School of Computing
National University of Singapore
CS5340: Uncertainty Modeling in AI
Semester 1, AY 2020/21

Exercise 2

Question 1

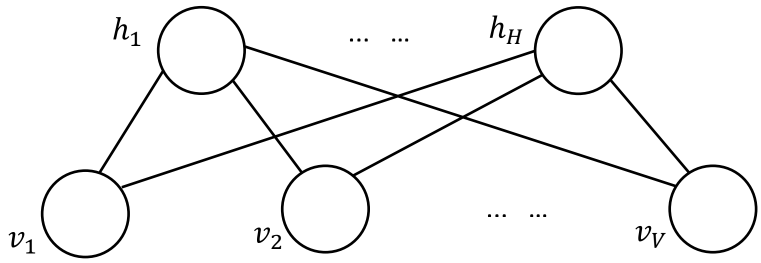


Fig. 1.1

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:

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

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.

Answer:

Using product rule, we have:

$$\begin{aligned}
p(\boldsymbol{h}|\boldsymbol{v}) &= \frac{p(\boldsymbol{h},\boldsymbol{v})}{\sum_{\boldsymbol{h}} p(\boldsymbol{h},\boldsymbol{v})} \\
&= \frac{\frac{1}{Z}exp\{\boldsymbol{v}^T\boldsymbol{W}\boldsymbol{h} + \boldsymbol{a}^T\boldsymbol{v} + \boldsymbol{b}^T\boldsymbol{h}\}}{\frac{1}{Z}\sum_{\boldsymbol{h}} exp\{\boldsymbol{v}^T\boldsymbol{W}\boldsymbol{h} + \boldsymbol{a}^T\boldsymbol{v} + \boldsymbol{b}^T\boldsymbol{h}\}} \\
&= \frac{exp\{(\boldsymbol{v}^T\boldsymbol{W} + \boldsymbol{h}^T)\boldsymbol{h}\}exp\{\boldsymbol{a}^T\boldsymbol{v}\}}{\sum_{(\boldsymbol{h}} exp\{\boldsymbol{v}^T\boldsymbol{W} + \boldsymbol{h}^T)\boldsymbol{h}\}exp\{\boldsymbol{a}^T\boldsymbol{v}\}} \\
&= \frac{exp\{(\boldsymbol{v}^T\boldsymbol{W} + \boldsymbol{h}^T)\boldsymbol{h}\}}{\sum_{(\boldsymbol{h}} exp\{\boldsymbol{v}^T\boldsymbol{W} + \boldsymbol{h}^T)\boldsymbol{h}\}}
\end{aligned}$$

Let $\boldsymbol{m}^T = \boldsymbol{v}^T\boldsymbol{W} + \boldsymbol{b}^T$ and since $\boldsymbol{h} = [h_1, h_2, ...h_H]^T$, we have:

$$\begin{aligned}
p(\boldsymbol{h}|\boldsymbol{v}) &= \frac{exp\{[m_1, m_2...m_H][h_1, h_2...h_H]^T\}}{\sum_{\boldsymbol{h}} exp\{[m_1, m_2...m_H][h_1, h_2...h_H]^T\}} \\
&= \frac{exp\{m_1h_1, m_2h_2...m_Hh_H\}}{\sum_{\boldsymbol{h}} exp\{m_1h_1, m_2h_2...m_Hh_H\}} \\
&= \frac{exp(m_1h_1)exp(m_2h_2)...exp(m_Hh_H)}{\sum_{h_1}\sum_{h_2}...\sum_{h_H} exp(m_1h_1)exp(m_2h_2)...exp(m_Hh_H)} \\
&= \frac{exp(m_1h_1)}{\sum_{h_1} exp(m_1h_1)}\frac{exp(m_2h_2)}{\sum_{h_2} exp(m_2h_2)}\cdots\frac{exp(m_Hh_H)}{\sum_{h_H} exp(m_Hh_H)} \\
&= \Pi_i\frac{exp(m_ih_i)}{\sum_{h_i} exp(m_ih_i)} \\
&= \Pi_i\frac{exp(m_ih_i)}{exp(m_ih_i = 0) + exp(m_ih_i = 1)} \\
&= \Pi_i\frac{exp(m_ih_i)}{1 + exp(m_ih_i = 1)} \\
&= \Pi_i p(h_i|\boldsymbol{v})
\end{aligned}$$

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. $\theta = \{\boldsymbol{W}_{2\times1}, \boldsymbol{a}_{2\times1}, b\}$.

Answer:

$$exp\{[v_1\ v_2]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} h + [a_1\ a_2]\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + bh\}$$

Case 1: $h = 0, v_1 = 0, v_2 = 1$

$$exp\{[0\ 1]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} 0 + [a_1\ a_2]\begin{bmatrix} 1 \\ 0 \end{bmatrix} + b\ 0\} = 2.13$$

$$\Rightarrow exp(a_2) = 2.13 \Rightarrow a_2 = 0.756$$

Case 2: $h = 0, v_1 = 1, v_2 = 0$

$$exp\{[0\ 1]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} 0 + [a_1\ a_2]\begin{bmatrix} 0 \\ 1 \end{bmatrix} + b\ 0\} = 4.65$$

$$\Rightarrow exp(a_1) = 4.65 \Rightarrow a_1 = 1.537$$

Case 3: $h = 1, v_1 = 0, v_2 = 0$

$$exp\{[0\ 0]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} 1 + [a_1\ a_2]\begin{bmatrix} 0 \\ 0 \end{bmatrix} + b\ 1\} = 3.65$$

$$\Rightarrow exp(b) = 3.65 \Rightarrow b = 1.2947$$

Case 4: $h = 1, v_1 = 0, v_2 = 1$

$$exp\{[0\ 1]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} 1 + [a_1\ a_2]\begin{bmatrix} 0 \\ 1 \end{bmatrix} + b\ 1\} = 8.66$$

$$\Rightarrow exp(w_2 + a_2 + b) = 8.66 \Rightarrow exp(w_2 + 0.756 + 1.2947) = 8.66$$

$$\Rightarrow w_2 + 2.0507 = 2.1587 \Rightarrow w_2 = 0.1080$$

Case 5: $h = 1, v_1 = 1, v_2 = 0$

$exp\{[1\ 0] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} 1 + [a_1\ a_2] \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b\ 1\} = 4.22$

$\Rightarrow exp(w_1 + a_1 + b) = 4.22 \Rightarrow exp(w_1 + 1.537 + 1.2947) = 4.22$

$\Rightarrow w_1 + 2.8317 = 1.4398 \Rightarrow w_1 = -1.3919$

Verifications:

Case 1: $h = 0, v_1 = 0, v_2 = 0 \Rightarrow exp(0) = 1.00$

Case 2: $h = 0, v_1 = 1, v_2 = 1 \Rightarrow exp(a_1 + a_2) = exp(1.537 + 0.756) = 9.90$

Case 3: $h = 1, v_1 = 1, v_2 = 1$

$\Rightarrow exp(v_1 W_1 h + v_2 W_2 h + a_1 v_1 + a_2 v_2 + bh) = exp(w_1 + w_2 + a_1 + a_2 + b)$

$= exp(-1.3919 + 0.1080 + 1.537 + 0.756 + 1.2947) = 10.0122$

Question 2

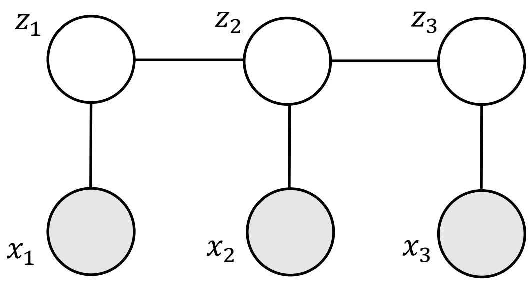


Fig. 2.1

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:

$\boldsymbol{\psi_t(z_n, z_{n-1})}$	$z_n = F$	$z_n = H$	$z_n = M$
$z_{n-1} = F$	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.

