UNIVERSITY COLLEGE LONDON

EXAMINATION FOR INTERNAL STUDENTS

MODULE CODE	:	**COMPGI08**
ASSESSMENT PATTERN	:	**COMPGI08A**
MODULE NAME	:	**Graphical Models**
DATE	:	**30-Apr-12**
TIME	:	**10:00**
TIME ALLOWED	:	**2 Hours 30 Minutes**

Graphical Models, GI08, 2012

Answer THREE of FIVE questions.

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

Calculators are NOT permitted

1. a. Consider three disjoint sets of variables $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$. The notation $\mathcal{X} \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{Z}$ means that $\mathcal{X}$ is independent of $\mathcal{Y}$ given the states of variables $\mathcal{Z}$. We also write $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}$ as shorthand notation for $\mathcal{X} \perp\!\!\!\perp \mathcal{Y} \mid \emptyset$. If $\mathcal{X} \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{Z}$ is false, we write $\mathcal{X} \top\!\!\!\top \mathcal{Y} \mid \mathcal{Z}$.

   i. Show that

$$\mathcal{X} \perp\!\!\!\perp \{\mathcal{Y} \cup \mathcal{Z}\} \Rightarrow \mathcal{X} \perp\!\!\!\perp \mathcal{Y}, \qquad \mathcal{X} \perp\!\!\!\perp \{\mathcal{Y} \cup \mathcal{Z}\} \Rightarrow \mathcal{X} \perp\!\!\!\perp \mathcal{Z}$$

[3 marks]

   ii. If $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}$ and $\mathcal{Y} \perp\!\!\!\perp \mathcal{Z}$ explain what you can infer about the truth of the statement $\mathcal{X} \perp\!\!\!\perp \mathcal{Z}$.

[3 marks]

   iii. If $\mathcal{X} \top\!\!\!\top \mathcal{Y}$ and $\mathcal{Y} \top\!\!\!\top \mathcal{Z}$ explain what you can infer about the truth of the statement $\mathcal{X} \top\!\!\!\top \mathcal{Z}$.

[3 marks]

   iv. Explain if the following statement is correct:

$$\mathcal{X} \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{Z} \Rightarrow \mathcal{X} \perp\!\!\!\perp \mathcal{Y}$$

[3 marks]

   b. Consider distributions of the form

$$p(x,y,z) = p(z|y)p(y|x)p(x)$$

   i. If $x \perp\!\!\!\perp y$ and $y \perp\!\!\!\perp z$ does it follow that $x \perp\!\!\!\perp z$?

[3 marks]

   ii. If $x \top\!\!\!\top y$ and $y \top\!\!\!\top z$ does it follow that $x \top\!\!\!\top z$? You may make reference to the lectures to justify your answer.

[3 marks]

c. Explain why any distribution on variables $x_1,\ldots,x_D$ can be written in the form

$$p(x_1,\ldots,x_D) = \prod_{i=1}^{D} p(x_i|x_{i+1},\ldots,x_D)$$

with the convention that when $i > D$ the set of variables is the empty set, $\{x_{i+1},\ldots,x_D\} = \emptyset$.

[3 marks]

d. A belief network on variables $x_1,\ldots,x_D$ is a distribution of the form

$$p(x_1,\ldots,x_D) = \prod_{i=1}^{D} p(x_i|x_{\mathrm{pa}(i)})$$

where $x_{\mathrm{pa}(i)} \subset \{x_1,\ldots,x_D\}$ are the parental variables of variable x_i.

i. Draw a graphical representation of a belief network and explain what restrictions the graph must have for it to represent a belief network.

[3 marks]

ii. Explain what the absence of an edge on a belief network represents in terms of conditional independence.

[3 marks]

e. Let T be the number of possible belief networks on D variables.

i. Show that

$$T \geq 2^{D-1} \times 2^{D-2} \times \ldots \times 2 = 2^{D(D-1)/2}$$

[4 marks]

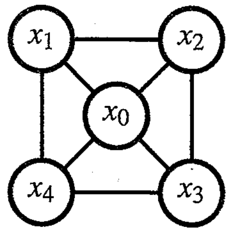
ii. Show that

$$T \leq 2^{D(D-1)/2} D!$$

[2 marks]

[Total 33 marks]

2. a. Consider the Markov network:



i. Write down the general functional form of distribution $p(x_0, x_1, x_2, x_3, x_4, x_5)$ that is consistent with the Markov network shown.

[3 marks]

ii. Write a list of independence statements the Markov network represents.

[4 marks]

iii. Draw a triangulated version of this Markov network.

[3 marks]

iv. Draw a Junction Tree for this Markov network.

