+1}, z_{n+1}, z_n) \\ &= \sum_{z_{n+1}} p(x_{n+1}, z_{n+1}|z_n) \cdot p(x_{n+2}, ..., x_N|z_{n+1}) \\ &= \sum_{z_{n+1}} p(x_{n+1}, z_{n+1}|z_n) \cdot \beta(z_{n+1})\end{aligned}$$

By using the independencies $x_{n+2}, ..., x_N \perp\!\!\!\perp z_n|z_{n+1}$ and $x_{n+2}, ..., x_N \perp\!\!\!\perp x_{n+1}|z_{n+1}$ inferred from the graph.

Problem 2

1. In order to apply the sum-product algorithm, we first define the factor graph that describes the chain of nodes model as shown in fig. 2.

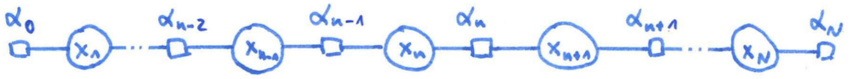

Figure 2: The factor graph corresponding to the chain of nodes model

Then, we define x_n to be the root node and start sending messages from the leaves towards this node. This gives us:

$$\begin{aligned}\mu_{\alpha_0 \to x_1}(x_1) &= \alpha_0(x_1) \\ \mu_{x_1 \to \alpha_1}(x_1) &= \mu_{\alpha_0 \to x_1}(x_1) \\ &\cdots \\ \mu_{\alpha_{n-1} \to x_n}(x_n) &= \sum_{x_{n-1}} \alpha_{n-1}(x_{n-1}, x_n) \mu_{n-1 \to \alpha_{n-1}}(x_{n-1}) \\ &\text{and} \\ \mu_{\alpha_N \to x_N}(x_N) &= \alpha_N(x_N) \\ \mu_{x_N \to \alpha_{N-1}}(x_N) &= \mu_{\alpha_N \to x_{N-1}}(x_N) \\ &\cdots \\ \mu_{\alpha_n \to x_n}(x_n) &= \sum_{x_{n+1}} \alpha_n(x_n, x_{n+1}) \mu_{n+1 \to \alpha_n}(x_{n+1})\end{aligned}$$

And for the marginal probability, we get:

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

Now, we can define:

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

And we see that the application of the sum-product algorithm to the chain of nodes model can result in the message passing algorithm given the right definitions.

2. The entities $\alpha(z_n)$ and $\mu_\alpha(x_n)$ clearly relate to one another, if we adjust the result of the sum-product algorithm as follows:

$$\begin{aligned}\alpha(z_n) &= \sum_{z_{n-1}} p(z_n|z_{n-1})\cdot p(x_n|z_n)\cdot \alpha(z_{n-1})\\ \mu_\alpha(z_n) &= \sum_{z_{n-1}} \alpha_{n-1}(z_{n-1},z_n)\mu_\alpha(z_{n-1})\end{aligned}$$

Where $\mu_\alpha(z_n)$ corresponds to $\alpha(z_n)$, as well as $\mu_\alpha(z_{n-1})$ to $\alpha(z_{n-1})$, $\alpha_{n-1}(z_{n-1},z_n)$ describes the function $p(z_n|z_{n-1})$ and we would have to introduce another factor to capture the dependency between x_n and z_n (i.e. $p(x_n|z_n)$, which does not exist in the chain of nodes model).

Similarly, we can relate $\beta(z_n)$ to $\mu_\beta(x_n)$ by setting

$$\begin{aligned}\beta(z_n) &= \sum_{z_{n+1}} p(x_{n+1},z_{n+1}|z_n)\cdot \beta(z_{n+1})\\ \mu_\beta(z_n) &= \sum_{z_{n+1}} \alpha_n(z_n,z_{n+1})\mu_\beta(z_{n+1})\end{aligned}$$

Where $\mu_\beta(z_n)$ corresponds to $\beta(z_n)$, $\mu_\beta(z_{n+1})$ to $\beta(z_{n+1})$ and the factor $a_n(z_n,z_{n+1})$ can be used to describe the function $p(x_{n+1},z_{n+1}|z_n)$. Here, we need to introduce the variable x_{n+1}, which was not present in the nodes of chain model.