Answer:

The solution can be evaluated as:

$$\max_{z_1,z_2,z_3} \psi(z_3,x_3)\psi(z_2,z_3)\psi(z_2,x_2)\,\psi(z_1,z_2)\psi(z_1,x_1) =$$

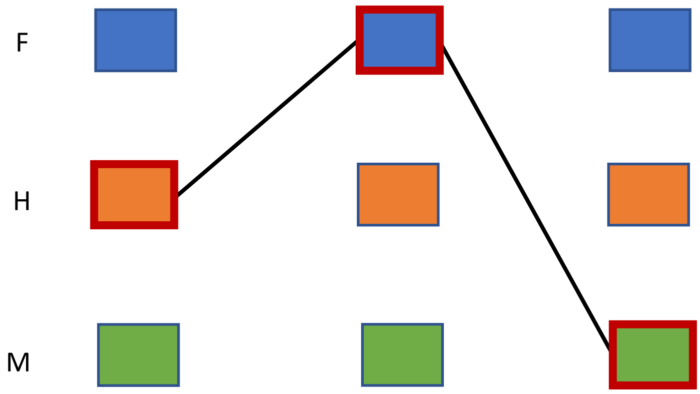
$$\max_{z_3} \psi(z_3,x_3) \max_{z_2} \psi(z_2,z_3)\psi(z_2,x_2) \max_{z_1} \psi(z_1,z_2)\psi(z_1,x_1)$$

$\boldsymbol{Z_2}$	$\max\limits_{z_1} \boldsymbol{\psi(z_1,z_2)\psi(z_1,x_1) = z_2^{max}(z_1)}$	$\boldsymbol{\delta^{max}(z_1)}$
F	$\max(2.0\times 1.0, 1.0\times 8.0, 4.5\times 1.0) = \max(2.0,8.0,4.5) = 8.0$	H
H	$\max(3.0\times 1.0, 6.0\times 8.0, 2.0\times 1.0) = \max(3.0,48.0,2.0) = 48.0$	H
M	$\max(5.0\times 1.0, 3.0\times 8.0, 2.5\times 1.0) = \max(5.0,24.0,2.5) = 24.0$	H

$\boldsymbol{Z_3}$	$\max\limits_{z_2} \boldsymbol{\psi(z_2,z_3)\psi(z_2,x_2)\, z_2^{max}(z_1) = z_3^{max}(z_2)}$	$\boldsymbol{\delta^{max}(z_2)}$
F	$\max(2.0\times 7.0\times 8.0, 1.0\times 1.0\times 48.0, 4.5\times 2.0\times 24.0)$ $= \max(112.0,48.0,216.0) = 216.0$	M
H	$\max(3.0\times 7.0\times 8.0, 6.0\times 1.0\times 48.0, 2.0\times 2.0\times 24.0)$ $= \max(168.0,288.0,96.0) = 288.0$	H
M	$\max(5.0\times 7.0\times 8.0, 3.0\times 1.0\times 48.0, 2.5\times 2.0\times 24.0)$ $= \max(280.0,144.0,120.0) = 280.0$	F

$\max\limits_{z_3} \boldsymbol{\psi(z_3,x_3)\, z_3^{max}(z_2)}$	$\boldsymbol{\delta^{max}(z_3)}$
$\max(216\times 2.0, 288.0\times 3.0, 280\times 5.0)$ $= \max(432.0,864.0,1400.0) = 1400.0$	M

Backtracking:



The code is: HFM

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 $\sum_k z_{nk} = 1$.

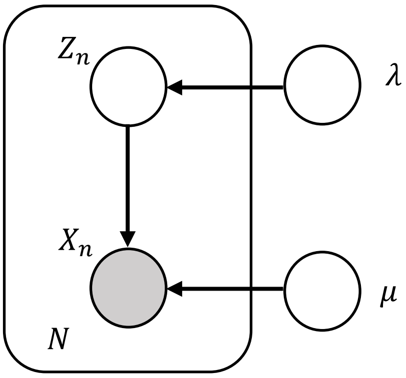


Fig. 3.1

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.

Answer:

Refer to Section 9.3.3 in “Pattern Recognition and Machine Learning”, Christopher Bishop.

Question 4

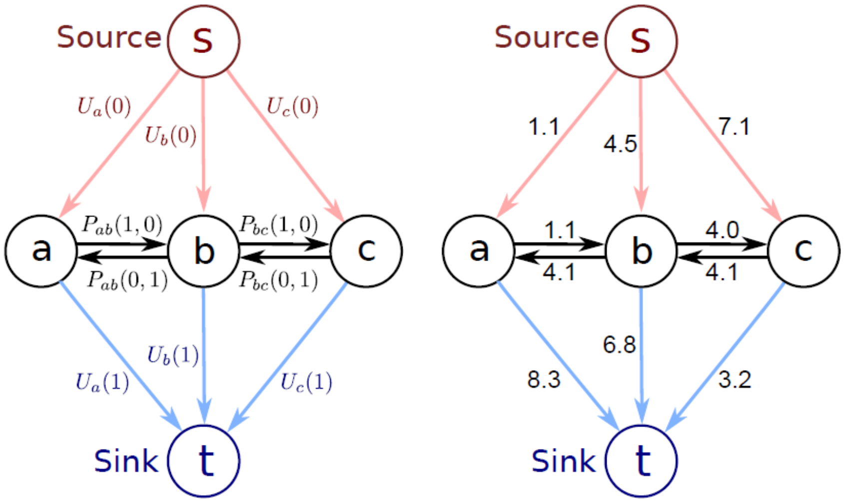


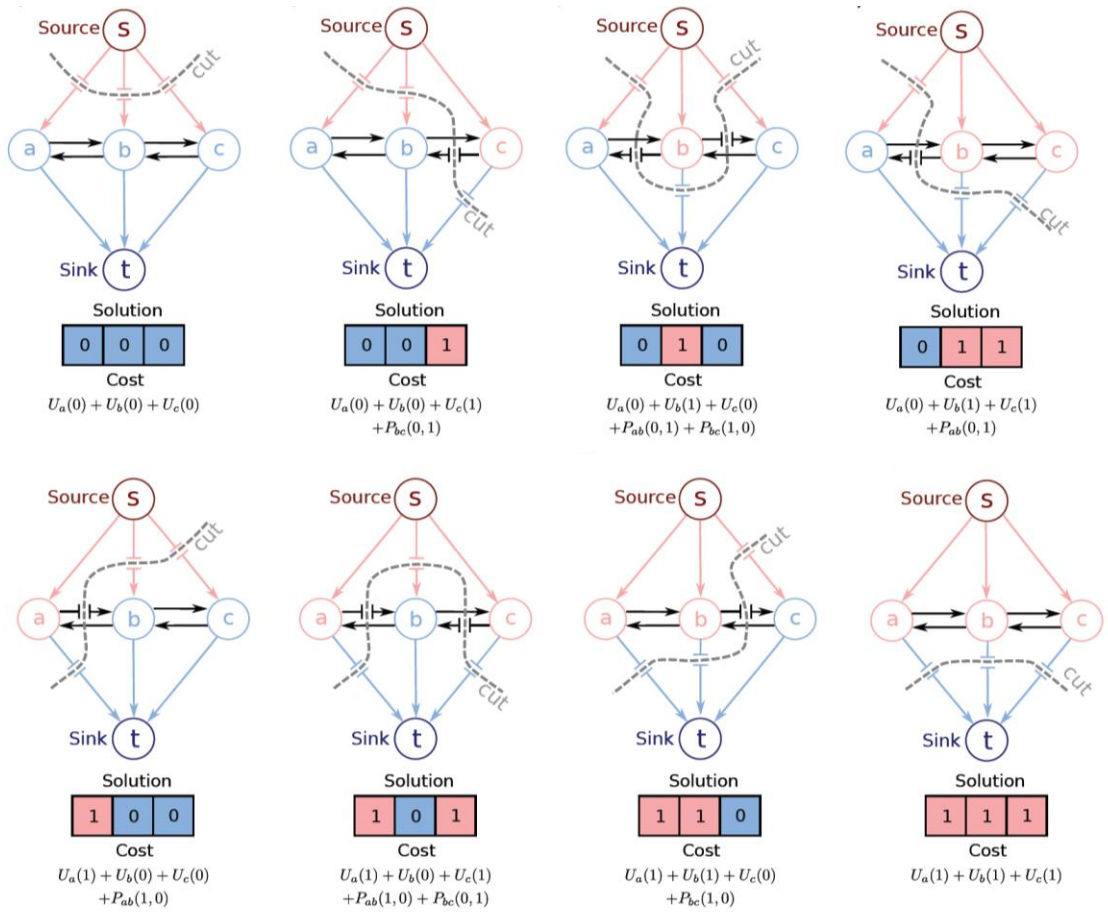
Fig 4.1

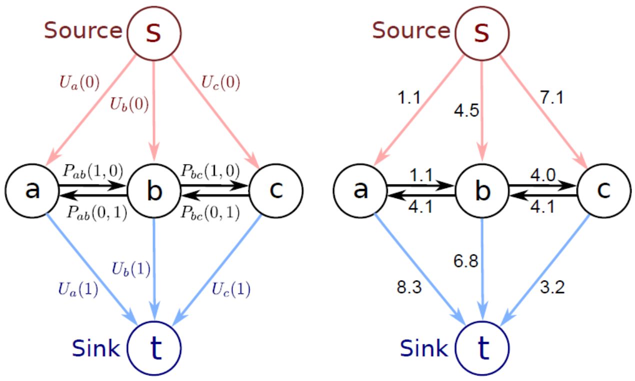
(Image source: "Computer Vision: Models, Learning and Inference", Simon Prince)

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

Answer:



