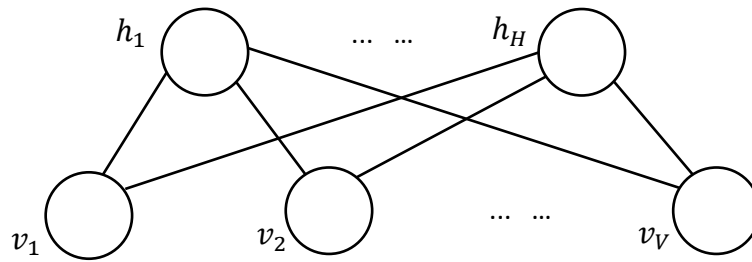


School of Computing  
National University of Singapore  
CS5340: Uncertainty Modeling in AI  
Semester 1, AY 2022/23

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  $\mathbf{v} = [v_1, \dots, v_V]^T$  and hidden variables  $\mathbf{h} = [h_1, \dots, h_H]^T$ , where all variables are binary taking states  $\{0,1\}$ . The joint distribution of the MRF is given by:

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

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

a) Given that:

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

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:

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

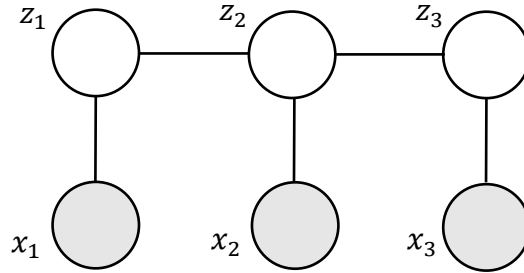
Show all your workings clearly.

- 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:

$h$	$v_1$	$v_2$	$\exp(\mathbf{v}^T \mathbf{W} \mathbf{h} + \mathbf{a}^T \mathbf{v} + b \mathbf{h})$
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.  $\theta = \{\mathbf{W}_{2 \times 1}, \mathbf{a}_{2 \times 1}, b\}$ .

## 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

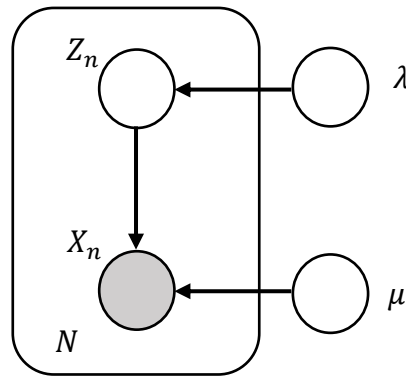
$z_2$	$\psi_e(x_2, z_2)$
$F$	7.0
$H$	1.0
$M$	2.0

$z_3$	$\psi_e(x_3, z_3)$
$F$	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:

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

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

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

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

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

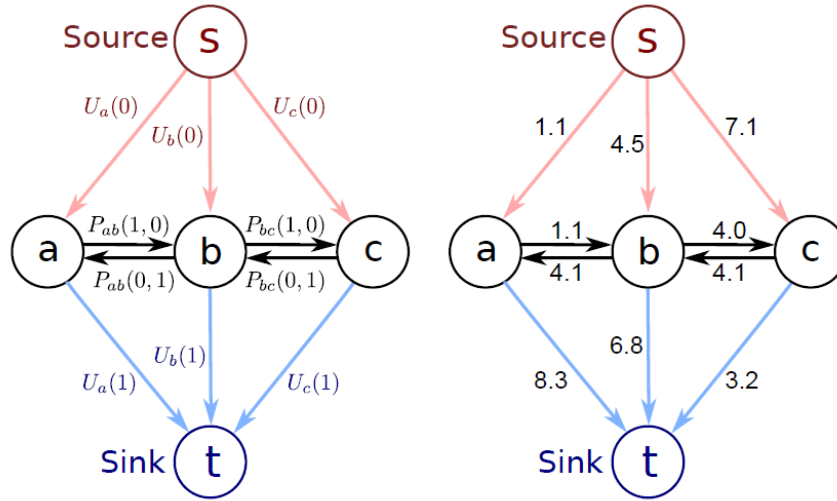
(b) Derive the responsibility  $\gamma(z_{nk}) = p(z_{nk} = 1 | x)$ , and show that the updates for the unknown parameters  $\mu$  and  $\lambda$  in the maximization step of the EM algorithm are given by:

