UNIVERSITY COLLEGE LONDON

EXAMINATION FOR INTERNAL STUDENTS

MODULE CODE : **COMPGI08**

ASSESSMENT PATTERN : **COMPGI08A**

MODULE NAME : **Graphical Models**

DATE : **19-Mar-09**

TIME : **14:30**

TIME ALLOWED : **2 Hours 30 Minutes**

2008/09-COMPGI08A-001-EXAM2-23

Graphical Models, GI08, 2008

Answer THREE of FIVE questions.

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

Calculators are NOT permitted

1. This question concerns the distribution

$$p(a,b,c,d,e,f) = p(a)p(b|a)p(c|b)p(d|c)p(e|d)p(f|a,e)$$

a. Draw the Directed Acyclic Graph that represents this distribution.

[4 marks]

b. Draw the moralised graph.

[2 marks]

c. Draw the triangulated graph. Your triangulated graph should contain cliques of the smallest size possible.

[5 marks]

d. Draw a Junction Tree for the above graph and verify that it satisfies the running intersection property.

[4 marks]

e. Describe a suitable initialisation of clique potentials.

[4 marks]

f. Describe the Absorption procedure and an appropriate message updating schedule.

[7 marks]

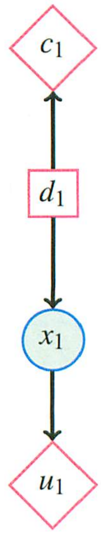
g. Show that the distribution can be expressed in the form

$$p(a|f)p(b|a,c)p(c|a,d)p(d|a,e)p(e|a,f)p(f)$$

[7 marks]

[Total 33 marks]

2. a. The Influence Diagram depicted describes the first stage of a game.



The decision variable $\text{dom}(d_1) = \{\text{play}, \text{not play}\}$, indicates the decision to either play the first stage or not. If you decide to play, there is a cost $c_1(\text{play}) = C_1$, but no cost otherwise, $c_1(\text{no play}) = 0$. The variable x_1 describes if you win or lose the game, $\text{dom}(x_1) = \{\text{win}, \text{lose}\}$, with probabilities:

$$p(x_1 = \text{win}|d_1 = \text{play}) = p_1, \qquad p(x_1 = \text{win}|d_1 = \text{no play}) = 0$$

The utility of winning/losing is

$$u_1(x_1 = \text{win}) = W_1, \qquad u_1(x_1 = \text{lose}) = 0$$

Show that the expected utility gain of playing this game is

$$U(d_1 = \text{play}) = p_1 W_1 - C_1$$

[5 marks]

b. Part (a) above describes the first stage of a new two-stage game. If you win the first stage $x_1 = \text{win}$, you have to make a decision d_2 as to whether or not play in the second stage $\text{dom}(d_2) = \{\text{play}, \text{not play}\}$. If you do not win the first stage, you cannot enter the second stage.

If you decide to play the second stage, you win with probability p_2:

$$p(x_2 = \text{win}|x_1 = \text{win}, d_2 = \text{play}) = p_2$$

If you decide not to play the second stage there is no chance to win:

$$p(x_2 = \text{win}|x_1 = \text{win}, d_2 = \text{not play}) = 0$$

The cost of playing the second stage is

$$c_2(d_2 = \text{play}) = C_2, \qquad c_2(d_2 = \text{no play}) = 0$$

and the utility of winning/losing the second stage is

$$u_2(x_2 = \text{win}) = W_2, \qquad u_2(x_2 = \text{lose}) = 0$$

i. Draw an Influence Diagram that describes this two-stage game. [5 marks]

ii. A gambler needs to decide if he should even enter the first stage of this two-stage game. Show that based on taking the optimal future decision d_2 the expected utility based on the first decision is:

$$U(d_1 = \mathsf{play}) = \begin{cases} p_1(p_2W_2 - C_2) + p_1W_1 - C_1 & \text{if } \ p_2W_2 - C_2 \geq 0 \\ p_1W_1 - C_1 & \text{if } \ p_2W_2 - C_2 \leq 0 \end{cases}$$

[12 marks]

c. You have £B in your bank account. You are asked if you would like to participate in a bet in which, if you win, your bank account will become £W. However, if you lose, your bank account will contain only £L. You win the bet with probability p_w.