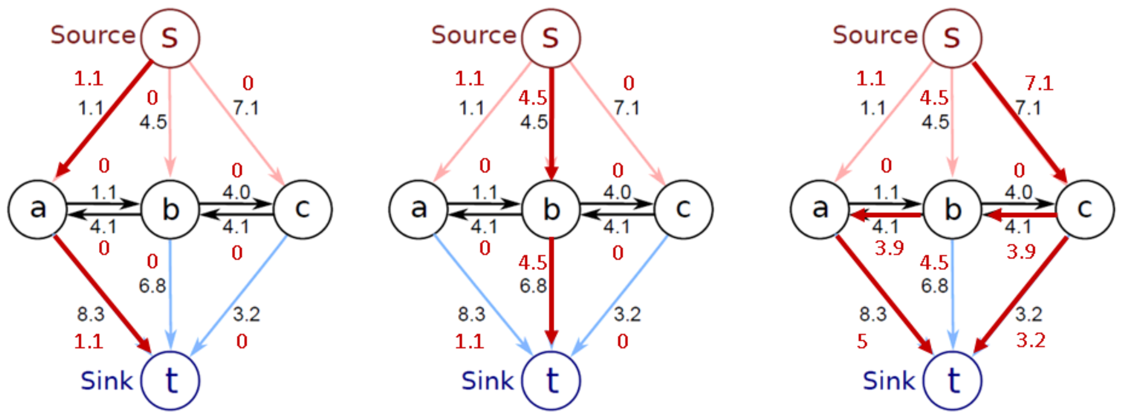


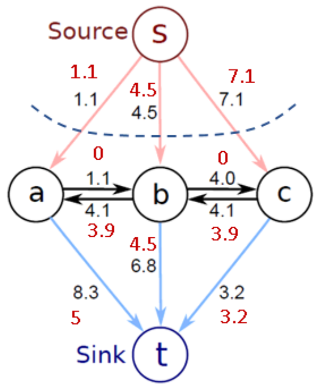
$\boldsymbol{U_a(0) + U_b(0) + U_c(0) = 1.1 + 4.5 + 7.1 = 12.7}$
$U_a(0) + U_b(0) + U_c(1) + P_{bc}(0,1) = 1.1 + 4.5 + 3.2 + 4.1 = 12.9$
$U_a(0) + U_b(1) + U_c(0) + P_{ab}(0,1) + P_{bc}(1,0) = 1.1 + 6.8 + 7.1 + 4.1 + 4.0 = 23.1$
$U_a(0) + U_b(1) + U_c(1) + P_{ab}(0,1) = 1.1 + 6.8 + 3.2 + 4.1 = 15.2$

$U_a(1) + U_b(0) + U_c(0) + P_{ab}(1,0) = 8.3 + 4.5 + 7.1 + 1.1 = 21$
$U_a(1) + U_b(0) + U_c(1) + P_{ab}(1,0) + P_{bc}(0,1) = 8.3 + 4.5 + 3.2 + 1.1 + 4.1 = 21.2$
$U_a(1) + U_b(1) + U_c(0) + P_{bc}(1,0) = 8.3 + 6.8 + 7.1 + 4.0 = 26.2$
$U_a(1) + U_b(1) + U_c(1) = 8.3 + 6.8 + 3.2 = 18.3$

(ii) running the augmenting paths algorithm on this graph by hand and interpreting the minimum cut.

Answer:





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 to apply Gibbs sampling to the posterior distribution $p(\mu,\tau \mid x)$.

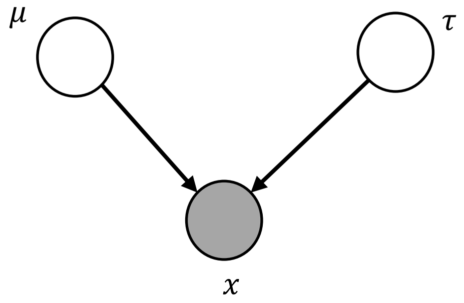


Fig. 5.1

Answer:

$$p(\mu|x,\tau)=\frac{p(\mu,x,\tau)}{\int p(\mu,x,\tau)d\mu}=\frac{p(\mu)p(\tau)p(x|\mu,\tau)}{\int p(\mu)p(\tau)p(x|\mu,\tau)d\mu}=\frac{p(\mu)p(x|\mu,\tau)}{\int p(\mu)p(x|\mu,\tau)d\mu}$$

$$p(x \mid \mu,\tau)=C_x\exp\{-0.5\tau(x-\mu)^2\}$$
$$p(\mu \mid \mu_0,s_0)=C_\mu\exp\{-0.5s_0(\mu_0-\mu)^2\}$$
$$p(\mu)p(x|\mu,\tau)=C_xC_\mu\exp\{-0.5\left[\mu^2(\tau+s_0)-2\mu(\tau x-s_0\mu_0)+(\tau x^2+s_0\mu_0^2)\right]\}$$

$$p(\mu|x,\tau)=\frac{p(\mu)p(x|\mu,\tau)}{\int p(\mu)p(x|\mu,\tau)d\mu}$$

$$=\frac{\exp\{-0.5\left[\mu^2(\tau+s_0)-2\mu(\tau x-s_0\mu_0)+(\tau x^2+s_0\mu_0^2)\right]\}}{\int\exp\{-0.5\left[\mu^2(\tau+s_0)-2\mu(\tau x-s_0\mu_0)+(\tau x^2+s_0\mu_0^2)\right]\}d\mu}$$

$$=\frac{\exp\{-0.5\left[\mu^2(\tau+s_0)-2\mu(\tau x-s_0\mu_0)\right]\}}{\int\exp\{-0.5\left[\mu^2(\tau+s_0)-2\mu(\tau x-s_0\mu_0)\right]\}d\mu}$$

$$=\frac{\exp\{-\alpha\mu^2+\beta\mu\}}{\int\exp\{-\alpha\mu^2+\beta\mu\}d\mu},\quad \text{where}\quad \alpha=0.5(\tau+s_0) \text{ and } \beta=\tau x-s_0\mu_0\ .$$

Since $\int_{-\infty}^{+\infty}\exp\{-\alpha x^2+\beta x\}\,dx=\sqrt{\frac{\pi}{\alpha}}\exp\left\{\frac{\beta^2}{4\alpha}\right\}$,

$$p(\mu|x,\tau)=\frac{\exp\{-\alpha\mu^2+\beta\mu\}}{\sqrt{\frac{\pi}{\alpha}}\exp\left\{\frac{\beta^2}{4\alpha}\right\}}$$

$$p(\tau|x,\mu) = \frac{p(\mu,x,\tau)}{\int p(\mu,x,\tau)d\tau} = \frac{p(\mu)p(\tau)p(x|\mu,\tau)}{\int p(\mu)p(\tau)p(x|\mu,\tau)d\tau} = \frac{p(\tau)p(x|\mu,\tau)}{\int p(\tau)p(x|\mu,\tau)d\tau}$$

$p(x \mid \mu,\tau) = C_x \exp\{-0.5\tau(x-\mu)^2\}$
$p(\tau \mid a,b) = C_\tau \tau^{a_0-1}\, exp(-b_0\tau)$
$p(\tau)p(x|\mu,\tau) = C_x C_\tau \tau^{a_0-1} \exp\{\tau[-0.5(x-\mu)^2 - b_0]\}$

$$p(\tau|x,\mu) = \frac{p(\tau)p(x|\mu,\tau)}{\int p(\tau)p(x|\mu,\tau)d\tau} = \frac{\tau^{a_0-1}\exp\{\tau[-0.5(x-\mu)^2 - b_0]\}}{\int \tau^{a_0-1}\exp\{\tau[-0.5(x-\mu)^2 - b_0]\}\,d\tau}$$

$$= \frac{\tau^n \exp\{-\alpha\tau\}}{\int \tau^n \exp\{-\alpha\tau\}d\tau}, \quad \text{where } n = a_0 - 1 \text{ and } \alpha = 0.5(x-\mu)^2 + b_0.$$

Since $\int_0^\infty x^n \exp\{-\alpha x\}\,dx = \begin{cases} \frac{\Gamma(n+1)}{\alpha^{n+1}}, & (n > -1, \alpha > 0) \\ \frac{n!}{a^{n+1}}, & (n = 0,1,2,\dots, \alpha > 0) \end{cases}$,

$$p(\tau|x,\mu) = \frac{\tau^n \exp\{-\alpha\tau\}}{\frac{\Gamma(n+1)}{\alpha^{n+1}}}, \quad \text{since } a_0 > 0.$$

Question 6

Figure 6.1 shows a Markov Random Field (MRF) with two random variables X_1 and X_2, where $x_i \in \{0,1\}$. Furthermore, let $\phi_1(x_1)$ and $\phi_2(x_2)$ denote the unary potentials, and $\psi_{12}(x_1,x_2)$ denotes the pairwise potential. Given the observations over 14 trials as shown in Table 6.1, find the unknown value of $\psi_{12}(x_1 = 0, x_2 = 0)$ in the potential tables shown in Table 6.2. Show all your workings clearly.

