CS 726: Advanced Machine Learning, Fall 2018, Mid-Semester exam

February 24, 2018.
1:30 – 3:30 pm

Roll: ______________________

Name: ______________________

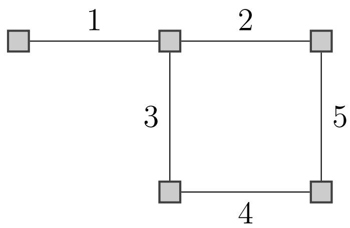
Mode: Credit/Audit/Sit-through ______________

Write all your answers in the space provided. Do not spend time/space giving irrelevant details or details not asked for. Use the marks as a guideline for the amount of time you should spend on a question. You are allowed to write elsewhere only under special circumstances like total cancellation of a previously written answer. Use the last sheet of this booklet in such cases. You are only allowed to refer your notes, no one else's notes or textbook.

---

1. Let R denote a road network graph where edges denote road segments and nodes denote intersection of two or more road segments. We are interested in modeling the congestion of road segments along time. We use $C_{e,t}$ to denote if at time t the road segment e is congested or not where e is an edge in R. We assume time t takes on discrete values e.g. $t = 1, 2, \ldots, T$. We want to model the joint distribution of congestions $C_{e,t}$ for all possible edges e in R and time t from 1 to T. Assume that congestion $C_{e,t}$ is conditionally independent of *all previous* and current congestions given the congestion at time $(t-1)$ of road segment e and road segments e' that intersect with e.

   For the example road network R below, draw the Bayesian network that can capture such conditional independence among the $C_{e,t}$ variables for $t = 1, 2, 3$. ..3 `Easy. Edges only`



`from` $C_{e,t}$ `to` $C_{e',t+1}$ `variables for` $t = 1, 2$ `and where` e, e' `are either same or they meet at an intersection.`

2. Let $P(x_1, \ldots, x_4)$ be a distribution defined over binary variables as follows

$$P(x_1, \ldots, x_4) = \frac{1}{Z} e^{x_1 \oplus x_2 \oplus x_3} e^{x_3 \oplus x_4} \tag{1}$$

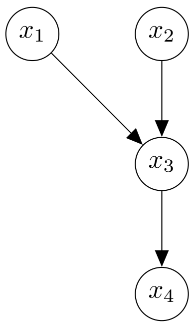
   where $\oplus$ denotes the XOR operation. XOR of two binary variables is 0 when both its arguments are the same and 1 otherwise. The value of the numerator for some of the entries have been filled in. You need to fill in the five missing entries.

x_1	x_2	x_3	x_4	$ZP(\mathbf{x})$
0	0	0	0	1
0	0	0	1	e
0	0	1	0	e^2
0	0	1	1	e
0	1	0	0	e
0	1	0	1	e^2
0	1	1	0	e
0	1	1	1	1
1	0	0	0	e
1	0	0	1	e^2
1	0	1	0	e
1	0	1	1	
1	1	0	0	
1	1	0	1	
1	1	1	0	
1	1	1	1	

..1

(a) Calculate the value of Z ..1 `Missing entries are` e, e^2 `and` e. `The value of` Z `is the sum of all table entries and is` $Z = 4e^2 + 4 + 8e$

(b) Draw a minimal Bayesian network representing the above distribution using the variable order x_1, x_2, x_3, x_4 to the right of the above table. ..2



`It is easy to see that` $x_1 \perp\!\!\!\perp x_2$ `but` x_3 `depends on both of them. Also, the form of the factorization implies that` $x_4 \perp\!\!\!\perp x_1, x_2 | x_3$.

(c) Write the CPD for $\Pr(x_1|\text{Pa}(x_1)), \Pr(x_2|\text{Pa}(x_2)), \Pr(x_3|\text{Pa}(x_3))$ in your Bayesian network above. ..3

$\Pr(x_1|\texttt{Pa}(x_1)) = \Pr(x_1)[\,0.5,\ 0.5\,]$

$\Pr(x_2|\texttt{Pa}(x_2)) = \Pr(x_2) = [\,0.5,\ 0.5\,]$

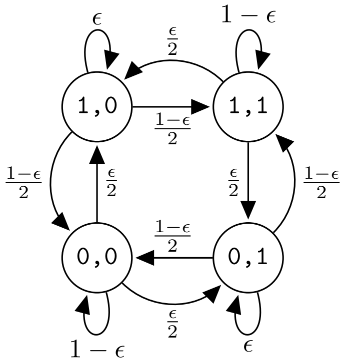
$$\Pr(x_3|x_1, x_2) = \begin{array}{cc} & \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{array}{c} 0 \\ 0 \end{array} & \begin{bmatrix} \frac{1}{1+e} & \frac{e}{1+e} & \frac{e}{1+e} & \frac{1}{1+e} \\ \frac{e}{1+e} & \frac{1}{1+e} & \frac{1}{1+e} & \frac{e}{1+e} \end{bmatrix} \end{array}$$