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n,$$

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

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

- computing the cost of all eight possible solutions explicitly and finding the one with the minimum cost, and
- 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  $\mu$  and precision  $\tau$ . Suppose that the marginal distributions over the mean and precision are given by  $\mathcal{N}(\mu|\mu_0, s_0)$  and  $\text{Gam}(\tau|a, b)$ , where  $\text{Gam}(\cdot|\cdot, \cdot)$  denotes a gamma distribution. Write down expressions for the conditional 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  $p(\mu, \tau | x)$ .

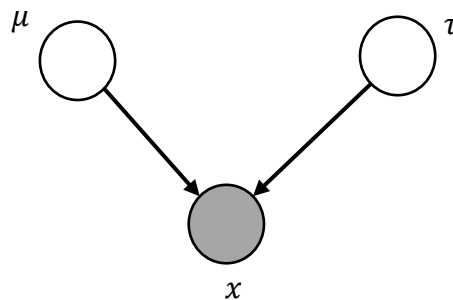


Fig. 5.1

### 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

### 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  $i = 3, 4, 5$ .

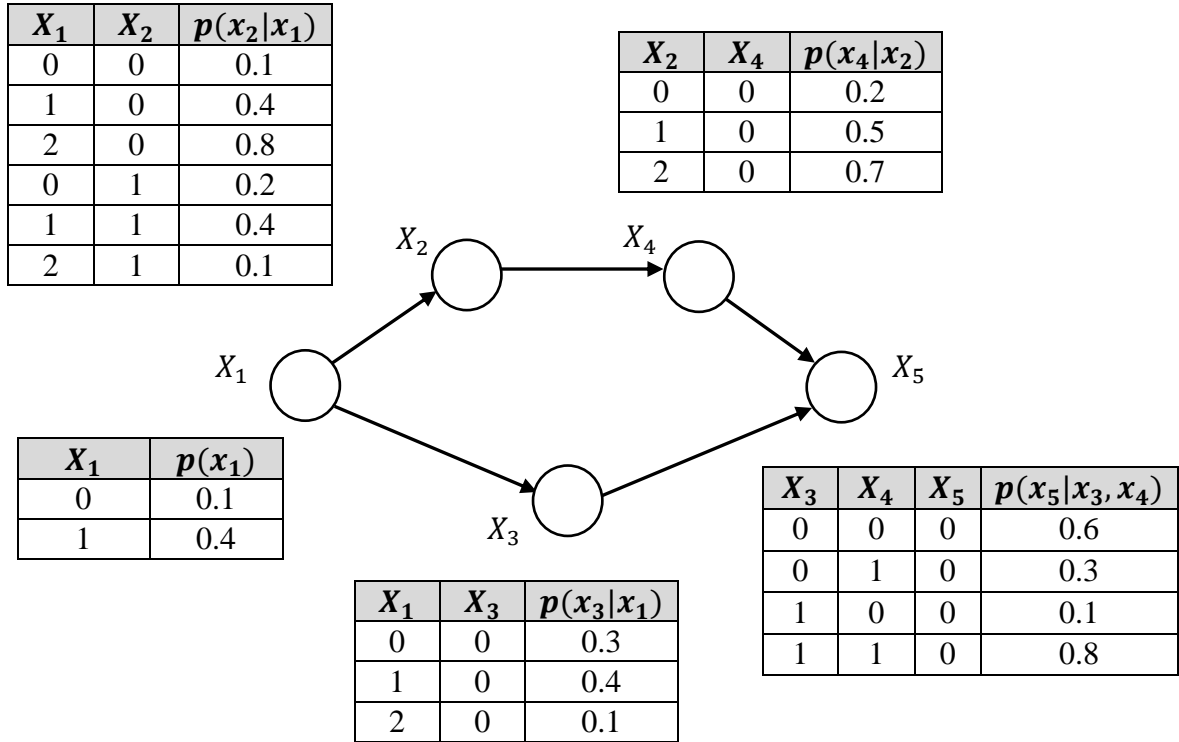


Figure 7.1

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

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

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:

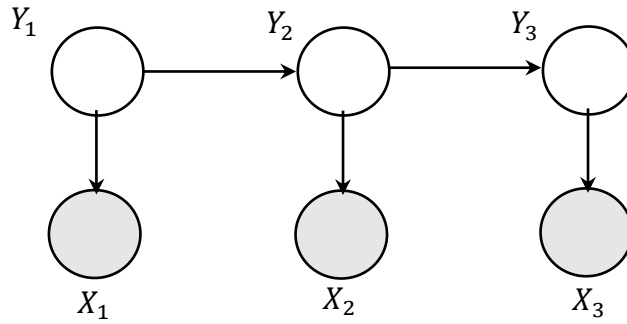
- $p(x_2)$
- $p(x_3, x_5)$
- $p(x_3, x_4 = 1, x_5 = 1)$
- $p(x_3|x_2 = 1)$

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**

### **Question 8**

- a. Figure 8.1 shows a homogeneous hidden Markov Model (HMM) over three timesteps. 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  $p(Y_1 | \pi) = \prod_k \pi_k^{y_{1k}}$ , where  $\pi = \{0.2, 0.5, 0.3\}$ . Furthermore, the transition probability is given by:

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

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

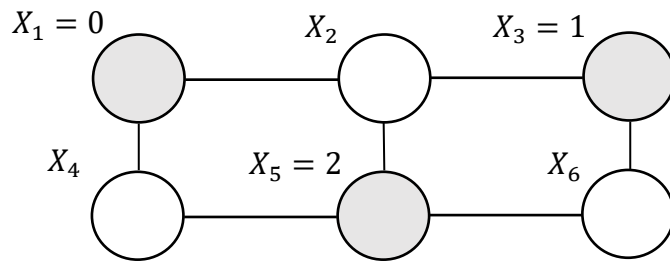


	$k = 0$	$k = 1$	$k = 2$
$X_1$	0.3	0.6	0.4
$X_2$	0.5	0.4	0.4
$X_3$	0.3	0.8	0.5

**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  $\alpha$  and  $\beta$  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**

### 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  $T$ :

$$\operatorname{argmax}_T \log p(T | X),$$

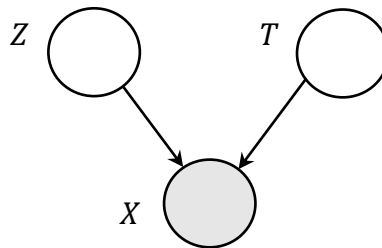
can be obtained from the Expectation-Maximization (EM) algorithm. The EM algorithm iterates between the Expectation step that evaluates the expected complete data log-likelihood with respect to  $p(Z | X, T^{old})$  and the Maximization step that maximizes  $T$  over the expected complete data log-likelihood with respect to  $p(Z | X, T^{old})$ .  $T^{old}$  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:

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

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

$$\mathcal{N}(X | w_0 + w_1T, \tau) = \sqrt{\frac{\tau}{2\pi}} \exp\{-0.5\tau(X - w_0 - w_1T)^2\},$$

$$p(T) = U.$$

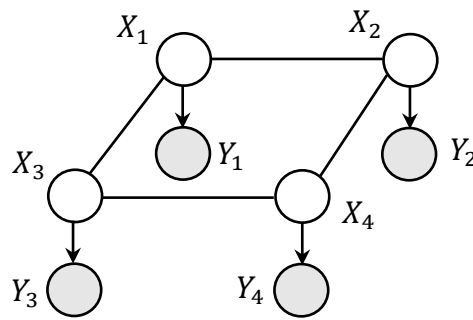


**Figure 3.1**

- Derive the expression for the posterior  $p(Z | X, T^{old})$  from the Bayesian Network.
- Derive the expression for  $T$  that maximizes the expected complete data log-likelihood with respect to  $p(Z | X, T^{old})$ .
- Given the initial value of  $T = 2.0$ , find the the value of  $T$  in the next EM iteration.

### 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:  $p(Y_i | X_i) = \mathcal{N}(Y_i | X_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-0.5 \frac{(Y_i - X_i)^2}{\sigma^2}\right\}$ .



**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  $u \sim \mathcal{U}(0,1)$ :

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

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

### Question 11

Table 11.1 shows nine observations  $\{\mathbf{x}_1, \dots, \mathbf{x}_9\}$  of 2-dimensional features  $[x, y]$ , where each observation is generated from an image of 1-out-of-3 handwritten alphabets. We further assume the sampling of each image is fully independent, and the observations given the alphabet follow a bivariate Gaussian distribution (see Equation 1). Figure 11.1 shows a plot of the nine 2-dimensional features in Table 1.1. Given a new observation  $\mathbf{x}_{\text{Test}} = [14.65, 11.00]$ , find the probability distribution of the alphabet on its corresponding image. **Explain and show all your workings clearly.**

$\mathbf{x}_n$	$[x, y]$
$\mathbf{x}_1$	[3.83, 14.48]
$\mathbf{x}_2$	[0.31, 2.06]
$\mathbf{x}_3$	[13.62, 8.89]
$\mathbf{x}_4$	[5.74, 1.35]
$\mathbf{x}_5$	[4.02, 15.69]
$\mathbf{x}_6$	[11.82, 9.88]
$\mathbf{x}_7$	[12.39, 10.8]
$\mathbf{x}_8$	[1.64, 15.22]
$\mathbf{x}_9$	[1.84, 0.68]

Table 11.1

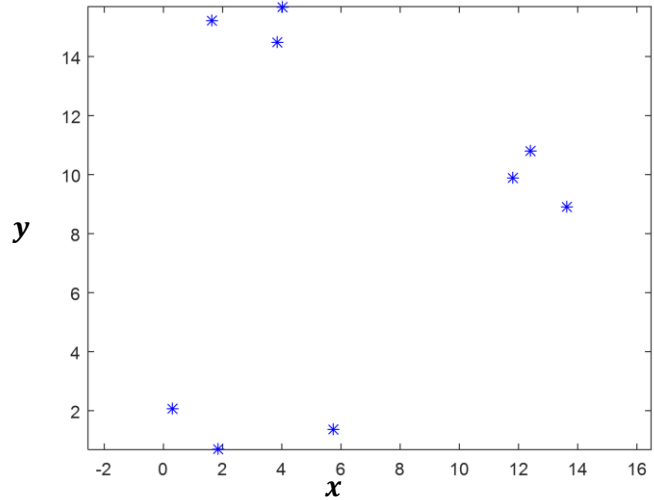


Figure 11.1

#### Useful Equations:

1.  $p(\mathbf{x}) = (2\pi)^{-1} \det(\Sigma)^{-0.5} \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\},$
2.  $\det \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21},$
3.  $\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21}} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{pmatrix}.$

### Question 12

Figure 12.1 shows a three time-step Hidden Markov Model (HMM) with binary-state latent  $Z_t \in \{0,1\}$  and observed  $X_t \in \{0,1\}$  random variables. The local conditional and prior probabilities of the HMM are shown in Table 12.1. Using variational inference, find the approximate posterior distribution of  $p(Z_1, Z_2, Z_3 | X_1, X_2, X_3)$  using the mean-field approximation, i.e.  $q(Z_1, Z_2, Z_3) = \prod_{t=1}^3 q_t(Z_t)$  in one iteration. Assume the initial value of  $q_2(Z_2 = 0) = 0.5$ , and  $X_1 = 0, X_2 = 1, X_3 = 0$ . **Explain and show all your workings clearly.**

