Machine Learning 2 **Due: 13:00, April 30, 2018**

Homework 4

Instructors: Stephan Bongers, Stephan Alaniz Kupsch
Email: s.r.bongers@uva.nl, s.a.alanizkupsch@uva.nl

You are allowed to discuss with your colleagues but you should write the answers in *your own words*. If you discuss with others, write down the name of your collaborators on top of the first page. No points will be deducted for collaborations. If we find similarities in solutions beyond the listed collaborations we will consider it as cheating.
We will not accept any late submissions under any circumstances. The solutions to the previous homework will be handed out in the class at the beginning of the next homework session. After this point, late submissions will be automatically graded zero.
$\star$ denotes bonus exercise. You earn bonus point for solving each bonus exercise. All bonus points earned will be added to your total homework points.

Problem 1. (0.5 + 0.5 + 0.5 + 2.5 = 4 pts)

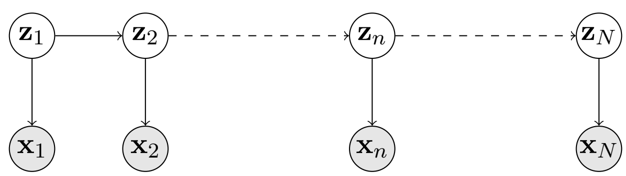

Figure 1: Markov chain of latent variables.

Given the Bayesian network in Figure 1, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ and $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$:

1. Write down the factorized joint probability distribution $p(\mathbf{Z}, \mathbf{X})$.
2. Draw the the corresponding factor graph.
3. Write down the the joint probability distribution using the factors introduced in 2.
4. Given $\mathbf{X}$, we want to infer z_n such that

$$p(z_n|\mathbf{X}) = \frac{p(\mathbf{X}|z_n)p(z_n)}{p(\mathbf{X})} \tag{1}$$

$$= \frac{\alpha(z_n)\beta(z_n)}{p(\mathbf{X})}. \tag{2}$$

   Using the conditional independencies of the graph in Figure 1, derive $\alpha(z_n)$ and $\beta(z_n)$ so that they are recursive definitions of themselves, i.e. $\alpha(z_n)$ is calculated from $\alpha(z_{n-1})$ and $\beta(z_n)$ is calculated from $\beta(z_{n+1})$. Indicate where you use independencies inferred from the graphical model.

Problem 2. (1.5 pts + 1 $\star$ pts)

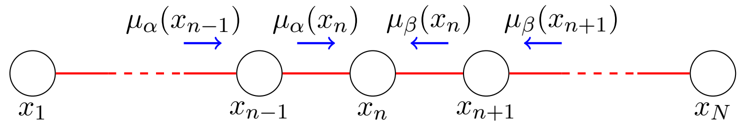

Figure 2: Chain of nodes model

1. Apply the sum-product algorithm (as in Bishop's section 8.4.4) to the chain of nodes model in Figure 2 and show that the results of message passing algorithm (as in Bishop's section 8.4.1) 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 $\{x_i, x_{i+1}\}$.

2. $\star$ Establish a relation of your results $\alpha(z_n)$ and $\beta(z_n)$ in 1.4 with the results of the sum-product algorithm $\mu_\alpha(x_n)$ and $\mu_\beta(x_n)$.

Problem 3. (1.5 pts) Consider the inference problem of evaluating $p(\mathbf{x}_n|\mathbf{x}_N)$ for the graph shown in Figure 2, 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.

Problem 4. (2 pts) 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