Figure 6.1

Trial Number	Outcomes	
	X_1	X_2
1	0	0
2	1	0
3	1	1
4	1	0
5	0	0
6	0	1
7	1	1
8	0	0
9	1	0
10	1	1
11	0	0
12	0	0
13	1	0
14	1	1

Table 6.1

X_1	$\phi_1(x_1)$
0	2
1	1

X_2	$\phi_2(x_2)$
0	1
1	2

X_1	X_2	$\psi_{12}(x_1, x_2)$
0	0	$\psi_{12}(x_1 = 0, x_2 = 0)$
0	1	1
1	0	2
1	1	2

Table 6.2

Answer:

Joint probability:

$p(x_1, x_2) = \prod_n \frac{1}{Z_p} \phi_1(x_{1,n})\phi_2(x_{2,n})\psi_{12}(x_{1,n}, x_{2,n})$,

where

$Z_p = \sum_{x_1} \sum_{x_2} \phi_1(x_1)\phi_2(x_2)\psi_{12}(x_1, x_2)$.

$\ln p(x_1, x_2) = \sum_n \ln \phi(x_{1,n}) + \sum_n \ln \phi_2(x_{2,n}) + \sum_n \ln \psi_{12}(x_{1,n}, x_{2,n}) - N \ln Z_p$,

where $N = \#$ observations.

Note that:

$\sum_n \ln \phi(x_{1,n}) = \ln \phi(x_{1,1}) + \ln \phi(x_{1,2}) + \ \dots \ + \ln \phi(x_{1,n}) = \sum_{x_1} N(x_1) \ln \phi(x_1)$,

where $N(x_1)$ is # times x_1 takes a state, e.g. $x_1 = 0$.

Hence,

$\ln p(x_1, x_2)$

$= \sum_n \ln \phi(x_{1,n}) + \sum_n \ln \phi_2(x_{2,n}) + \sum_n \ln \psi_{12}(x_{1,n}, x_{2,n}) - N \ln Z_p$

$= \sum_{x_1} N(x_1) \ln \phi(x_1) + \sum_{x_2} N(x_2) \ln \phi(x_2) + \sum_{x_1} \sum_{x_2} N(x_1, x_2) \ln \psi(x_1, x_2) - N \ln Z_p$

To find: $\underset{\psi_{12}}{\operatorname{argmax}} \ln p(x_1, x_2)$,

$$\Rightarrow \frac{\partial \ln p(x_1,x_2)}{\partial \psi(x_1=0,x_2=0)} = \sum_{x_1} \sum_{x_2} \frac{N(x_1,x_2)}{\psi(x_1=0,x_2=0)} - \frac{N}{Z_p} \frac{\partial Z_p}{\partial \psi(x_1=0,x_2=0)}$$

$$= \frac{N(x_1=0,x_2=0)}{\psi(x_1=0,x_2=0)} - \frac{N\phi_1(x_1=0)\phi_2(x_2=0)}{\sum_{x_1}\sum_{x_2} \phi_1(x_1)\phi_2(x_2)\psi_{12}(x_1,x_2)} = 0$$

$$\frac{N(x_1 = 0, x_2 = 0)}{\psi(x_1 = 0, x_2 = 0)} - N \frac{\phi_1(x_1 = 0)\phi_2(x_2 = 0)}{k} = 0$$

$$k = \phi_1(x_1 = 0)\phi_2(x_2 = 0)\psi_{12}(x_1 = 0, x_2 = 0) + \phi_1(x_1 = 0)\phi_2(x_2 = 1)\psi_{12}(x_1 = 0, x_2 = 1) + \phi_1(x_1 = 1)\phi_2(x_2 = 0)\psi_{12}(x_1 = 1, x_2 = 0) + \phi_1(x_1 = 1)\phi_2(x_2 = 1)\psi_{12}(x_1 = 1, x_2 = 1)$$

From Table 1.1,

$N(x_1 = 0, x_2 = 0) = 5$,

$N(x_1 = 0, x_2 = 1) = 1$,

$N(x_1 = 1, x_2 = 0) = 4$,

$N(x_1 = 1, x_2 = 1) = 4$.

Now, we have:

$$\frac{5}{\psi(x_1 = 0, x_2 = 0)} - 14 \times \frac{(2)(1)}{(2)(1)\psi_{12}(x_1 = 0, x_2 = 0) + (2)(2)(1) + (1)(1)(2) + (1)(2)(2)} = 0$$

$$\Rightarrow \frac{5}{\psi_{12}(x_1 = 0, x_2 = 0)} - \frac{28}{2\psi_{12}(x_1 = 0, x_2 = 0) + 10} = 0$$

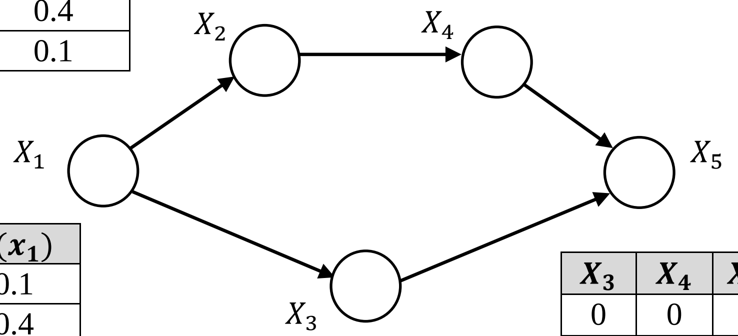
$$\Rightarrow \psi_{12}(x_1 = 0, x_2 = 0) = \frac{25}{9}$$

Question 7

The Bayesian network shown in Figure 7.1 has five random variables X_1, X_2, X_3, X_4, X_5, where $x_i \in \{0,1,2\}$ for $i = 1, 2$ and $x_i \in \{0,1\}$ for $i = 3, 4, 5$.

X_1	X_2	$p(x_2\|x_1)$
0	0	0.1
1	0	0.4
2	0	0.8
0	1	0.2
1	1	0.4
2	1	0.1

X_2	X_4	$p(x_4\|x_2)$
0	0	0.2
1	0	0.5
2	0	0.7



X_1	$p(x_1)$
0	0.1
1	0.4

X_3	X_4	X_5	$p(x_5\|x_3, x_4)$
0	0	0	0.6
0	1	0	0.3
1	0	0	0.1
1	1	0	0.8

X_1	X_3	$p(x_3\|x_1)$
0	0	0.3
1	0	0.4
2	0	0.1

Figure 7.1

(a) Given the following numbers drawn from a uniform distribution $u \sim \mathcal{U}(0,1)$:

$$u = [\quad 0.4387 \quad 0.4898 \quad 0.7513 \quad 0.4984 \quad 0.2760 \quad],$$

generate one set of samples from the joint distribution $p(x_1, x_2, x_3, x_4, x_5)$ using Gibbs sampling. Use $x_1 = 0$, $x_2 = 0$, $x_3 = 0$, $x_4 = 0$, $x_5 = 0$ as the initialization. Show all your workings clearly.

(b) Table 7.1 shows 10 sets of samples drawn from Gibbs sampling. Ignoring the burn-in effect and initialization, find the approximation for the following probabilities using the generated samples:

i. $p(x_2)$
ii. $p(x_3, x_5)$
iii. $p(x_3, x_4 = 1, x_5 = 1)$
iv. $p(x_3|x_2 = 1)$

Sample #	X_1	X_2	X_3	X_4	X_5
0	0	0	0	0	0
1	2	0	1	1	0
2	2	0	0	1	0
3	0	0	0	1	1
4	1	1	1	0	0
5	2	2	1	1	0
6	2	0	1	0	1
7	1	2	0	0	0
8	2	1	0	0	0
9	1	0	1	1	0
10	1	0	1	1	1

Table 7.1

Answer:

(a) Write down the expressions for the conditional probabilities:

$$p(x_1|x_2,x_3,x_4,x_5) = \frac{p(x_1,x_2,x_3,x_4,x_5)}{p(x_2,x_3,x_4,x_5)} = \frac{p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_3,x_4)}{p(x_4|x_2)p(x_5|x_3,x_4)\sum_{x_1} p(x_1)p(x_2|x_1)p(x_3|x_1)}$$
$$\propto p(x_1)p(x_2|x_1)p(x_3|x_1)$$

$$p(x_2|x_1,x_3,x_4,x_5) = \frac{p(x_1,x_2,x_3,x_4,x_5)}{p(x_1,x_3,x_4,x_5)} = \frac{p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_3,x_4)}{p(x_1)p(x_3|x_1)p(x_5|x_3,x_4)\sum_{x_2} p(x_2|x_1)p(x_4|x_2)}$$
$$\propto p(x_2|x_1)p(x_4|x_2)$$

$$p(x_3|x_1,x_2,x_4,x_5) = \frac{p(x_1,x_2,x_3,x_4,x_5)}{p(x_1,x_3,x_4,x_5)} = \frac{p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_3,x_4)}{p(x_1)p(x_2|x_1)p(x_4|x_2)\sum_{x_3} p(x_3|x_1)p(x_5|x_3,x_4)}$$
$$\propto p(x_3|x_1)p(x_5|x_3,x_4)$$

$$p(x_4|x_1,x_2,x_3,x_5) = \frac{p(x_1,x_2,x_3,x_4,x_5)}{p(x_1,x_3,x_4,x_5)} = \frac{p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_3,x_4)}{p(x_1)p(x_2|x_1)p(x_3|x_1)\sum_{x_4} p(x_4|x_2)p(x_5|x_3,x_4)}$$
$$\propto p(x_4|x_2)p(x_5|x_3,x_4)$$

$$p(x_5|x_1,x_2,x_3,x_4) = \frac{p(x_1,x_2,x_3,x_4,x_5)}{p(x_1,x_3,x_4,x_5)} = \frac{p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_3,x_4)}{p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)\sum_{x_5} p(x_5|x_3,x_4)}$$
$$\propto p(x_5|x_3,x_4)$$

