UNIVERSITY COLLEGE LONDON

EXAMINATION FOR INTERNAL STUDENTS

MODULE CODE	:	**COMPM056**
ASSESSMENT PATTERN	:	**COMPM056C**
MODULE NAME	:	**Graphical Models (Masters Level)**
DATE	:	**09-May-14**
TIME	:	**14:30**
TIME ALLOWED	:	**2 Hours 30 Minutes**

Graphical Models, GI08/M056, 2014

Answer ALL FOUR questions.

Marks for each part of each question are indicated in square brackets

Calculators are NOT permitted

1. a. i. Show that any joint distribution $p(x_1,\ldots,x_N)$ can be expressed in the form

$$p(x_1)\prod_{n=2}^{N} p(x_n|x_1,\ldots,x_{n-1})$$

[2 marks]

ii. Give a mathematical and graphical description of a belief network (also known as a Bayesian network) and explain why it must correspond to a Directed Acyclic Graph.

[2 marks]

iii. For a joint distribution $p(x_1,\ldots,x_N)$, give a mathematical definition of $\mathcal{X}_i$ being independent of $\mathcal{X}_j$, where $\mathcal{X}_i$ and $\mathcal{X}_j$ are disjoint subsets of $\mathcal{X} = \{x_1,\ldots,x_N\}$.

[2 marks]

iv. Similarly, give a mathematical definition of $\mathcal{X}_i$ being independent of $\mathcal{X}_j$, given $\mathcal{X}_k$.

[2 marks]

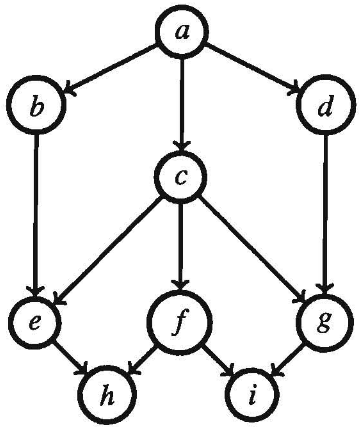
v. If x_i is independent of x_j, given x_k ($i \neq j \neq k$), does this imply that x_i and x_j are independent? Explain your answer.

[3 marks]

b. Describe an algorithm that will determine if a collection of variables $\mathcal{X}_i$ is independent of $\mathcal{X}_j$, given $\mathcal{X}_k$.

[4 marks]

c. Consider the distribution represented in the figure:



For each of the following questions, give a full explanation as to how you reached your answer.

i. Is $b \perp\!\!\!\perp d | a$?

[2 marks]

ii. Is $a \perp\!\!\!\perp h | e$?

[2 marks]

iii. Is $a \perp\!\!\!\perp h | (e, c)$?

[2 marks]

iv. Is $(b, d) \perp\!\!\!\perp (a, f) | \emptyset$?

[2 marks]

v. Is $d \perp\!\!\!\perp h | a$?

[2 marks]

[Total 25 marks]

2. a. Give a mathematical and graphical description of a Factor Graph.

[3 marks]

b. Derive the sum-product algorithm for marginal inference in singly-connected Factor Graphs and explain the computational complexity of the algorithm.

[14 marks]

c. Derive the max-product algorithm for maximal inference in singly-connected Factor Graphs.

[3 marks]

d. Consider a joint distribution with singly-connected Factor Graph over a collection of variables $\mathcal{X}$. Consider two disjoint sets of variables $\mathcal{A}$ and $\mathcal{B}$ which are both subsets of $\mathcal{X}$ such that $\mathcal{X} = \mathcal{A} \cup \mathcal{B}$. We wish to find

$$\max_{\mathcal{A}} \sum_{\mathcal{B}} p(\mathcal{X})$$

Explain if and when this can be carried out efficiently.

[5 marks]

[Total 25 marks]

3. a. You are playing a board game against your friend Randy. The board is in state x and it is your turn to play. You need to decide to make a move m from the available set of moves $\mathcal{M}(x)$; the set of available moves $\mathcal{M}(x)$ depends on the current board position x. After making your move, the board will be in a state y and it is then Randy's turn to move. He makes a move from the available set of moves $\mathcal{M}(y)$. The board is then in a new position x' and you score a number of points $u(x')$ for being in this position. The moves Randy makes can be characterised as follows: If the board is in state x and you make a move m, the board will be in the state x' with probability $p(x'|x,m)$ the next time you have to make a move. The game continues in this way until you have made a total of T moves, specifically moves $m_1,\ldots,m_T$. Mathematically, this can be described as a distribution over states of the board

$$p(x_{2:T+1}|x_1) \equiv \prod_{t=1}^{T} p(x_{t+1}|x_t,m_t)$$

The total number of points that you will be awarded for this sequence of moves is

$$\sum_{t=2}^{T+1} u(x_t)$$

Given that at timestep 1 the board is in position x_1, the optimal expected number of points you can gain by making move m_1 is

$$U(m_1) \equiv \sum_{x_2} \max_{m_2} \sum_{x_3} \ldots \max_{m_{T-1}} \sum_{x_T} \max_{m_T} \sum_{x_{T+1}} p(x_{2:T+1}|x_1)\left(u(x_2)+\ldots+u(x_T)+u(x_{T+1})\right)$$

so that the optimal move at timestep 1 is

$$\arg\max_{m_1} U(m_1)$$

Derive a recursion that will enable you to compute $U(m_1)$ in time that scales linearly with T.

[15 marks]

b. For the above game, rather than stopping after a finite time T, the game will continue indefinitely. Explain in full detail how you can modify your approach and solve the problem of determining the optimal move m from a board position x.

[10 marks]

[Total 25 marks]

4. a. The Hidden Markov Model defines a distribution on observations $v_{1:T}$ and 'hidden' variables $h_{1:T}$:

$$p(v_{1:T}, h_{1:T}) = p(v_1|h_1)p(h_1)\prod_{t=2}^{T} p(v_t|h_t)p(h_t|h_{t-1})$$

i. Derive a recursive algorithm for computing $p(h_t|v_{1:t})$.

[3 marks]

ii. Derive a recursive algorithm for computing $p(h_t|v_{1:T})$.

[3 marks]

iii. Explain how to efficiently compute $p(v_{1:T})$.

[2 marks]

iv. Explain how to compute the prediction of an (as yet) unobserved future observation $p(v_{T+1}|v_{1:T})$.

[2 marks]

v. Explain how to compute $p(h_t, h_{t+1}|v_{1:T})$.

[2 marks]

b. Given a set of training observations $\mathcal{V} = \{v^n_{1:T}, n = 1\ldots, N\}$ we wish to learn the transition A and emission matrix parameters of a Hidden Markov Model with transition distribution

$$p(h_t = i|h_{t-1} = j) = A_{ij}, \qquad i = 1,\ldots,H, \quad j = 1,\ldots,H$$

$$p(v_t = k|h_t = i) = B_{ki}, \qquad k = 1,\ldots,V, \quad i = 1,\ldots,H$$

Explain how to compute exactly the gradient of the log likelihood and describe the computational complexity of computing this gradient

$$\frac{\partial}{\partial A_{ij}} \log p(\mathcal{V}|A,B), \qquad \frac{\partial}{\partial B_{ki}} \log p(\mathcal{V}|A,B)$$

[13 marks]

[Total 25 marks]

END OF PAPER