UNIVERSITY COLLEGE LONDON

EXAMINATION FOR INTERNAL STUDENTS

MODULE CODE : **COMPGI08**

ASSESSMENT PATTERN : **COMPGI08C**

MODULE NAME : **Graphical Models**

DATE : **21-May-13**

TIME : **14:30**

TIME ALLOWED : **2 Hours 30 Minutes**

2012/13-COMPGI08C-001-EXAM-31

Graphical Models, GI08/M056, 2013

Answer ALL FOUR questions.

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

Calculators are NOT permitted

1. a. i. Explain the concepts of a singly-connected and a multiply-connected graph, giving an example graph in each case.

[2 marks]

ii. Explain the concept of a connected graph, giving also an example of a connected graph and a non-connected graph. Explain also what is meant by a connected component.

[2 marks]

iii. Describe how to find the connected components of a graph.

[4 marks]

iv. For a singly-connected graph with N nodes and E edges, it is suggested that $N = E + 1$ is always true. If you believe this suggestion is correct, explain why this must be the case; otherwise give a counter-example.

[3 marks]

v. Explain how to verify if a graph is singly-connected. Your explanation should be of the form of an algorithm that could be easily implemented by a computer programmer.

[3 marks]

b. i. Describe the concept of a clique and a maximal-clique.

[2 marks]

ii. For a graph with N nodes, explain whether it will always be computationally easy to find all the maximal cliques of the graph. If this is computationally easy, give an algorithm; otherwise give an example of why you believe this may not be computationally easy.

[5 marks]

c. If two people i and j know each other, we set the adjacency matrix element $A_{ij} = A_{ji} = 1$; otherwise if they don't know each other we set $A_{ij} = 0$. We do this for every adult on the planet to create a (very large) adjacency matrix A. It is commonly believed that any two people on the planet are connected by at most '6 degrees of separation' – that is that there is a path of at length 6 between any two people. Explain how you could check this efficiently, giving an explicit algorithm, and describe the approximate computational cost of performing this check.

[4 marks]

[Total 25 marks]

2. a. i. For a distribution $p(x_1,\dots,x_N)$, explain what is meant by two variables x_i and x_j, $i \neq j$ being dependent.

[2 marks]

ii. Describe what is meant by a belief network and why its graphical representation must be a Directed Acyclic Graph (DAG).

[2 marks]

iii. Explain why any distribution can be written as a belief network.

[2 marks]

iv. Explain why the lack of an edge in a DAG implies some independence statement in all distributions consistent with the DAG factorisation. Explain also why the presence of an edge in a DAG does not imply a dependence statement in all distributions consistent with the DAG factorisation.

[3 marks]

b. i. Explain what is meant by a Markov network.

[2 marks]

ii. Show that any distribution can be written as a Markov network.

[3 marks]

iii. Consider a belief network whose DAG has adjacency matrix A. Now form an undirected graph with adjacency matrix B containing elements $B_{ij} = 1$ if either $A_{ij} = 1$ or $A_{ji} = 1$; $B_{ij} = 0$ otherwise. Can every belief network with DAG A be represented by a Markov network with adjacency matrix B? If so, explain why. If not, give a counter-example.

[3 marks]

c. i. Explain what is meant by the term that two graphs are 'Markov equivalent'.

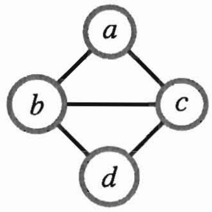
[2 marks]

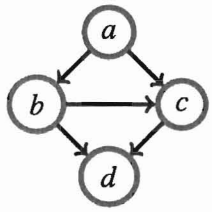
ii. Describe an algorithm to determine if two graphs are Markov equivalent.

[2 marks]

iii. Explain if the following two graphs are Markov equivalent.

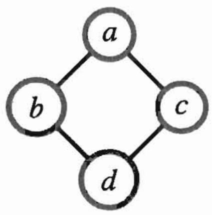
[2 marks]

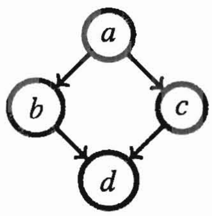




iv. Explain if the following two graphs are Markov equivalent.

[2 marks]





[Total 25 marks]

3. We have a dataset which contains N datapoints. Each datapoint is a D-dimensional vector with components $x_1, \ldots, x_D$. Each variable x_i is discrete, $x_i \in \{1, 2, \ldots, X\}$.

a. i. Explain how the PC algorithm works to learn the structure of a belief network based on data.

[3 marks]

ii. Explain how 'network scoring' works to learn the structure of a belief network based on data. In particular, describe how this can be achieved using a Bayesian approach based on placing Dirichlet priors over the conditional probability tables.

[4 marks]

b. We wish to fit a belief network with D variables to a dataset under the constraint that the number of parents for each variable x_i is less than or equal to P.

i. Explain why finding the fully optimal (maximum likelihood) belief network is computationally difficult, giving both upper and lower bounds on the number of possible belief networks.

[5 marks]

ii. For the case that the number of parents is less than or equal to 1, $P = 1$, derive the Chow-Liu algorithm for learning the maximum likelihood structure from the dataset, and discuss its computational complexity.

[8 marks]

[Total 20 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

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

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

Derive both the E and the M steps of the Expectation Maximization algorithm for maximising the likelihood $p(\mathcal{V}|A, B)$ with respect to A and B.

[6 marks]

c. Consider a distribution on continuous real-valued observations v_t, and discrete hidden variables h_t:

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

where

$$p(v_t|v_{t-1}, h_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(v_t - f(v_{t-1}, h_t))^2\right\}$$

Here $f(v_{t-1}, h_t)$ is fixed function of the previous continuous observation and the current discrete hidden variable state.

i. Derive a recursive algorithm for computing the filtered distribution $p(h_t|v_{1:t})$. [3 marks]

ii. The 'return' of a stock is represented by v_t. Traders are interested to know if the 'return' is increasing by an amount δ, staying the same, or going down by an amount δ. Trader Alice suggests that one can use the model above for suitably defined $f(v_{t-1}, h_t)$. Explain what might be a suitable function $f(v_{t-1}, h_t)$ [3 marks]

iii. Another trader Bob suggests there is a flaw with this idea, namely that it is preferable to consider the returns v_t as noisy versions of the 'clean' underlying return $c_t \in \{-1, -1+\delta, \ldots, -\delta, 0, \delta, \ldots, 1-\delta, 1\}$, with

$$v_t = c_t + \varepsilon_t$$

for normally distributed 'noise', $p(\varepsilon_t) = \mathcal{N}(\varepsilon_t|0, \sigma^2)$. He suggests then to define a model

$$p(v_{1:T}, c_{1:T}, h_{1:T}) = p(v_1|c_1)p(c_1|h_1)p(h_1)\prod_{t=2}^{T} p(v_t|c_t)p(c_t|c_{t-1}, h_t)p(h_t|h_{t-1})$$

Explain how to set up the transition $p(c_t|c_{t-1}, h_t)$ and explain in detail how to compute $p(h_t|v_{1:t})$.

[6 marks]

[Total 30 marks]

END OF PAPER