Gibbs sampling:

t	X_1	X_2	X_3	X_4	X_5
0	0	0	0	0	0
1	1	1	0	0	0

Iteration 1:

$$p(x_1 = 0|x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 0) \propto p(x_1 = 0)p(x_2 = 0|x_1 = 0)p(x_3 = 0|x_1 = 0)$$
$$= (0.1)(0.1)(0.3) = 0.003$$

$$p(x_1 = 1|x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 0) \propto p(x_1 = 1)p(x_2 = 0|x_1 = 1)p(x_3 = 0|x_1 = 1)$$
$$= (0.4)(0.4)(0.4) = 0.064$$

$$p(x_1 = 2|x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 0) \propto p(x_1 = 2)p(x_2 = 0|x_1 = 2)p(x_3 = 0|x_1 = 2)$$
$$= (0.5)(0.8)(0.1) = 0.04$$

Normalization,

$$p(x_1 = 0|x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 0) = \frac{0.003}{0.107} = 0.028$$

$$p(x_1 = 1|x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 0) = \frac{0.064}{0.107} = 0.598$$

$$p(x_1 = 2|x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 0) = \frac{0.04}{0.107} = 0.374$$

$$\boldsymbol{u = 0.4387 \Rightarrow x_1 = 1.}$$

$$p(x_2 = 0|x_1 = 1, x_3 = 0, x_4 = 0, x_5 = 0) \propto p(x_2 = 0|x_1 = 1)p(x_4 = 0|x_2 = 0)$$
$$= (0.4)(0.2) = 0.08$$

$$p(x_2 = 1|x_1 = 1, x_3 = 0, x_4 = 0, x_5 = 0) \propto p(x_2 = 1|x_1 = 1)p(x_4 = 0|x_2 = 1)$$
$$= (0.4)(0.5) = 0.20$$

$$p(x_2 = 2|x_1 = 1, x_3 = 0, x_4 = 0, x_5 = 0) \propto p(x_2 = 2|x_1 = 1)p(x_4 = 0|x_2 = 2)$$
$$= (0.2)(0.7) = 0.14$$

Normalization,

$$p(x_2 = 0|x_1 = 1, x_3 = 0, x_4 = 0, x_5 = 0) = \frac{0.08}{0.42} = 0.191$$

$$p(x_2 = 1|x_1 = 1, x_3 = 0, x_4 = 0, x_5 = 0) = \frac{0.20}{0.42} = 0.476$$

$$p(x_2 = 2|x_1 = 1, x_3 = 0, x_4 = 0, x_5 = 0) = \frac{0.14}{0.42} = 0.333$$

$\boldsymbol{u = 0.4898 \Rightarrow x_2 = 1}$

$$p(x_3 = 0|x_1 = 1, x_2 = 1, x_4 = 0, x_5 = 0) \propto p(x_3 = 0|x_1 = 1)p(x_5 = 0|x_3 = 0, x_4 = 0)$$
$$= (0.4)(0.6) = 0.24$$

$$p(x_3 = 1|x_1 = 1, x_2 = 1, x_4 = 0, x_5 = 0) \propto p(x_3 = 1|x_1 = 1)p(x_5 = 0|x_3 = 1, x_4 = 0)$$
$$= (0.6)(0.1) = 0.06$$

Normalization,

$$p(x_3 = 0|x_1 = 1, x_2 = 1, x_4 = 0, x_5 = 0) = \frac{0.24}{0.30} = 0.8$$

$$p(x_3 = 1|x_1 = 1, x_2 = 1, x_4 = 0, x_5 = 0) = \frac{0.06}{0.30} = 0.2$$

$\boldsymbol{u = 0.7513 \Rightarrow x_3 = 0}$

$$p(x_4 = 0|x_1 = 1, x_2 = 1, x_3 = 0, x_5 = 0) \propto \ p(x_4 = 0|x_2 = 1)p(x_5 = 0|x_3 = 0, x_4 = 0)$$
$$= (0.5)(0.6) = 0.3$$

$$p(x_4 = 1|x_1 = 1, x_2 = 1, x_3 = 0, x_5 = 0) \propto \ p(x_4 = 1|x_2 = 1)p(x_5 = 0|x_3 = 0, x_4 = 1)$$
$$= (0.5)(0.3) = 0.15$$

Normalization,

$$p(x_4 = 0|x_1 = 1, x_2 = 1, x_3 = 0, x_5 = 0) = \frac{0.30}{0.45} = 0.667$$

$$p(x_4 = 1|x_1 = 1, x_2 = 1, x_3 = 0, x_5 = 0) = \frac{0.15}{0.45} = 0.333$$

$\boldsymbol{u = 0.4984 \Rightarrow x_4 = 0}$

$$p(x_5 = 0|x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 0) \propto p(x_5 = 0|x_3 = 0, x_4 = 0)$$
$$= 0.6$$

$$p(x_5 = 1|x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 0) \propto p(x_5 = 1|x_3 = 0, x_4 = 0)$$
$$= 0.4$$

No need for normalization.

$\boldsymbol{u = 0.2760 \Rightarrow x_5 = 0}$

(b)

i. $p(x_2 = 0) = \frac{6}{10} = 0.6$
$p(x_2 = 1) = \frac{2}{10} = 0.2$
$p(x_2 = 2) = \frac{2}{10} = 0.2$

ii. $p(x_3 = 0, x_5 = 0) = \frac{3}{10} = 0.3$
$p(x_3 = 0, x_5 = 1) = \frac{1}{10} = 0.1$
$p(x_3 = 1, x_5 = 0) = \frac{4}{10} = 0.4$
$p(x_3 = 1, x_5 = 1) = \frac{2}{10} = 0.2$

iii. $p(x_3 = 0, x_4 = 1, x_5 = 1) = \frac{1}{10} = 0.1$

iv. $p(x_3 = 0\ |x_2 = 1) = \frac{1}{2} = 0.5$
$p(x_3 = 1\ |x_2 = 1) = \frac{1}{2} = 0.5$

Question 8

a. Figure 8.1 shows a homogeneous hidden Markov Model (HMM) over three time steps. The latent random variables are Y_1, Y_2, Y_3, where $Y_n \in \{0, 1, 2\}$, and the observed random variables are X_1, X_2, X_3, where $X_n \in \mathbb{R}$.

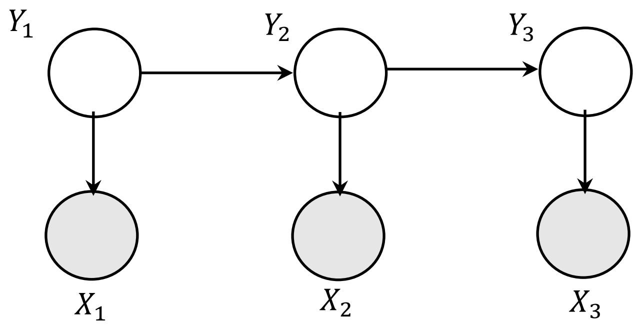


Figure 2.1

The prior probability of the random variable Y_1 is $p(Y_1 \mid \pi) = \prod_k \pi_k^{y_{1k}}$, where $\pi = \{0.2, 0.5, 0.3\}$. Furthermore, the transition probability is given by:

$$p(Y_n \mid Y_{n-1}, A) = \prod_k \prod_j A_{jk}^{y_{n-1,j}y_{nk}}, \text{ where } A = \begin{bmatrix} 0.2 & \alpha & \beta \\ 0.1 & 0.6 & 0.3 \\ 0.4 & 0.5 & 0.1 \end{bmatrix}, \text{ and}$$

the emission probabilities of the respective observed random variables X_n are shown in Table 8.1.

	$k = 0$	$k = 1$	$k = 2$
X_1	0.3	0.6	0.4
X_2	0.5	0.4	0.4
X_3	0.3	0.8	0.5

Table 8.1

Given that the minimum probability of the joint distribution $p(Y_1, Y_2, Y_3, X_1, X_2, X_3)$ is 0.000216 and occurs at $Y_1 = 0, Y_2 = 1, Y_3 = 0$, find the unknown values α and β in the transition probability.

b. Figure 8.2 shows an undirected graphic model with six random variables X_1, X_2, X_3, X_4, X_5 and X_6, where $X_i \in \{0,1,2\}$. The potential $\psi(X_i, X_j)$ between any pair of nodes X_i and X_j, where $i < j$ is given in Table 2.2. Given $X_1 = 0, X_3 = 1$ and $X_5 = 2$, find the states of X_2, X_4 and X_6 that maximizes the joint distribution $p(X_1, X_2, X_3, X_4, X_5, X_6)$.

