School of Computing
National University of Singapore
CS5340: Uncertainty Modeling in AI
Semester 1, AY 2018/19

Exercise 2

Question 1

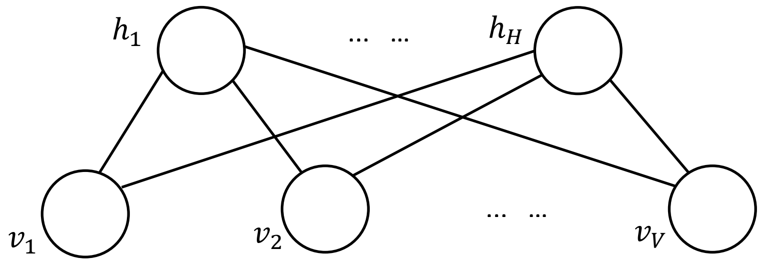


Fig. 1.1

The restricted Boltzmann machine is a Markov Random Field (MRF) defined on a bipartite graph as shown in Fig. 3.1. It consists of a layer of visible variables $\boldsymbol{v} = [v_1, \dots, v_V]^T$ and hidden variables $\boldsymbol{h} = [h_1, \dots, h_H]^T$, where all variables are binary taking states $\{0,1\}$. The joint distribution of the MRF is given by:

$$p(\boldsymbol{v}, \boldsymbol{h}) = \frac{1}{Z(\boldsymbol{W}, \boldsymbol{a}, \boldsymbol{b})} \exp(\boldsymbol{v}^T \boldsymbol{W} \boldsymbol{h} + \boldsymbol{a}^T \boldsymbol{v} + \boldsymbol{b}^T \boldsymbol{h}),$$

where $\theta = \{\boldsymbol{W}_{V\times H}, \boldsymbol{a}_{V\times 1}, \boldsymbol{b}_{H\times 1}\}$ are the parameters of the potential functions, and $Z(.)$ is the partition function.

a) Given that:

$$p(h_i = 1 \mid \boldsymbol{v}) = \sigma\big(b_i + \textstyle\sum_j W_{ji} v_j\big),$$

where $\sigma(x) = \frac{e^x}{1+e^x}$ is the sigmoid activation function. Show that the distribution of hidden units conditioned on the visible units factorizes as:

$$p(\boldsymbol{h} \mid \boldsymbol{v}) = \textstyle\prod_i p(h_i \mid \boldsymbol{v}).$$

Show all your workings clearly.

b) Assuming that the restricted Boltzmann machine consists of only 2 visible and 1 hidden variables, and the joint distribution of the MRF is given by:

h	v_1	v_2	$\exp(\boldsymbol{v}^T\boldsymbol{W}h + \boldsymbol{a}^T\boldsymbol{v} + bh)$
0	0	0	1.00
0	0	1	2.13
0	1	0	4.65
0	1	1	9.90
1	0	0	3.65
1	0	1	8.66
1	1	0	4.22
1	1	1	10.01

Find the unknown parameters, i.e. $\theta = \{\boldsymbol{W}_{2\times1}, \boldsymbol{a}_{2\times1}, b\}$.

Question 2

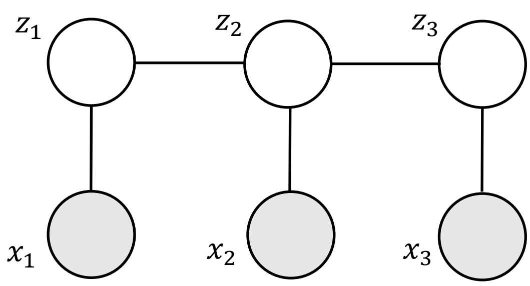


Fig. 2.1

Fig. 4.1 shows a Markov Random Field (MRF) representation of a Hidden Markov Model (HMM) over three time steps. The hidden variables z_1, z_2, z_3 are discrete random variables that take three possible states $z_n \in \{F, H, M\}$, and x_1, x_2, x_3 are the observed variables that take on real values $x_n \in \mathbb{R}$. The joint distribution is given by:

$$p(z_1, z_2, z_3, x_1, x_2, x_3) = \frac{1}{Z}\prod_{n=2}^{3} \psi_t(z_n, z_{n-1}) \prod_{n=1}^{3} \psi_e(x_n, z_n),$$

where Z is the partition function, and the transition potential $\psi_t(z_n, z_{n-1})$ and the emission potentials $\psi_e(x_n, z_n)$ are given by:

$\psi_t(z_n, z_{n-1})$	$z_n = F$	$z_n = H$	$z_n = M$
$z_{n-1} = F$	2.0	3.0	5.0
$z_{n-1} = H$	1.0	6.0	3.0
$z_{n-1} = M$	4.5	2.0	2.5

z_1	$\psi_e(x_1, z_1)$
F	1.0
H	8.0
M	1.0

z_2	$\psi_e(x_2, z_2)$
F	7.0
H	1.0
M	2.0

z_3	$\psi_e(x_3, z_3)$
F	2.0
H	3.0
M	5.0

Decode the message that corresponds to the states of the hidden variables that give the maximal probability. Show all your workings clearly.