3. One problem with Gibbs sampling is that it cannot handle highly correlated variables very well.

(a) We will see this with an example of a distribution over two binary variables. For some small ϵ, let

$$\begin{aligned} P(x_1, x_2) &= \frac{1-\epsilon}{2} \quad \text{if} \;\; x_1 = x_2 \\ &= \frac{\epsilon}{2} \quad \text{otherwise.} \end{aligned} \tag{2}$$

Draw the Markov Chain (MC) with edge weights denoting transition probabilities between states when using Gibbs sampling on the above distribution? ..3



(b) If your MC is correct you will find that the probability of reaching from a state $(0,0)$ to $(1,1)$ is small even after two steps. To get around this problem, we will do block Gibbs sampling. Say we have a graphical model G with n variables and r edges and potentials over edges. If current sample is $\mathbf{x}^t$, we transition to the next sample by first uniformly randomly choosing one edge in G. Say, that edge is (i,j). We then sample values of x_i and x_j conditioned on the values of all other variables fixed to their values in $\mathbf{x}^t$ using $\Pr(X_i, X_j | X_{-i,-j} = \mathbf{x}^t_{-i,-j})$.

For example, in the above we will allow direct transition from the (0,0) to (1,1) state and back along with all the previous transitions. For this block Gibbs sampler, draw the modified Markov Chain and the transition probabilities for the $P(x_1, x_2)$ distribution above. ..2

All states will have direct edges to all other states and itself. The transition probability to a state (x_1, x_2) will be equal to $P(x_1, x_2)$ since there is no conditioned value left for this trivial case.

(c) We will next apply the above block Gibbs sampling on a 3×3 pairwise grid network with binary labels $0, 1$ and all edges having the same potentials of the form $\psi(x_i, x_j) = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 5 \end{smallmatrix}\right]$. For example, $\psi(1,1) = 5$ as per this potential.

A state X is the assignment of 0/1 labels to the nine variables of the network and we write them in row major order. For example, $X = 111\ 000\ 000$ denotes that the first row has all 1s and the last two rows has all 0s. Work out the following transition probabilities $T(X \to X')$ for this network.

i. $T(000\ 000\ 000 \to 000\ 011\ 000)$ ..3

The graph has 12 edges.

$$T(000\ \ 000\ \ 000 \to 000\ \ 011\ \ 000) = \frac{1}{12}\frac{52^6}{52^6 + 2^8 + 2^9 + 2^{12}}$$

ii. What is the condition on G in order for the corresponding Markov Chain to be regular. ..2

Graph has to be connected.

4. Consider the undirected graphical model below over binary variables.

$$P(x_1, \ldots, x_5) = \psi_a(x_1, x_3)\psi_b(x_2, x_3)\psi_c(x_3, x_5)\psi_d(x_4, x_5)$$

where the potentials are defined as follows. For example, the first entry says that $\psi_a(x_1 = 1, x_3 = 1) = 2^3$ and $\psi_d(x_4 = 0, x_5 = 1) = 2^2$

$\psi_a(x_1,x_3)$	$\psi_b(x_2,x_3)$	$\psi_c(x_3,x_5)$	$\psi_d(x_4,x_5)$
$\begin{bmatrix}1 & 2^2\\ 2 & 2^3\end{bmatrix}$	$\begin{bmatrix}2 & 2^2\\ 2^3 & 1\end{bmatrix}$	$\begin{bmatrix}2^4 & 2^3\\ 2^2 & 2^3\end{bmatrix}$	$\begin{bmatrix}2 & 2^2\\ 2^4 & 1\end{bmatrix}$

(a) Draw the junction tree for the above distribution calling the cliques as C_a, C_b, C_c, C_d corresponding to the four potentials above. ..1

(b) Our goal is to find the assignment of values to $x_1, \ldots, x_5$ for which the probability is maximum. For this we run the max-product message passing algorithm. Compute the minimal messages needed to be sent so that we can compute the maximizing assignment at clique node $C_d = (4, 5)$. You need to show the values of the minimal set of intermediate messages to be computed before this step can be executed by filling in the values for the question marks below. [Remember that these are messages for computing Max not Sum over variables. Also plugin numerical values for messages and not just formula.] ..4

i. Step 1: $m_{? \to ?}(?) =?$
   $m_{a\to c}(x_3) = [2, 8]$
ii. Step 2: $m_{? \to ?}(?) =?$
   $m_{b\to c}(x_3) = [8, 4]$
iii. Step 3: $m_{C_c \to C_d}(x_5) =?$
   $\max_{x_3} m_{a\to c}(x_3) m_{b\to c}(x_3) \psi_c(x_3, x_5) = [2^8, 2^8]$
iv. Step 4: Multiply message $m_{C_c \to C_d}(x_5)$ with potential ψ_4 at clique C_d and report $\text{argmax} P(x_1, \ldots, x_4)$. State the maximizing assignment on all five variables after this step.
   `argmax`$_{x_4,x_5} m_{c\to d}(x_5)\psi_d(x_4, x_5) = (x_4 = 1, x_5 = 0)$ `maximizing assignment after retracing becomes:` $x_5 = 0, x_4 = 1, x_3 = 0, x_1 = 2, x_2 = 1$