[3 marks]

b. Consider a distribution on variables $\mathcal{X} = \{x_0, x_1, \ldots, x_D\}$

$$p(\mathcal{X}) = \left\{ \prod_{i=1}^{D} \phi(x_0, x_i) \right\} \phi(x_D, x_1) \prod_{j=1}^{D-1} \phi(x_j, x_{j+1})$$

i. Draw a Markov network for distributions of this form.

[3 marks]

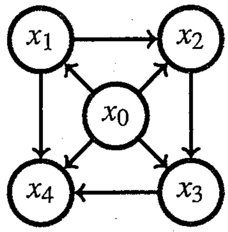
ii. Draw a Junction Tree for Markov networks of this form.

[3 marks]

iii. If all variables x_i, $i = 0, \ldots, D$ are binary, estimate the computational complexity of computing the marginal $p(x_0)$.

[4 marks]

c. Consider the belief network:



i. Write a list of independence statements this belief network represents.

[4 marks]

ii. Draw a triangulated version of this belief network.

[2 marks]

iii. Draw a Junction Tree for this belief network.

[2 marks]

iv. For binary variables x_i, $i = 0, 1, 2, 3, 4$, estimate the computational complexity of finding the marginal $p(x_4)$.

[2 marks]

[Total 33 marks]

3. The KL divergence between two distributions $q(x)$ and $p(x)$ is defined as

$$KL(q|p) = \langle \log q(x) \rangle_{q(x)} - \langle \log p(x) \rangle_{q(x)}$$

where $\langle f(x) \rangle_{q(x)}$ means taking the expectation of the function $f(x)$ with respect to the distribution $q(x)$.

a. Consider the bound

$$\log x \geq x - 1$$

and show how this can be used to prove $KL(q|p) \geq 0$.

[5 marks]

b. For a Chow-Liu belief network each variable has at most one parent. Writing $x \doteq \{x_1, \ldots, x_D\}$, we can define a Chow-Liu network as

$$q(x) = \prod_{i=1}^{D} q(x_i | x_{pa(i)}), \qquad pa(i) < i, \text{ or } pa(i) = \emptyset$$

where $pa(i)$ is the parent index of variable i. Our interest is to find the best Chow-Liu approximation to a distribution $p(x)$ based on minimising $KL(p|q)$.

i. Show that

$$KL(p|q) = \langle \log p(x) \rangle_{p(x)} - \sum_{i=1}^{D} \left\langle \log q(x_i | x_{pa(i)}) \right\rangle_{p(x_i, x_{pa(i)})}$$

[3 marks]

ii. Explain why, optimally,

$$q(x_i | x_{pa(i)}) = p(x_i | x_{pa(i)})$$

[2 marks]

iii. Using this setting, show that the dependence of $KL(p|q)$ on the parent indices $pa(i)$ is given by

$$KL(p|q) = -\sum_{i=1}^{D} \mathrm{MI}\left(x_i; x_{pa(i)}\right) + \text{const.}$$

where the mutual information is defined as

$$\mathrm{MI}\left(x_i; x_j\right) = \left\langle \log \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \right\rangle_{p(x_i, x_j)}$$

[5 marks]

iv. Hence describe an efficient algorithm to find the best Chow-Liu approximation to $p(x)$ and describe its computational complexity.

[5 marks]

c. Consider a distribution $p(v, h|\theta)$ with visible (observable) variable v, hidden (latent) variable h and parameter θ. Our interest is to find the maximum likelihood parameter setting

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \ \log p(v|\theta)$$

i. By considering $KL(q(h|v)|p(h|v,\theta)) \geq 0$, show that

$$\log p(v|\theta) \geq -\langle \log q(h|v) \rangle_{q(h|v)} + \langle \log p(v, h|\theta) \rangle_{q(h|v)}$$

[3 marks]

ii. The Expectation Maximisation (EM) algorithm is an iterative procedure that updates the parameter θ_i on iteration i by the following steps:

$$q_{i-1}(h|v) = p(h|v, \theta_{i-1})$$

$$\theta_i = \underset{\theta}{\operatorname{argmax}} \ \langle \log p(v, h|\theta) \rangle_{q_{i-1}(h|v)}$$

By defining

$$B(\theta_i, \theta_{i-1}) = -\langle \log p(h|v, \theta_{i-1}) \rangle_{p(h|v,\theta_{i-1})} + \langle \log p(v, h|\theta_i) \rangle_{p(h|v,\theta_{i-1})}$$

show that

$$\log p(v|\theta_i) = B(\theta_i, \theta_{i-1}) + KL\left(p(h|v, \theta_i)|p(h|v, \theta_{i-1})\right)$$

[5 marks]