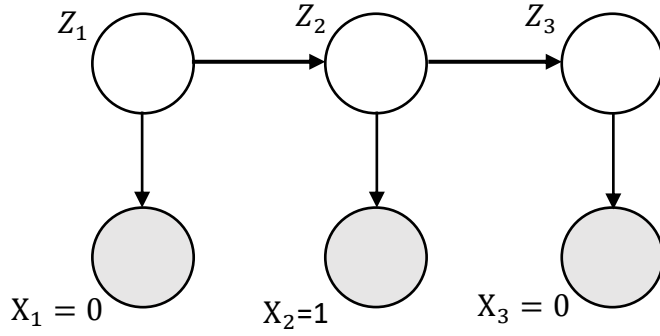


Figure 12.1

$z_1$	$p(z_1)$
0	0.2
1	0.8

$z_2$	$z_3$	$p(z_3   z_2)$
0	0	0.4
0	1	0.6
1	0	0.8
1	1	0.2

$z_1$	$z_2$	$p(z_2   z_1)$
0	0	0.3
0	1	0.7
1	0	0.9
1	1	0.1

$x_t$	$z_t$	$p(x_t   z_t)$
0	0	0.6
0	1	0.7
1	0	0.4
1	1	0.3

Table 12.1

### Question 13

Figure 13.1 shows a three-node undirected graphical model, where  $X_i \in \mathbb{R}_{\geq 0}$ ,  $\psi(X_1, X_2) = \exp\{-\alpha X_1 X_2\}$ ,  $\psi(X_i) = \exp\{-\beta X_i\}$ , and  $\alpha = 0.5$  and  $\beta = 2.5$  are constants.

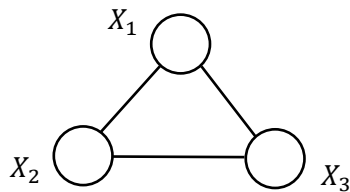


Figure 13.1.

- a) Using variational inference, find the expressions of the expectation of  $X_1, X_2$ , and  $X_3$  under  $q(X_1), q(X_2)$ , and  $q(X_3)$ , respectively, where  $q(X_1, X_2, X_3) = q(X_1)q(X_2)q(X_3)$  is the mean-field approximation of the posterior distribution  $p(X_1, X_2, X_3)$ .

(15 marks)

- b) Taking the initial expected values of  $X_2$ , and  $X_3$  under  $q(X_2)$ , and  $q(X_3)$  to be 2.0, and 1.0, respectively, find the mean-field approximation  $q(X_1, X_2, X_3)$  after one iteration.

(10 marks)

Show all your workings clearly.

Useful equations:

1.  $\int \exp\{kx\} dx = \frac{1}{k} \exp kx + \text{const}$ , where  $k \neq 0$  is a constant.
2.  $\int x \exp\{kx\} dx = \left(\frac{kx-1}{k^2}\right) \exp\{kx\}$ , where  $k \neq 0$  is a constant.

### Question 14

Figure 14.1 shows a three-node directed graphical model where  $X_i \in \{0,1\}$ . The prior and conditional probabilities are given by the following Bernoulli's distributions:

$$p(X_1 | \lambda) = \lambda^{X_1} (1 - \lambda)^{1-X_1}, \text{ where } 0 \leq \lambda \leq 1;$$

$$p(X_2 | X_1, \beta) = \beta_{X_1}^{X_2} (1 - \beta_{X_1})^{1-X_2}, \text{ where } \beta_{X_1} = \begin{cases} \beta_0, & 0 \leq \beta_0 \leq 1, \text{ if } X_1 = 0 \\ \beta_1, & 0 \leq \beta_1 \leq 1, \text{ otherwise} \end{cases};$$

$$p(X_3 | X_1, \gamma) = \gamma_{X_1}^{X_3} (1 - \gamma_{X_1})^{1-X_3}, \text{ where } \gamma_{X_1} = \begin{cases} \gamma_0, & 0 \leq \gamma_0 \leq 1, \text{ if } X_1 = 0 \\ \gamma_1, & 0 \leq \gamma_1 \leq 1, \text{ otherwise} \end{cases}.$$