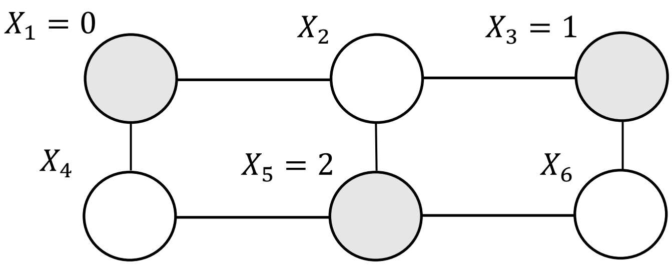


Figure 8.2

X_i	X_j	$\psi(X_i, X_j)$
0	0	1
0	1	5
0	2	7

1	0	2
1	1	4
1	2	8
2	0	3
2	1	6
2	2	9

Table 8.2

Answer:

a.

Joint probability: $p(X, Y) = p(Y_1) \prod_{n=2} p(Y_n \mid Y_{n-1}) \prod_{n=1} p(X_n \mid Y_n)$

$$\min_Y p(X, Y) = \min_Y p(Y_1) \prod_{n=2} p(Y_n \mid Y_{n-1}) \prod_{n=1} p(X_n \mid Y_n)$$

$$= \min_{Y_1} \min_{Y_2} \min_{Y_3} p(Y_1) p(Y_2 \mid Y_1) p(Y_3 \mid Y_2) p(X_1 \mid Y_1) p(X_2 \mid Y_2) p(X_3 \mid Y_3)$$

$$= \min_{Y_3} p(X_3 \mid Y_3) \min_{Y_2} p(X_2 \mid Y_2) p(Y_3 \mid Y_2) \min_{Y_1} p(Y_1) p(X_1 \mid Y_1) p(Y_2 \mid Y_1)$$

Given that the minimum probability equals 0.000216 and occurs at $Y_1 = 0, Y_2 = 1, Y_3 = 0$, this imply:

$$\min_Y p(X, Y) = 0.000216$$

$$p(Y_1 = 0) p(Y_2 = 1 \mid Y_1 = 0) p(Y_3 = 0 \mid Y_2 = 1) p(X_1 \mid Y_1 = 0) p(X_2 \mid Y_2 = 1) p(X_3 \mid Y_3 = 0)$$

$$= (0.2)(\alpha)(0.1)(0.3)(0.4)(0.3) = 0.00072\alpha = 0.000216 \Longrightarrow \alpha = 0.3$$

Since each row of the transition matrix sums to one, we have $0.2 + \alpha + \beta = 1 \Rightarrow \beta = 0.5$

b.

Joint probability:

$$p(X) = \frac{1}{Z} \psi(X_1 = 0, X_2) \psi(X_1 = 0, X_4) \psi(X_2, X_3 = 1) \psi(X_2, X_5 = 2) \psi(X_3 = 1, X_6) \psi(X_4, X_5 = 2) \psi(X_5 = 2, X_6)$$

$$\max_{X_2} \max_{X_4} \max_{X_6} p(X) =$$

$\max_{X_2} \psi(X_1 = 0, X_2)\psi(X_2, X_3 = 1)\psi(X_2, X_5 = 2) \max_{X_4} \psi(X_1 = 0, X_4)\psi(X_4, X_5 = 2)$

$\max_{X_6} \psi(X_3 = 1, X_6)\psi(X_5 = 2, X_6)$

Consider

$\max_{X_6} \psi(X_3 = 1, X_6)\psi(X_5 = 2, X_6) = \max[(2)(3), (4)(6), (8)(9)] = \max[6, 24, 72]$

$= 72 \ \ (\boldsymbol{X_6 = 2})$

$\max_{X_4} \psi(X_1 = 0, X_4)\psi(X_4, X_5 = 2) = \max[(1)(7), (5)(8), (7)(9)] = \max[\ 7{,}40{,}63]$

$= 63\ (\boldsymbol{X_4 = 2})$

$\max_{X_2} \psi(X_1 = 0, X_2)\psi(X_2, X_3 = 1)\psi(X_2, X_5 = 2)$

$= \max[(1)(5)(7), (5)(4)(8), (7)(6)(9)] = \max[35{,}160{,}378] = 378 \ \ (\boldsymbol{X_2 = 2})$

Question 9

Figure 9.1 shows a Bayesian network with both binary and continuous state latent random variables, i.e., $Z \in \{0,1\}$ and $T \in \mathbb{R}$. In addition, $X = 0.5$ is the observed random variable. The maximum log-likelihood of T:

$$\underset{T}{\operatorname{argmax}} \log p(T \mid X),$$

can be obtained from the Expectation-Maximization (EM) algorithm. The EM algorithm iterates between the Expectation step that evaluates the expected complete data log-likelihood with respect to $p(Z \mid X, T^{old})$ and the Maximization step that maximizes T over the expected complete data log-likelihood with respect to $p(Z \mid X, T^{old})$. T^{old} is the value of T from the previous iteration of the EM algorithm. $\{\lambda = 0.1, w_{a0} = 0.5, w_{a1} = 0.5, w_{b0} = 0.8, w_{b1} = 0.2, \tau_a = 1.0, \tau_b = 1.2, U = 0.6\}$ are known hyperparameters of the following distributions:

$p(Z) = \lambda^Z (1 - \lambda)^{(1-Z)}$,

$p(X \mid T, Z) = \mathcal{N}(X \mid w_{a0} + w_{a1}T, \tau_a)^Z \mathcal{N}(X \mid w_{b0} + w_{b1}T, \tau_b)^{(1-Z)}$,

$\mathcal{N}(X \mid w_0 + w_1 T, \tau) = \sqrt{\frac{\tau}{2\pi}} \exp\{-0.5\tau(X - w_0 - w_1 T)^2\}$,

$p(T) = U$.

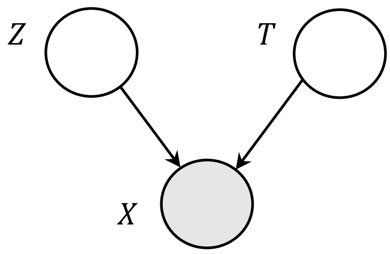


Figure 9.1

a. Derive the expression for the posterior $p(Z \mid X, T^{old})$ from the Bayesian Network.

b. Derive the expression for T that maximizes the expected complete data log-likelihood with respect to $p(Z \mid X, T^{old})$.

c. Given the initial value of $T = 2.0$, find the the value of T in the next EM iteration.

Answer:

Joint distribution:

$$p(X, Z, T) = p(T)\, p(Z) p(\, X \mid T, Z\,) = \; p(Z) p(\, X \mid T, Z\,)$$

a. $p(\, Z \mid X, T^{old}\,) = \frac{p(Z)p(X|T,Z)}{\sum_Z p(Z)p(X|T,Z)} = \frac{p(Z)p(X|T,Z)}{\sum_Z p(Z)p(X|T,Z)}$

$$= \frac{\lambda^Z (1-\lambda)^{(1-Z)} \mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,)^Z \mathcal{N}(\, X \mid w_{b0} + w_{b1}T, \tau_b\,)^{(1-Z)}}{\sum_Z \lambda^Z (1-\lambda)^{(1-Z)} \mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,)^Z \mathcal{N}(\, X \mid w_{b0} + w_{b1}T, \tau_b\,)^{(1-Z)}}$$

$$= \frac{\lambda^Z (1-\lambda)^{(1-Z)} \mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,)^Z \mathcal{N}(\, X \mid w_{b0} + w_{b1}T, \tau_b\,)^{(1-Z)}}{\lambda \mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,) + \; (1-\lambda) \mathcal{N}(\, X \mid w_{b0} + w_{b1}T, \tau_b\,)}$$

Let's define:

$$\gamma(Z=0) = p(\, Z=0 \mid X, T^{old}\,)$$

$$= \frac{(1-\lambda) \mathcal{N}(\, X \mid w_{b0} + w_{b1}T, \tau_b\,)}{\lambda \mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,) + \; (1-\lambda) \mathcal{N}(\, X \mid w_{b0} + w_{b1}T, \tau_b\,)}$$

$$\gamma(Z=1) = p(\, Z=1 \mid X, T^{old}\,)$$

$$= \frac{\lambda \mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,)}{\lambda \mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,) + \; (1-\lambda) \mathcal{N}(\, X \mid w_{b0} + w_{b1}T, \tau_b\,)}$$

b. $Q = \sum_Z p(Z \mid X, T^{old}) \ln p(X, Z \mid T)$

$= \sum_Z \gamma(Z) \ln p(Z) p(X \mid T, Z)$

$= \gamma(Z=0) \ln p(Z=0) p(X \mid T, Z=0) + \gamma(Z=1) \ln p(Z=1) p(X \mid T, Z=1)$

