Machine Learning Principles and Methods **Due: March 26, 2016**

Homework 3

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

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

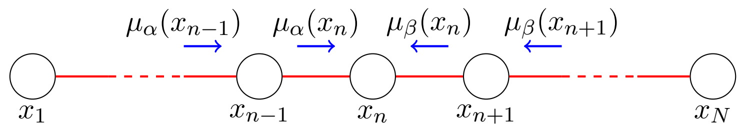

Figure 1: Chain of nodes model

Solution: When x_N is observed, the potential function $\psi(x_{N-1}, x_N)$ is the only one that depends on x_N. Assume that $x_N = x^*$, we can modify $\psi(x_{N-1}, x_N)$ as $\psi(x_{N-1}, x_N)\mathbb{I}[x_N = x^*]$ where $\mathbb{I}[x_N = x^*]$ is the identity function

$$\mathbb{I}[x_N = x^*] = \begin{cases} 1 & \text{if } x_N = x^* \\ 0 & \text{if } x_N \neq x^* \end{cases}$$

By doing so, we can treat x_N as an unobserved variable, thus we can keep using message passing algorithm discussed in Bishop.

Another possibility is to remove x_N from the chain, and the potential function $\psi(x_{N-2}, x_{N-1})$ is modified as

$$\psi(x_{N-2}, x_{N-1})\psi(x_{N-1}, x^*)$$

■

Problem 2. 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

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

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

Solution: Let $\mathbf{x}_s = \{x_1, x_2, \ldots, x_M\}$ be the set of variables associated with factor $f_s(\mathbf{x}_s)$. The joint distribution can be written as

$$p(\mathbf{x}) = f_s(\mathbf{x}_s) \prod_{i=1}^{M} \prod_{t \in \mathsf{ne}(x_i)} F_t(x_i, X_t)$$

$\mathsf{ne}(x_i)$ denotes the set of factor nodes that are neighbors of x_i except f_s, and X_t denotes the set of all variables in the subtree connected to the variable node x_i via the factor node f_t, and $F_t(x_i, X_t)$

represents the product of all the factors in the group associated with factor f_t. The marginal distribution $p(\mathbf{x}_s)$ is obtained by summing all over $\mathbf{x}\backslash\mathbf{x}_s$.

$$\begin{aligned}
p(\mathbf{x}_s) &= \sum_{\mathbf{x}\backslash\mathbf{x}_s} p(\mathbf{x}) && (1)\\
&= \sum_{\mathbf{x}\backslash\mathbf{x}_s} f_s(\mathbf{x}_s) \prod_{i=1}^{M} \prod_{t\in\mathsf{ne}(x_i)} F_t(x_i, X_t) && (2)\\
&= f_s(\mathbf{x}_s) \prod_{i=1}^{M} \prod_{t\in\mathsf{ne}(x_i)} \left[\sum_{X_t} F_t(x_i, X_t)\right] && (3)\\
&= f_s(\mathbf{x}_s) \prod_{i=1}^{M} \prod_{t\in\mathsf{ne}(x_i)} \mu_{f_t\to x_i}(x_i) && (4)\\
&= f_s(\mathbf{x}_s) \prod_{i=1}^{M} \mu_{x_i\to f_s} && (5)
\end{aligned}$$

■