Problem 3

We can evaluate a conditional probability using the message-passing algorithm by introducing the indicator function to all factors involving the variable that we condition on. This is based on the general rule where $p(x_n|x_N=\xi_N) = p(x_n,x_N)\cdot \mathbb{1}(x_N=\xi_N)$ and enables us to condition the algorithm for this variable to take on a specific value (ξ_N). Starting from node x_N this gives us:

$$\begin{aligned}\mu_{\alpha_N\to x_N}(x_N) &= \alpha_N(x_N)\cdot \mathbb{1}(x_N=\xi_N)\\ \mu_{x_N\to\alpha_{N-1}}(x_N) &= \mu_{\alpha_N\to x_{N-1}}(x_N)\\ &\dots\\ \mu_{\alpha_n\to x_n}(x_n) &= \sum_{x_{n+1}} \alpha_n(x_n,x_{n+1})\mu_{n+1\to\alpha_n}(x_{n+1})\end{aligned}$$

Due to the iterative nature of the algorithm all messages going from the leaf node x_N towards the root node will change accordingly to encompass the newly introduced indicator function. Messages that are sent from the leaf node x_1 to the root node remain unchanged.

When sending the messages back from the root node to the leaves the reverse situation arises. Since we forward the messages that the root receives from the opposite side, now the messages going from the root towards the leaf x_1 include the indicator term. The messages that are being sent from the root towards x_N remain unchanged, since they simply forward the information from the front part of the graph.

Problem 4

By definition the marginal is obtained by summing the joint distribution over all variables except $\mathbf{x}_s$ so that

$$p(\mathbf{x}_s) = \sum_{\mathbf{x}\setminus\mathbf{x}_s} p(\mathbf{x}) \tag{1}$$

Since we are working on a factor graph, $p(\mathbf{x})$ is given as the product over all factors and can be rewriten as:

$$\begin{align} p(\mathbf{x}) &= \prod_{\alpha} f_\alpha(\mathbf{x}_\alpha) \tag{2} \\ &= f_s(\mathbf{x}_s)\left[\prod_{i\in\mathrm{ne}(f_s)}\left(\prod_{l\in\mathrm{ne}(x_i)\setminus f_s} F_l(x_i, X_l)\right)\right] \tag{3} \end{align}$$

Using the definition of the function $F_l(x_i, X_l)$ as in Bishop (8.65) as a representation of the product of all the factors in the group associated with factor f_l. Substituting (3) into (1) gives us:

$$p(\mathbf{x}_s) = \sum_{\mathbf{x}\setminus\mathbf{x}_s} f_s(\mathbf{x}_s)\left[\prod_{i\in\mathrm{ne}(f_s)}\left(\prod_{l\in\mathrm{ne}(x_i)\setminus f_s} F_l(x_i, X_l)\right)\right]$$

Now, we can push the sum inside the product, as it does not affect the terms that are dependent on variables in x_s

$$\begin{align*} p(\mathbf{x}_s) &= f_s(\mathbf{x}_s)\left[\prod_{i\in\mathrm{ne}(f_s)}\left(\prod_{l\in\mathrm{ne}(x_i)\setminus f_s} \sum_{\mathbf{x}\setminus\mathbf{x}_s} F_l(x_i, X_l)\right)\right] \\ &= f_s(\mathbf{x}_s)\left[\prod_{i\in\mathrm{ne}(f_s)}\left(\prod_{l\in\mathrm{ne}(x_i)\setminus f_s} \sum_{X_l} F_l(x_i, X_l)\right)\right] \end{align*}$$

Using Bishop (8.64) and (8.69) gives

$$\begin{align*} p(\mathbf{x}_s) &= f_s(\mathbf{x}_s)\left[\prod_{i\in\mathrm{ne}(f_s)}\left(\prod_{l\in\mathrm{ne}(x_i)\setminus f_s} \mu_{f_l\to x_i}(x_i)\right)\right] \\ &= f_s(\mathbf{x}_s)\left[\prod_{i\in\mathrm{ne}(f_s)} \mu_{x_i\to f_s}(x_i)\right] \end{align*}$$