i. Assuming that the utility is given by the number of pounds in your bank account, write down a formula for the expected utility of taking the bet, $U(\mathsf{bet})$ and also the expected utility of not taking the bet, $U(\mathsf{no\ bet})$. [3 marks]

ii. The above situation can be formulated differently. If you win the bet you gain £$(W - B)$. If you lose the bet you lose £$(B - L)$. Compute the expected amount of money you gain if you bet $U_{gain}(\mathsf{bet})$ and if you don't bet $U_{gain}(\mathsf{no\ bet})$. [3 marks]

iii. Show that $U(\mathsf{bet}) - U(\mathsf{no\ bet}) = U_{gain}(\mathsf{bet}) - U_{gain}(\mathsf{no\ bet})$. [5 marks]

[Total 33 marks]

3. A Hidden Markov Model (HMM) of first order is defined by the distribution

$$p(h_1,\dots,h_T,v_1,\dots,v_T) = p(h_1)p(v_1|h_1)\prod_{t=2}^{T}[p(h_t|h_{t-1})p(v_t|h_t)]$$

where each h_t, $t = 1,\dots,T$ is a discrete random variable with states $1,\dots,H$. That is $dom(h_t) = \{1,\dots,H\}$. Each variable v_t is a discrete random variable with states $1,\dots,V$. That is $dom(v_t) = \{1,\dots,V\}$.

a. Draw a Belief Network of the HMM.

[3 marks]

b. When the states of the variables $v_1,\dots,v_T$ are observed (that is, we know their states), show that the conditional distribution of the HMM can be written as

$$p(h_1,\dots,h_T|v_1,\dots,v_T) \propto \prod_{t=2}^{T}\phi_t(h_t,h_{t-1})$$

and express explicitly the potentials $\phi_t(h_t,h_{t-1})$ in terms of the HMM.

[5 marks]

c. Your task is to derive an algorithm that will find the most likely joint state

$$\arg\max_{h_1,\dots,h_T}\prod_{t=2}^{T}\phi_t(h_t,h_{t-1})$$

for arbitrarily defined potentials $\phi_t(h_t,h_{t-1})$.

i. First consider

$$\max_{h_1,\dots,h_T}\prod_{t=2}^{T}\phi_t(h_t,h_{t-1})$$

Show that how the maximisation over h_T may be pushed inside the product and that the result of the maximisation can be interpreted as a message

$$\gamma_{T-1\leftarrow T}(h_{T-1})$$

[3 marks]

ii. Derive the recursion

$$\gamma_{t-1\leftarrow t}(h_{t-1}) = \max_{h_t} \phi_t(h_t, h_{t-1}) \gamma_{t\leftarrow t+1}(h_t)$$

[7 marks]

iii. Explain how the above recursion enables the computation of

$$\arg\max_{h_1} \prod_{t=2}^{T} \phi_t(h_t, h_{t-1})$$

[3 marks]

iv. Explain how once the most likely state for h_1 is computed, one may efficiently compute the remaining optimal states $h_2, \dots, h_T$.

[4 marks]

d. A second order HMM is defined as

$$p^{HMM2}(h_1, \dots, h_T, v_1, \dots, v_T) \\ = p(h_1)p(v_1|h_1)p(h_2|h_1)p(v_2|h_2) \prod_{t=3}^{T} [p(h_t|h_{t-1}, h_{t-2})p(v_t|h_t)]$$

Following a similar approach to the first order HMM, derive explicitly a message passing algorithm to compute the most likely joint state

$$\arg\max_{h_1,\dots,h_T} p^{HMM2}(h_1, \dots, h_T|v_1, \dots, v_T)$$

[8 marks]

[Total 33 marks]

4. a. Explain why any distribution $p(x_1,\dots,x_n)$ can be expressed as a Belief Network.

[3 marks]

b. A 'trivial' Belief Network is one in which all possible links in the graph exist. Explain the meaning of a missing link in a Belief Network.

[2 marks]

c. Consider the distribution

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

i. Write $p(x|z)$ using a formula involving (all or some of) $p(z|w)$, $p(w|x,y)$, $p(x)$, $p(y)$.

