UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2005

MEng Honours Degree in Electrical Engineering Part IV
MSc in Computing for Industry
MEng Honours Degree in Information Systems Engineering Part IV
MSci Honours Degree in Mathematics and Computer Science Part IV
MEng Honours Degrees in Computing Part IV
MSc in Advanced Computing
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the Associateship of the City and Guilds of London Institute
This paper is also taken for the relevant examinations for the Associateship of the Royal College of Science

PAPER C493=I4.48=E4.41

INTELLIGENT DATA AND PROBABILISTIC INFERENCE

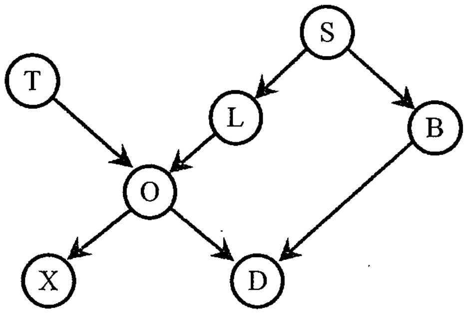
Thursday 28 April 2005, 10:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators required

1 Probability Propagation

The following network is used for reasoning about patients with suspected lung disease



B	Bronchitis
D	Dyspnea
L	Lung Cancer
O	Reduced Lung Capacity
S	Smoker
T	Tuberculosis
X	Positive XRay

All the nodes are binary, and the prior and conditional probabilities are as follows:

$$P(O|T\&L) = \begin{pmatrix} P(O1|T1\&L1) & P(O1|T1\&L2) & P(O1|T2\&L1) & P(O1|T2\&L2) \\ P(O2|T1\&L1) & P(O2|T1\&L2) & P(O2|T2\&L1) & P(O2|T2\&L2) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P(D|O\&B) = \begin{pmatrix} P(D1|O1\&B1) & P(D1|O1\&B2) & P(D1|O2\&B1) & P(D1|O2\&B2) \\ P(D2|O1\&B1) & P(D2|O1\&B2) & P(D2|O2\&B1) & P(D2|O2\&B2) \end{pmatrix} = \begin{pmatrix} 0.9 & 0.8 & 0.7 & 0.1 \\ 0.1 & 0.2 & 0.3 & 0.9 \end{pmatrix}$$

$$P(L|S) = \begin{pmatrix} P(L1|S1) & P(L1|S2) \\ P(L2|S1) & P(L2|S2) \end{pmatrix} = \begin{pmatrix} 0.2 & 0.1 \\ 0.8 & 0.9 \end{pmatrix}$$

$P(T) = (0.1, 0.9)$

$$P(B|S) = \begin{pmatrix} P(B1|S1) & P(B1|S2) \\ P(B2|S1) & P(B2|S2) \end{pmatrix} = \begin{pmatrix} 0.1 & 0.1 \\ 0.9 & 0.9 \end{pmatrix}$$

$P(S) = (0.3, 0.7)$

$$P(X|O) = \begin{pmatrix} P(X1|O1) & P(X1|O2) \\ P(X2|O1) & P(X2|O2) \end{pmatrix} = \begin{pmatrix} 0.4 & 0.1 \\ 0.6 & 0.9 \end{pmatrix}$$

Given that the evidence propagated between nodes is defined by the following equations:

For one parent only

$$\lambda_c(a_k) = \sum_{j=1}^{m} P(c_j \mid a_k)\, \lambda(c_j)$$

For two parents

$$\lambda_c(a_k) = \sum_{i=1}^{n} \pi_C(b_i) \sum_{j=1}^{m} P(c_j \mid a_k \& b_i)\, \lambda(c_j)$$

$$\pi(c_i) = \sum_{j=1}^{n} \sum_{k=1}^{m} P(c_i \mid a_j \& b_k)\, \pi_C(a_j)\, \pi_C(b_k)$$

a Calculate the π evidence for the nodes L and O before any measurements are made.

b A new patient arrives and has an X-Ray taken. Fortunately for him, it is negative (state X2). Given just this evidence, calculate the probability of his suffering from Lung cancer (ie the probability distribution over L).

c He is now examined and found to be suffering from Dyspnea (D is in state D1). Explain why it is no longer possible to compute a probability of his suffering from lung cancer.

d He now admits to being a smoker (S is in state s1). Calculate the probability of his suffering from Lung Cancer.

e Explain briefly the advantages and disadvantages of using a join tree for calculating probabilities rather than simple λ and π messages.