$= \gamma(Z=0) \ln\{(1-\lambda)\mathcal{N}(X \mid w_{b0} + w_{b1}T, \tau_b)\} +$

$\gamma(Z=1) \ln\{\lambda \mathcal{N}(X \mid w_{a0} + w_{a1}T, \tau_a)\}$

$$\underset{T}{\operatorname{argmax}}\, \gamma(Z=0) \ln\{(1-\lambda)\mathcal{N}(X \mid w_{b0} + w_{b1}T, \tau_b)\} + \gamma(Z=1) \ln\{\lambda \mathcal{N}(X \mid w_{a0} + w_{a1}T, \tau_a)\}$$

$$0 = \gamma(Z=0) \frac{\partial}{\partial T}\{\ln(1-\lambda) + \ln \mathcal{N}(X \mid w_{b0} + w_{b1}T, \tau_b)\} + \gamma(Z=1) \frac{\partial}{\partial T}\{\ln \lambda + \ln \mathcal{N}(X \mid w_{a0} + w_{a1}T, \tau_a)\}$$

$$0 = \gamma(Z=0) \frac{\partial}{\partial T}\left\{\ln \sqrt{\frac{\tau_b}{2\pi}} - 0.5\tau_b(X - w_{b0} - w_{b1}T)^2\right\} + \gamma(Z=1) \frac{\partial}{\partial T}\left\{\ln \sqrt{\frac{\tau_a}{2\pi}} - 0.5\tau_a(X - w_{a0} - w_{a1}T)^2\right\}$$

$0 = \gamma(Z=0)\{w_{b1}\tau_b(X - w_{b0} - w_{b1}T)\} + \gamma(Z=1)\{w_{a1}\tau_a(X - w_{a0} - w_{a1}T)\}$

$= X(\gamma(Z=0)w_{b1}\tau_b) - w_{b0}(\gamma(Z=0)w_{b1}\tau_b) - w_{b1}T(\gamma(Z=0)w_{b1}\tau_b) +$

$X(\gamma(Z=1)w_{a1}\tau_a) - w_{a0}(\gamma(Z=1)w_{a1}\tau_a) - w_{a1}T(\gamma(Z=1)w_{a1}\tau_a)$

$= X(\gamma(Z=0)w_{b1}\tau_b + \gamma(Z=1)w_{a1}\tau_a) - T(\gamma(Z=0)w_{b1}^2\tau_b + \gamma(Z=0)w_{a1}^2\tau_a) -$

$\gamma(Z=0)w_{b0}w_{b1}\tau_b - \gamma(Z=1)w_{a0}w_{a1}\tau_a$

$T\big(\gamma(Z=0)w_{b1}^2\tau_b + \gamma(Z=0)w_{a1}^2\tau_a\big) =$

$X(\gamma(Z=0)w_{b1}\tau_b + \gamma(Z=1)w_{a1}\tau_a) - \gamma(Z=0)w_{b0}w_{b1}\tau_b - \gamma(Z=1)w_{a0}w_{a1}\tau_a$

$$T = \frac{X(\gamma(Z=0)w_{b1}\tau_b + \gamma(Z=1)w_{a1}\tau_a) - \gamma(Z=0)w_{b0}w_{b1}\tau_b - \gamma(Z=1)w_{a0}w_{a1}\tau_a}{\gamma(Z=0)w_{b1}^2\tau_b + \gamma(Z=0)w_{a1}^2\tau_a}$$

c.

$\mathcal{N}(X \mid w_{b0} + w_{b1}T, \tau_b) = \mathcal{N}(X = 0.5 \mid 0.8 + (0.2)(2.0), 1.2)$

$$= \mathcal{N}(\, X = 0.5 \mid 1.2, 1.2\,)$$

$$= \sqrt{\frac{1.2}{2\pi}} \exp\{-0.5(1.2)(0.5-1.2)^2\} = 0.3257$$

$$\mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,) = \mathcal{N}(\, X = 0.5 \mid 0.5 + (0.5)(2.0), 1.0\,)$$

$$= \mathcal{N}(\, X = 0.5 \mid 1.5, 1.0\,)$$

$$= \sqrt{\frac{1.0}{2\pi}} \exp\{-0.5(1.0)(0.5-1.5)^2\} = 0.242$$

$$\gamma(Z=0) = p(\, Z = 0 \mid X, T^{old}\,)$$

$$= \frac{(1-\lambda)\mathcal{N}(\, X \mid w_{b0} + w_{b1}T, \tau_b\,)}{\lambda\mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,) + \ (1-\lambda)\mathcal{N}(\, X \mid w_{b0} + w_{b1}T, \tau_b\,)}$$

$$= \frac{(1-0.1)0.3257}{(0.1)0.242 + \ (1-0.1)0.3257} = \frac{0.29313}{0.31733} = 0.924$$

$$\gamma(Z=1) = p(\, Z = 1 \mid X, T^{old}\,)$$

$$= \frac{\lambda\mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,)}{\lambda\mathcal{N}(\, X \mid w_{a0} + w_{a1}T, \tau_a\,) + \ (1-\lambda)\mathcal{N}(\, X \mid w_{b0} + w_{b1}T, \tau_b\,)}$$

$$= \frac{(0.1)0.242}{(0.1)0.242 + \ (1-0.1)0.3257} = \frac{0.0242}{0.31733} = 0.076$$

$$X(\gamma(Z=0)w_{b1}\tau_b + \gamma(Z=1)w_{a1}\tau_a)$$

$$= (0.5)\{(0.924)(0.2)(1.2) + (0.076)(0.5)(1.0)\} = 0.130$$

$$-\gamma(Z=0)w_{b0}w_{b1}\tau_b - \gamma(Z=1)w_{a0}w_{a1}\tau_a$$

$$= -(0.924)(0.8)(0.2)(1.2) - (0.076)(0.5)(0.5)(1.0) = -0.196$$

$$\gamma(Z=0)w_{b1}^2\tau_b + \ \gamma(Z=0)w_{a1}^2\tau_a = (0.924)(0.04)(1.2) + (0.076)(0.25)(1.0) = 0.063$$

$$T = \frac{X(\gamma(Z=0)w_{b1}\tau_b + \gamma(Z=1)w_{a1}\tau_a) - \gamma(Z=0)w_{b0}w_{b1}\tau_b - \gamma(Z=1)w_{a0}w_{a1}\tau_a}{\gamma(Z=0)w_{b1}^2\tau_b + \gamma(Z=0)w_{a1}^2\tau_a}$$

$$= \frac{0.130 - 0.196}{0.063} = \mathbf{-1.048}$$

Question 10

a. The objective of image denoising is to recover the clean image (noise-free) from a given noisy image. Figure 10.1 shows a Markov Random Field (MRF) to solve a four-pixel binary image denoising problem. The latent random variable $X_i \in \{-1, +1\}$ represents the pixels of the desired clean image, and the observed random variables $Y_i \in \{-1, +1\}$ represents the pixels of the noisy image. We use the Ising model, i.e., $\psi(X_i, X_j) = \exp(JX_iX_j)$ as the edge potentials, where J is the coupling strength of the smoothness prior between neighboring pixels X_i and X_j. The observation model follows a Gaussian distribution: $p(Y_i \mid X_i) = \mathcal{N}(Y_i \mid X_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-0.5\frac{(Y_i - X_i)^2}{\sigma^2}\right\}$.

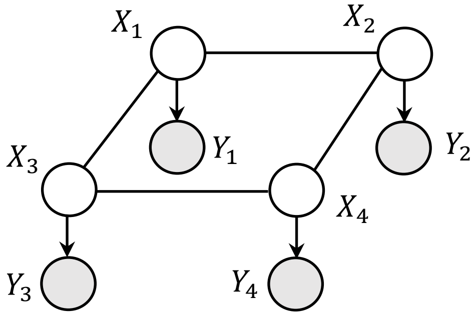


Figure 10.1

Given that we observe $Y_1 = +1, Y_2 = -1, Y_3 = -1, Y_4 = +1$, and the following random numbers are drawn from a uniform distribution $u \sim \mathcal{U}(0,1)$:

$$u = [\ 0.6557 \quad 0.0357 \quad 0.9340 \quad 0.8491\ ],$$

generate one set of samples from the joint distribution $p(X, Y)$ using Gibbs sampling. Use $X_1 = -1, X_2 = -1, X_3 = -1, X_4 = -1$ as the initialization and set $J = 0.01$, $\sigma^2 = 1.0$. Show all workings clearly.

b. Draw the Bayesian Network and write down the factorized joint probability distribution that encodes all the following conditional independences:

1. $X_4 \perp \{X_1, X_2, X_5\} \mid X_3$
2. $X_5 \perp \{X_1, X_3, X_4\} \mid X_2$
3. $X_3 \perp X_5 \mid \{X_1, X_2\}$

4. $X_1 \perp X_2 \mid \emptyset$
5. $X_1 \perp \{X_2, X_5\} \mid \emptyset$

Answers:

Joint probability: $p(X,Y) = \frac{1}{Z} \prod_i p(Y_i \mid X_i) \prod_{j \in \mathrm{nbr}(i)} \psi(X_i, X_j)$

$$= \frac{1}{Z} p(Y_1 \mid X_1) p(Y_2 \mid X_2) p(Y_3 \mid X_3) p(Y_4 \mid X_4) \psi(X_1, X_2) \psi(X_1, X_3) \psi(X_3, X_4) \psi(X_2, X_4)$$

