Machine Learning 2 **Due: April 26, 2016**

Homework 3

Instructor: Ke Tran Email: m.k.tran@uva.nl

Student: Andrea Jemmett UvA ID: 11162929
Collaborators: N/A Email: andreajemmett@gmail.com

Problem 1. Consider the inference problem of evaluating $p(\mathbf{x}_n|\mathbf{x}_N)$ for the graph shown in Figure 1, for all nodes $n \in \{1, \dots, N-1\}$. Show that the message passing algorithm can be used to solve this efficiently, and discuss which messages are modified and in what way.

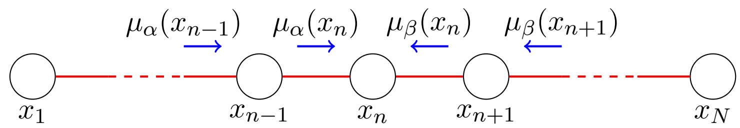

Figure 1: Chain of nodes model

Problem 2. Apply the sum-product algorithm to the chain of nodes model in Figure 1 and show that the results of message passing algorithm are recovered as a special case, that is

$$\begin{aligned}
p(x_n) &= \frac{1}{Z}\mu_\alpha(x_n)\mu_\beta(x_n) \\
\mu_\alpha(x_n) &= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n)\mu_\alpha(x_{n-1}) \\
\mu_\beta(x_n) &= \sum_{x_{n+1}} \psi_{n+1,n}(x_{n+1}, x_n)\mu_\beta(x_{n+1})
\end{aligned}$$

where $\psi_{i,i+1}(x_i, x_{i+1})$ is a potential function defined over clique $\{x_i, x_{i+1}\}$.

Solution:

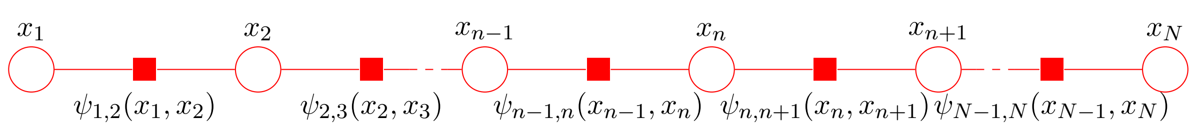

Figure 2: Factor graph for chain of nodes model

The first thing to do to apply the sum-product algorithm to the chain model of Figure 1 is to transform it into a factor graph. The factor graph for the chain model is shown in Figure 2.

Once we have a factor graph, we need to chose a root node, *e.g.*, x_n, so that we can start propagating messages from the leaf nodes (x_1 and x_N) to the root. We start by propagating messages from x_1 to the root:

$$\mu_{x_1\to\psi_{1,2}}(x_1) = 1 \tag{1}$$
$$\mu_{\psi_{1,2}\to x_2}(x_2) = \sum_{x_1} \psi_{1,2}(x_1, x_2) \tag{2}$$
$$\mu_{x_2\to\psi_{2,3}}(x_2) = \mu_{\psi_{1,2}\to x_2}(x_2) \tag{3}$$
$$\mu_{\psi_{2,3}\to x_3}(x_3) = \sum_{x_2} \psi_{2,3}(x_2, x_3)\mu_{x_2\to\psi_{2,3}}(x_2) \tag{4}$$
$$\vdots \qquad \vdots$$
$$\mu_{x_{n-1}\to\psi_{n-1,n}}(x_{n-1}) = \mu_{\psi_{n-2,n-1}\to x_{n-1}}(x_{n-1}) \tag{5}$$
$$\mu_{\psi_{n-1,n}\to x_n}(x_n) = \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n)\mu_{x_{n-1}\to\psi_{n-1,n}}(x_{n-1}) \tag{6}$$

and we note that a message that a variable node sends on one link is equal to the message that node has received on its other link. That is variable nodes act as *proxies* for messages created at factor nodes.

If we define $\mu_\alpha(x_n)$ as the incoming from the left factor node to variable node x_n, we have that

$$\begin{aligned}\mu_\alpha(x_n) &\equiv \mu_{\psi_{n-1,n}\to x_n}(x_n) &(7)\\ &= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n)\mu_{x_{n-1}\to\psi_{n-1,n}}(x_{n-1}) &(8)\\ &= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n)\mu_{\psi_{n-2,n-1}\to x_{n-1}}(x_{n-1}) &(9)\\ &= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n)\mu_\alpha(x_{n-1}) &(10)\end{aligned}$$

If we compute in the same way the messages from leaf node x_N to the root node x_n, we obtain that

$$\begin{aligned}\mu_\beta(x_n) &\equiv \mu_{\psi_{n,n+1}\to x_n}(x_n) &(11)\\ &= \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1})\mu_{x_{n+1}\to\psi_{n,n+1}} &(12)\\ &= \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1})\mu_{\psi_{n+1,n+2}\to x_{n+1}}(x_{n+1}) &(13)\\ &= \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1})\mu_\beta(x_{n+1}) &(14)\end{aligned}$$

Finally applying eq. 8.63 from Bishop (with the addition of the normalizing constant Z that we know from the full joint distribution or we can compute locally on the node x_n) we have that

$$\begin{aligned}p(x_n) &= \frac{1}{Z}\mu_{\psi_{n-1,n}\to x_n}(x_n)\mu_{\psi_{n,n+1}\to x_n}(x_n) &(15)\\ &= \frac{1}{Z}\mu_\alpha(x_n)\mu_\beta(x_n) &(16)\end{aligned}$$

■

Problem 3. Run sum-product algorithm on the graph in Figure 3 with node x_3 designed as the root. Using the computed messages given in *p.409 Bishop*.

1. Show that the correct marginals are obtained for x_1 and x_3.

2. Show that the sum-product algorithm on this graph gives the correct joint distribution for x_1, x_2.

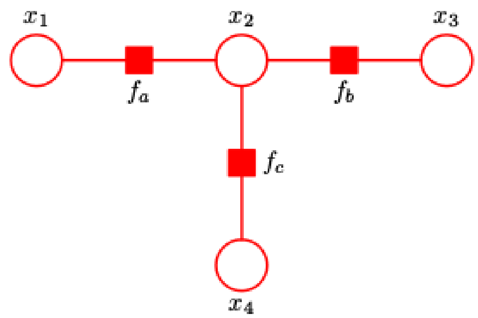

Figure 3: A simple factor graph

Problem 4. Show that the marginal distribution for the variables $\mathbf{x}_s$ in a factor $f_s(\mathbf{x}_s)$ in a tree-structured factor graph, after running the sum-product message passing algorithm, can be written as

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

where $\mathrm{ne}(f_s)$ denotes the set of variable nodes that are neighbors of the factor node f_s