[3 marks]

ii. Write $p(y|z)$ using a formula involving (all or some of) $p(z|w)$, $p(w|x,y)$, $p(x)$, $p(y)$.

[3 marks]

iii. Using the above results, derive an explicit condition for $I(x,y|z)$ (x and y to be independent, conditioned on z) and explain if this is satisfied for this distribution.

[4 marks]

d. Consider the distribution

$$p(t_1,t_2,y_1,y_2,h) = p(y_1|t_1,h)p(y_2|t_2,h)p(t_1)p(t_2)p(h)$$

i. Draw a Belief Network for this distribution.

[2 marks]

ii. Can the distribution

$$p(t_1,t_2,y_1,y_2) = \sum_h p(y_1|t_1,h)p(y_2|t_2,h)p(t_1)p(t_2)p(h)$$

be written as a (non-trivial) Belief Network?

[3 marks]

iii. Show that for $p(t_1,t_2,y_1,y_2)$ as defined above $I(t_1,y_2|\emptyset)$.

[4 marks]

e. Consider the distribution

$$p(a,b,c,d) = \phi_{ab}(a,b)\phi_{bc}(b,c)\phi_{cd}(c,d)\phi_{da}(d,a)$$

where the ϕ are potentials.

i. Draw a Markov Network for this distribution.

[2 marks]

ii. Explain if the distribution can be represented as a (non-trivial) Belief Network.

[4 marks]

iii. Derive explicitly if $I(a,c|\emptyset)$.

[3 marks]

[Total 33 marks]

5. This question concerns spam filtering. Each email is represented by a vector

$$x = (x_1, \dots, x_D)$$

where $x_i \in \{0,1\}$. Each entry of the vector indicates if a particular symbol or word appears in the email. The symbols/words are

$$money, cash, !!!, viagra, \dots, etc$$

So that $x_2 = 1$ if the word 'cash' appears in the email. The training dataset consists of a set of vectors along with the class label c, where $c = 1$ indicates the email is spam, and $c = 0$ not spam. Hence, the training set consists of a set of pairs $(x^n, c^n), n = 1, \dots, N$.

a. For the Naive Bayes model

$$p(c,x) = p(c) \prod_{i=1}^{D} p(x_i|c)$$

i. Draw a Belief Network for this distribution.

[3 marks]

ii. Derive expressions for the parameters of this model in terms of the training data using Maximum Likelihood. Assume that the data is independent and identically distributed $p(c^1, \dots, c^N, x^1, \dots, x^N) = \prod_{n=1}^{N} p(c^n, x^n)$. Explicitly, the parameters are

$$p(c=1), p(x_i = 1|c=1), p(x_i = 1|c=0), i = 1, \dots, D$$

[5 marks]

iii. Given a trained model $p(x,c)$, explain how to form a classifier $p(c|x)$.

[2 marks]

iv. If 'viagra' never appears in the spam training data, discuss what effect this will have on the classification for a new email that contains the word 'viagra'.

[2 marks]

v. Write down an expression for the decision boundary

$$p(c=1|x) = p(c=0|x)$$

and show that it can be written in the form

$$\sum_{d=1}^{D} u_d x_d - b = 0$$

for suitably defined u and b.

[6 marks]

b. The Logistic regression model is

$$p(c=1|x,w) = \sigma(\sum_{i=1}^{D} w_i x_i)$$

where

$$\sigma(s) = \frac{e^s}{1+e^s}$$

i. Write down a formula for the log-likelihood

$$L(w) = \sum_{n=1}^{N} \log p(c^n|x^n, w)$$

[4 marks]

ii. Compute

$$\frac{d}{dw_i} L(w)$$

and suggest an algorithm to find the vector w that maximises $L(w)$.

[7 marks]

iii. Show that the decision boundary for Logistic Regression can be written as

$$\sum_{i=1}^{D} w_i x_i - f = 0$$

for suitably defined f and discuss the geometric picture of the decision boundary of Logistic Regression compared to Naive Bayes. In particular, are the decision boundaries exactly the same for the two models?

[4 marks]

[Total 33 marks]

END OF PAPER