(c) Suppose now we wish to design a forward sampling step on the above graphical model by first converting it into a perfect Bayesian network and second using the messages on the sum-product message passing to initialize the potentials. Draw a *perfect* Bayesian network for P with variable x_1 as the first variable and a simplicial order after that.[Hint: Draw the undirected graphical model starting from the above potentials and convert that into a BN.] ..3

(d) In terms of above potentials and sum-product messages $m_{i\to j}(X_{s_{ij}})$ write the CPDs for node x_1, and node x_3 in this Bayesian network. ..3

$\Pr(x_1) = \sum_{x_3} \psi_1(x_1, x_3) m_{c_3\to c_1}(x_3)$

$\Pr(x_3|x_1) = \frac{\psi_1(x_1,x_3) m_{c_3\to c_1}(x_3)}{\sum_{x_3} \psi_1(x_1,x_3) m_{c_3\to c_1}(x_3)}$

(e) Provide an efficient method to compute the following expected value

$$E_P[f(x_1, \ldots, x_5)] = \sum_{x_1,x_2,\ldots x_5} f(x_1, \ldots, x_5) P(x_1, \ldots, x_5)$$

where $f(x_1, \ldots, x_5) = \sum_{i=1}^{5} x_i$ ..3

`We exploit the fact that expectation of sum is equal to sum of expectation and rewrite` the above as

$$E_P[f(x_1,\ldots,x_5)] = \sum_{x_1,x_2,\ldots x_5} x_1 P(x_1,\ldots,x_5) + \ldots + \sum_{x_1,x_2,\ldots x_5} x_5 P(x_1,\ldots,x_5) \tag{3}$$

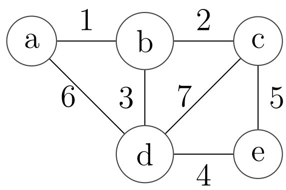
$$= \sum_{x_1} x_1 P(x_1) + \ldots + \sum_{x_5} x_5 P(x_5) \tag{4}$$

$$= P(x_1 = 1) + \ldots + P(x_5 = 1) \tag{5}$$

$$\tag{6}$$

The last equality is because the variables are binary. The marginal probabilities can be found efficiently using sum-product message passing on the junction tree.

5. A Hamiltonion cycle in an undirected graph G is a cycle that visits each vertex exactly once. For example, in the graph below a Hamiltonian cycle comprises of edges 1,2,5,4,6. This is a



well-known NP-hard problem. Reduce it to an inference problem on a suitably constructed undirected graphical model H. The number of nodes in H should be polynomial in the number of nodes and edges in G and the size of the largest potential should be a constant. [Hint: The inference task will check if the maximum probability assignment (MAP) in H has score greater than a threshold.] Specify each of the steps below to provide your reduction. ..6

(a) Nodes in H along with the set of values each node is allowed to take. Each edge in G is a node in H. Each edge variable can take one of $n+1$ labels where n is the number of nodes in G. Label 0 denotes the edge is not selected in Hamiltonian cycle. Other labels denote the order of selected edges around the cycle.

(b) Edges in H. Fully connected graph, that is, for each edge pair in G, there is an edge in H.

(c) Potentials in H over its nodes and cliques [Need not be maximal cliques] For each node e in H, $\psi_e(0) = 1$ and $\psi_e(l) = 3$ for $l > 0$.

Let adj(l, l') denote the condition that l, l' are consecutive in a circular sense (for e.g. n and 1 are consecutive and so are 1 and 2, 4 and 5 etc.).

For each pair of nodes e_i, e_j in H, define a potential that is takes a value of 1 in all but the following three cases.

$\psi_{i,j}(l, l) = 0$ if $l > 0$. This ensures that each label between 1 to n is assigned to exactly one edge.

$\psi_{i,j}(l, l') = 0$ if $l, l' > 0$ and not adj(l, l') and i, j are co-incident on a node. This condition makes sure that the two edges selected around each node have consecutive labels denoting its position in the Hamiltonian cycle.

$\psi_{i,j}(l, l') = 0$ if $l, l' > 0$ and adj(l, l') and i, j are not co-incident on a node. This makes sure that when two edges have adjacent labels, they have to be co-incident on a node. Thus, we cannot get disconnected cycles as solutions.

(d) Specify the test you will do on the MAP assignment of the above graphical model, to check if you found a Hamiltonion path in G or not. `If the MAP score is equal to` 3^n `we found a Hamiltonian cycle, otherwise not.`

Total: 40