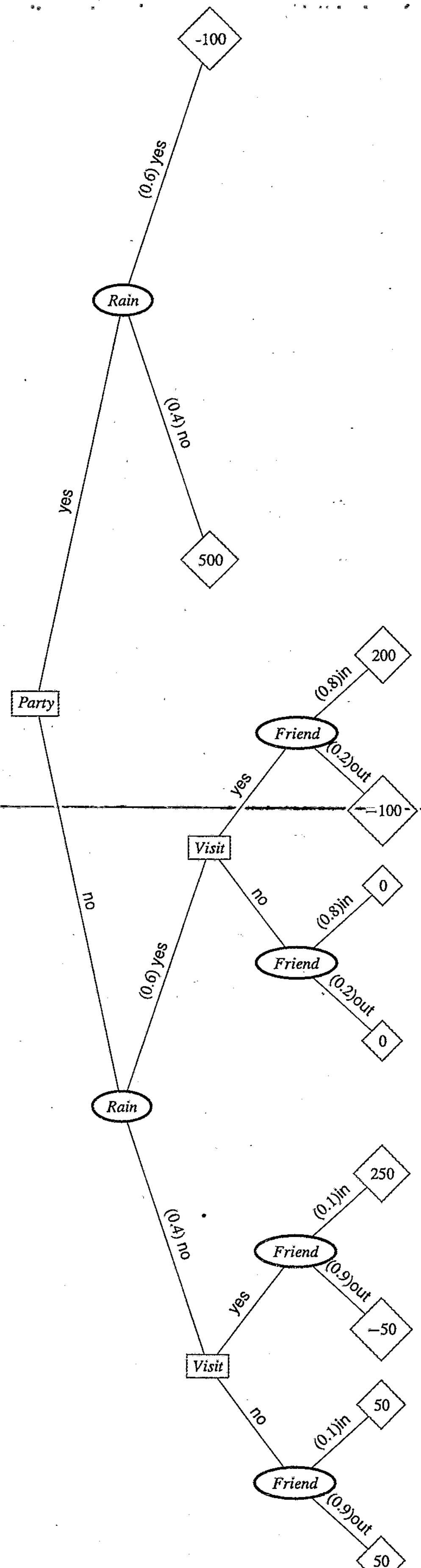
iii. Hence show that, for the EM algorithm

$$\log p(v|\theta_i) \geq \log p(v|\theta_{i-1})$$

[5 marks]

[Total 33 marks]

4. a.



The graph on the left represents a decision tree with decisions, probabilities and utilities as shown. What is the optimal expected utility at the party root node?

[8 marks]

b. Given an action ('decision') d, distribution $p(x)$ and utility $u(x,d)$, explain how to take the best action to maximise the expected utility.

[3 marks]

c. Based on your knowledge of human behaviour, give a numerical scenario that suggests that the utility of money $u(m)$ is not likely to be considered a linear function of the quantity of money m.

[5 marks]

d. A Markov Decision Process is a belief network on states s_t and actions ('decisions') a_t defined by

$$p(s_{1:T}|a_{1:T}) = p(s_1)\prod_{t=2}^{T} p(s_t|s_{t-1}, a_{t-1})$$

The rewards $r(s_t)$ are assumed only a function of the state s_t. The optimal expected reward for the MDP is given by the expression

$$r^* = \sum_{s_1} \max_{a_1} \sum_{x_2} \max_{a_2} \dots \max_{a_{T-1}} \sum_{s_T} \sum_{t=1}^{T} p(s_{1:T}|a_{1:T}) r(s_t)$$

i. Draw an influence diagram that represents this Markov Decision Process.

[3 marks]

ii. Show that by defining 'value' messages

$$v_{t-1}(s_{t-1}) = r(s_{t-1}) + \max_{a_{t-1}} \sum_{s_t} p(s_t|s_{t-1}, a_{t-1}) v_t(s_t), \qquad v_T(s_T) = r(s_T)$$

the optimal utility is given by

$$r^* = \sum_{s_1} p(s_1) v_1(s_1)$$

[8 marks]

iii. Explain what is meant by an infinite horizon $T \to \infty$ Markov Decision Process and derive a suitable set of equations that the value $v(s)$ must satisfy.

[3 marks]

iv. For infinite horizon Markov Decision Processes, describe the value iteration and policy iteration algorithms for finding $v(s)$.

[3 marks]

[Total 33 marks]

5. a. Describe in detail the max-sum algorithm for marginal inference in singly-connected factor graphs.

[10 marks]

b. Describe in detail the junction tree algorithm for marginal inference in multiply-connected belief networks.

[10 marks]

c. For the hidden Markov model

$$p(h_1)\left\{\prod_{t=2}^{T} p(h_t|h_{t-1})\right\}\left\{\prod_{t=1}^{T} p(v_t|h_t)\right\}$$

describe how the max-sum and junction tree algorithms can be applied to find the smoothed posterior marginals $p(h_t|v_{1:T})$.

[10 marks]

d. Can the max-sum algorithm on a factor graph be used to perform exact inference in a multiply-connected Markov network? Give an example to justify your answer.

[3 marks]

[Total 33 marks]

END OF PAPER