Question 3

Fig. 3.1 shows a Bayesian network of the mixture of Bernoulli Distribution. X_n is a binary random variable, i.e. $x_n \in \{0,1\}$. N is the total number of observations. Z_n is the 1-of-k indicator random variable, $z_{nk} = 1 \Rightarrow z_{n,j\neq k} = 0$ indicates the assignment of the random variable x to the k^{th} Bernoulli density. $z_{nk} \in \{0,1\}$ and $\sum_k z_{nk} = 1$.

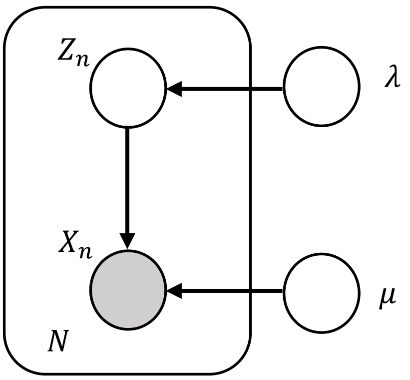


Fig. 3.1

Given the expressions for the Bernoulli distribution:

$$p(\mathrm{x} \mid \mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{(1-x_n)},$$

and marginal distribution of Z_n, which is a categorical distribution specified in terms of the mixing coefficients λ_k:

$$p(\mathrm{z_n}) = \prod_{k=1}^{K} \lambda_k^{z_{nk}} = \mathrm{cat}_{\mathrm{z_n}}[\lambda]\text{ , where } 0 \le \lambda_k \le 1 \text{ and } \sum_k \lambda_k = 1.$$

(a) Show that the mixture of Bernoulli distribution is given by:

$$p(\mathrm{x} \mid \mu, \lambda) = \prod_{n=1}^{N} \sum_{k=1}^{K} \lambda_k \mu_k^{x_n} (1 - \mu_k)^{(1-x_n)}.$$

(b) Derive the responsibility $\gamma(z_{nk}) = p(z_{nk} = 1 \mid \mathrm{x})$, and show that the updates for the unknown parameters μ and λ in the maximization step of the EM algorithm are given by:

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathrm{x}_n,$$
$$\lambda_k = \frac{N_k}{N}\text{, where } N_k = \sum_{n=1}^{N} \gamma(z_{nk}).$$

Show all your workings clearly.

Question 4

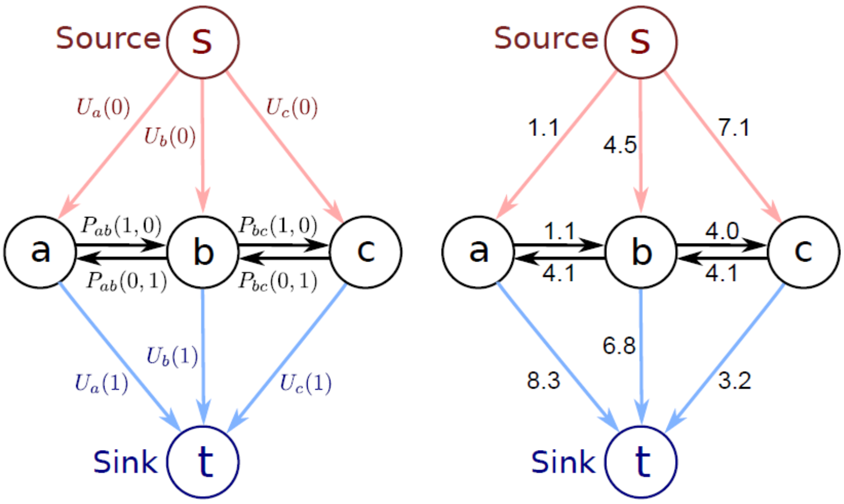


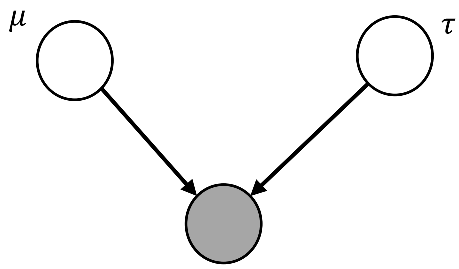
Fig 4.1
(Image source: “Computer Vision: Models, Learning and Inference”, Simon Prince)

Compute the **MAP solution** to the three-pixel graph cut problem in Fig. 4.1 by

(i) computing the cost of all eight possible solutions explicitly and finding the one with the minimum cost, and
(ii) running the augmenting paths algorithm on this graph by hand and interpreting the minimum cut.

Question 5

Consider the simple 3-node graph shown in Fig. 5.1 in which the observed node X is given by a Gaussian distribution $\mathcal{N}(x|\mu,\tau^{-1})$ with mean μ and precision τ. Suppose that the marginal distributions over the mean and precision are given by $\mathcal{N}(\mu|\mu_0,s_0)$ and $\mathrm{Gam}(\tau|a,b)$, where $\mathrm{Gam}(.|.,.)$ denotes a gamma distribution. Write down expressions for the conditions distributions for the conditional distributions $p(\mu|x,\tau)$ and $p(\tau|x,\mu)$ that would be required in order to apply Gibbs sampling to the posterior distribution $p(\mu,\tau \mid x)$.



x

Fig. 5.1

--End--