Conditional distribution for each pixel X_i:

$$p(X_i \mid X_{\setminus i}, Y) = \frac{\frac{1}{Z} \prod_i p(Y_i \mid X_i) \prod_{j \in \mathrm{nbr}(i)} \psi(X_i, X_j)}{\sum_{X_i} \frac{1}{Z} \prod_i p(Y_i \mid X_i) \prod_{j \in \mathrm{nbr}(i)} \psi(X_i, X_j)}$$

$$= \frac{p(Y_i \mid X_i) \prod_{j \in \mathrm{nbr}(i)} \psi(X_i, X_j)}{p(Y_i \mid X_i = -1) \prod_{j \in \mathrm{nbr}(i)} \psi(X_i = -1, X_j) + p(Y_i \mid X_i = +1) \prod_{j \in \mathrm{nbr}(i)} \psi(X_i = +1, X_j)}$$

$$p(X_i = +1 \mid X_{\setminus i}, Y)$$
$$= \frac{p(Y_i \mid X_i = +1) \prod_{j \in \mathrm{nbr}(i)} \psi(X_i = +1, X_j)}{p(Y_i \mid X_i = -1) \prod_{j \in \mathrm{nbr}(i)} \psi(X_i = -1, X_j) + p(Y_i \mid X_i = +1) \prod_{j \in \mathrm{nbr}(i)} \psi(X_i = +1, X_j)}$$

$$= \frac{\mathcal{N}(Y_i \mid +1, \sigma^2) \prod_{j \in \mathrm{nbr}(i)} \exp(J x_j)}{\mathcal{N}(Y_i \mid -1, \sigma^2) \prod_{j \in \mathrm{nbr}(i)} \exp(-J x_j) + \mathcal{N}(Y_i \mid +1, \sigma^2) \prod_{j \in \mathrm{nbr}(i)} \exp(J x_j)}$$

$$= \frac{1}{1 + \frac{\mathcal{N}(Y_i \mid -1, \sigma^2)}{\mathcal{N}(Y_i \mid +1, \sigma^2)} \prod_{j \in \mathrm{nbr}(i)} \exp(-2 J x_j)}$$

$$p(X_i = -1 \mid X_{\setminus i}, Y) = 1 - p(X_i = +1 \mid X_{\setminus i}, Y)$$

$$\mathcal{N}(\,Y_i = +1 \mid +1, \sigma^2\,) = \frac{1}{\sqrt{2\pi 1^2}} \exp\left\{-0.5\,\frac{(1-1)^2}{1^2}\right\} = \frac{1}{\sqrt{2\pi}}$$

$$\mathcal{N}(\,Y_i = +1 \mid -1, \sigma^2\,) = \frac{1}{\sqrt{2\pi 1^2}} \exp\left\{-0.5\,\frac{(1+1)^2}{1^2}\right\} = \frac{1}{\sqrt{2\pi}} \exp\{-2.0\}$$

$$\mathcal{N}(\,Y_i = -1 \mid +1, \sigma^2\,) = \frac{1}{\sqrt{2\pi 1^2}} \exp\left\{-0.5\,\frac{(-1-1)^2}{1^2}\right\} = \frac{1}{\sqrt{2\pi}} \exp\{-2.0\}$$

$$\mathcal{N}(\,Y_i = -1 \mid -1, \sigma^2\,) = \frac{1}{\sqrt{2\pi 1^2}} \exp\left\{-0.5\,\frac{(-1+1)^2}{1^2}\right\} = \frac{1}{\sqrt{2\pi}}$$

Consider $\boldsymbol{X_1}$:

Markov blanket: $X_2 = -1, X_3 = -1, Y_1 = +1,$

$$p\big(X_1 = +1 \mid X_{\setminus 1}, Y\big) = \frac{1}{1 + \frac{\mathcal{N}(Y_1 \mid -1, \sigma^2)}{\mathcal{N}(Y_1 \mid +1, \sigma^2)} \prod_{j \in \mathrm{nbr}(1)} \exp(-2Jx_j)}$$

$$= \frac{1}{1 + \frac{\mathcal{N}(Y_1 = +1 \mid -1, \sigma^2)}{\mathcal{N}(Y_1 = +1 \mid +1, \sigma^2)} \exp(-2J(x_2 + x_3))}$$

$$= \frac{1}{1 + \exp\{-2.0\} \exp\{-(2)(0.01)(-2)\,\}\,)}$$

$$= \frac{1}{1 + \exp\{-2.0 + 0.04\}\,)} = \frac{1}{1 + \exp\{-1.96\}\,)} = \mathbf{0.8765}$$

Since $u_1 = 0.6557 < p\big(X_1 = +1 \mid X_{\setminus 1}, Y\big) \Rightarrow \boldsymbol{X_1} = +\mathbf{1}$

Consider $\boldsymbol{X_2}$:

Markov blanket: $X_1 = +1,\ X_4 = -1, Y_2 = -1,$

$$p\big(X_2 = +1 \mid X_{\setminus 2}, Y\big) = \frac{1}{1 + \frac{\mathcal{N}(Y_2 \mid -1, \sigma^2)}{\mathcal{N}(Y_2 \mid +1, \sigma^2)} \prod_{j \in \mathrm{nbr}(2)} \exp(-2Jx_j)}$$

$$= \frac{1}{1 + \frac{\mathcal{N}(Y_2 = -1 \mid -1, \sigma^2)}{\mathcal{N}(Y_2 = -1 \mid +1, \sigma^2)} \exp(-2J(x_1 + x_4))}$$

$$= \frac{1}{1 + \exp\{+2.0\} \exp\{0\}\,)}$$

$$= \frac{1}{1 + \exp\{+2.0\}} = \mathbf{0.1192}$$

Since $u_2 = 0.0357 < p\big(X_2 = +1 \mid X_{\setminus 2}, Y\big) \Rightarrow \boldsymbol{X_2} = +\mathbf{1}$

Consider $\boldsymbol{X_3}$:

Markov blanket: $X_1 = +1,\ X_4 = -1, Y_3 = -1,$

$$p\big(X_3 = +1 \mid X_{\setminus 3}, Y\big) = \frac{1}{1 + \frac{\mathcal{N}(Y_3 \mid -1, \sigma^2)}{\mathcal{N}(Y_3 \mid +1, \sigma^2)} \prod_{j \in \mathrm{nbr}(3)} \exp(-2Jx_j)}$$

$$= \frac{1}{1 + \frac{\mathcal{N}(Y_3 = -1 \mid -1, \sigma^2)}{\mathcal{N}(Y_3 = -1 \mid +1, \sigma^2)} \exp(-2J(x_1 + x_4))}$$

$$= \frac{1}{1 + \exp\{+2.0\} \exp\{0\}\,)}$$

$$= \frac{1}{1 + \exp\{+2.0\}} = \mathbf{0.1192}$$

Since $u_3 = 0.9340 \geq p\big(X_3 = +1 \mid X_{\setminus 3}, Y\big) \Rightarrow \boldsymbol{X_3} = -\mathbf{1}$

Consider $\boldsymbol{X_4}$:

Markov blanket: $X_2 = +1,\ X_3 = -1, Y_4 = +1,$

$$p\big(X_4 = +1 \mid X_{\setminus 4}, Y\big) = \frac{1}{1 + \frac{\mathcal{N}(Y_4 \mid -1, \sigma^2)}{\mathcal{N}(Y_4 \mid +1, \sigma^2)} \prod_{j \in \mathrm{nbr}(4)} \exp(-2Jx_j)}$$

$$= \frac{1}{1 + \frac{\mathcal{N}(Y_4 = +1 \mid -1, \sigma^2)}{\mathcal{N}(Y_4 = +1 \mid +1, \sigma^2)} \exp(-2J(x_2 + x_3))}$$

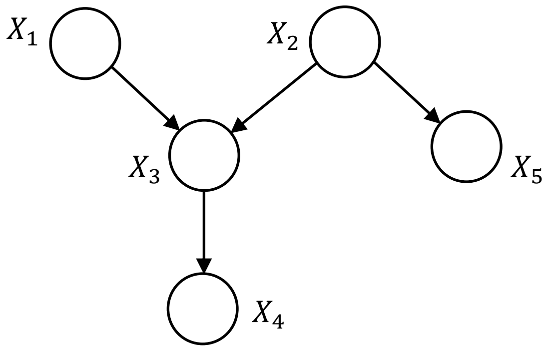
$$= \frac{1}{1 + \exp\{-2.0\} \exp\{0\}\,)}$$

$$= \frac{1}{1 + \exp\{-2.0\}} = \mathbf{0.8808}$$

Since $u_4 = 0.8491 \geq p\big(X_4 = +1 \mid X_{\setminus 4}, Y\big) \Rightarrow \boldsymbol{X_4} = -\mathbf{1}$

b.

$$p(X) = p(x_1)p(x_2)p(x_3|x_1, x_2)p(x_4|x_3)p(x_5|x_2)$$



--End--