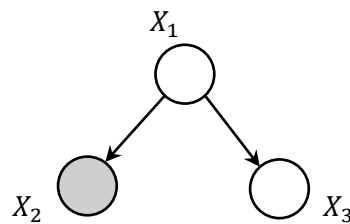


Figure 14.1.

- a) Given the observation of  $X_2 = 1$  and the initial conditions:  $\lambda = 0.5, \beta_0 = 0.2, \beta_1 = 0.6, \gamma_0 = 0.3$ , and  $\gamma_1 = 0.4$ . Draw *one set of samples* for  $X_1$  and  $X_3$  using Gibbs sampling. Assume initial samples of  $X_1 = 0, X_2 = 1$  and  $X_3 = 1$ , and  $u = [0.7655, 0.1869]$  are two numbers drawn from a uniform distribution in the range of  $[0,1]$ , i.e.,  $u \sim \text{Uniform}(0,1)$ .

(10 marks)

- b) Suppose that the subsequent samples drawn from Gibbs sampling are:  $\{[X_1 = 1, X_3 = 0], [X_1 = 0, X_3 = 1], [X_1 = 1, X_3 = 1], [X_1 = 0, X_3 = 0]\}$  under the observations of  $X_2 = \{0, 1, 1, 0\}$ , respectively. Find the updated optimal parameters  $\lambda, \beta_0, \beta_1, \gamma_0$ , and  $\gamma_1$  with all

the samples drawn from Gibbs sampling. Ignore the initial set of samples and the burn-in effect.

(15 marks)

Show all your workings clearly.

Useful equation:  $\frac{d}{dx} \ln(f(x)) = \frac{1}{f(x)} \frac{df(x)}{dx}$ .

### Question 15

Figure 15.1 shows a hidden Markov model with binary latent and observed random variables, i.e.  $Z_i \in \{0,1\}$  and  $X_i \in \{0,1\}$ . The prior, transition and emission probabilities are given in Table 3.1, 3.2 and 3.3, respectively.

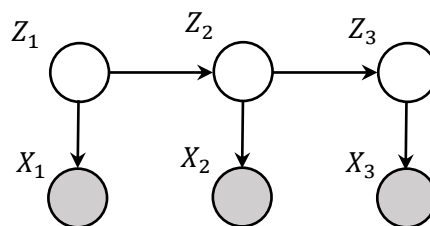


Figure. 15.1

Table. 15.1

$Z_1$	$p(Z_1)$
0	0.2
1	0.8

Table. 15.2

$Z_j$	$Z_i$	$p(Z_j   Z_i)$
0	0	0.2
1	0	0.8
0	1	0.7
1	1	0.3

Table. 15.3

$X_i$	$Z_i$	$p(X_i   Z_i)$
0	0	0.4
1	0	0.6
0	1	0.1
1	1	0.9

- a) Find the values of  $a, b$  and  $c$  that give the conditional probability of  $p(Z_2 = 0 \mid X_1 = a, X_2 = b, X_3 = c) = 0.579$ .

(15 marks)

- b) Given  $X_1 = 0, X_2 = 1, X_3 = 0$ , find  $Z_1, Z_2$  and  $Z_3$  that give the maximum joint probability of  $p(X_1, X_2, X_3, Z_1, Z_2, Z_3)$ .

(10 marks)

**Show all your workings clearly.**

--End--