The five parts carry equal marks

2 The maximum Weighted Spanning Tree

A data warehouse contains the following vast data set connecting three variables A B and C:

A	B	C
a1	b1	c1
a1	b1	c2
a1	b1	c2
a1	b2	c1
a2	b2	c2
a2	b2	c1
a2	b2	c2
a2	b2	c1

a Construct co-occurrence matrices for the three possible pairings AB, BC and AC:

	a1	a2
b1		
b2		

etc

b From the co-occurrences construct the joint probability table for each pair, and the marginalisations, using the following format:

	a1	a2	P(B)
b1			
b2			
P(A)			

etc.

c Calculate the L1 metric for each possible pair of nodes

$$\mathrm{Dep}(A,B) = \sum_{A \times B} |P(ai\&bj) - P(ai)P(bj)|$$

d Given that A is the root node construct the tree.

e Calculate the prior probability distribution P(A).

f Calculate the two conditional probability matrices for the tree found in part 2d.

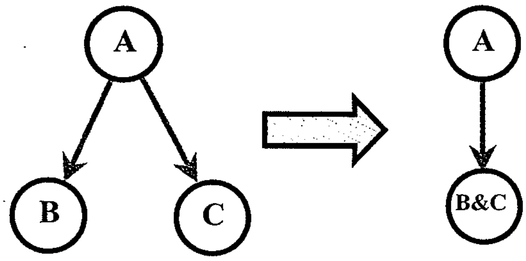
g The tree of part (d) is being used in the case where it is not possible to measure node B. The state of C is found to be c2. Estimate the probabilities of nodes A and B.

The seven parts carry, respectively, 15%,20%,15%,10%,10%,15% and 15% of the marks

3 Join Trees

Joining two variables is one technique that can be used to improve the accuracy of a network in cases where there are dependencies that cannot be accommodated in a singly connected graph. Consider the following simple network and data set:

a0	b0	c1
a0	b0	c1
a0	b1	c0
a0	b1	c0
a1	b0	c0
a1	b0	c0
a1	b1	c0
a1	b1	c1

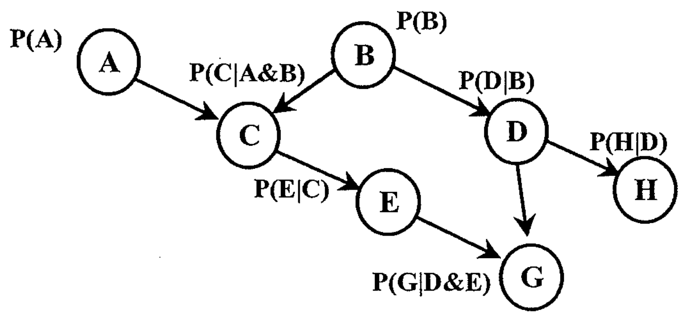


a Find the conditional probability matrices for:
  i) the original network (P(B|A) and P(C|A))
  ii) the joined network (P(B&C|A)).

b The network is being initialised and the prior probability of A is known to be P(A) = {0.3,0.7}. This prior probability is treated as π evidence and propagated. What are the posterior distributions over B and C in:
  i) the original network
  ii) the joined network

c The node B is instantiated to state b1, and node C is instantiated to c0. Calculate the λ evidence at node A using
  i) the original network
  ii) the joined network

d Given the network:



find a join tree using the following steps from the algorithm by Lauritzen and Spiegelhalter:
  i) Find the moral graph
  ii) Triangulate the moral graph
  iii) Identify the cliques of the resulting triangulated moral graph
  iv) Define a join tree in which the ordering of the cliques satisfies the running intersection property.
  v) Draw up a table showing the potential functions and the R and S sets for each node

The four parts carry, respectively, 20%, 20%, 20% and 40% of the marks

4 Co-variance and Principal Component Analysis

The following artificial problem has three variables and just two data points set out in the table below

x	y	z
1	0	3
5	2	1

a Find the mean-centred data matrix $U = [\mathbf{x\text{-}x_m}, \mathbf{y\text{-}y_m}, \mathbf{z\text{-}z_m}]$, hence, find the co-variance matrix using U^TU

b For real problems, where sample sizes are much smaller than the number of variables, it is common to use the Karhunen-Lowe transformation to find the principal components. Using the answer to part a, find the reduced matrix $A = UU^T$ and calculate its principal eigenvector. First calculate the eigenvalues using $\det(A-\lambda I)=0$, then solve $(A-\lambda I)\, e = 0$ to find the eigenvectors.

c Find the principal eigenvector of the covariance matrix of part a by multiplying the eigenvector found in part b by the data matrix U. Check that this is indeed an eigenvector of the covariance matrix, and find its corresponding eigenvalue.

d Explain what is meant by an eigenface in face recognition. How many eigenfaces could be calculated from a face data base in which the images are 128 by 128 resolution and there are three images of twenty subjects in the data base.

e The Mahalanobis distance between two two-dimensional points (x_1,y_1) (x_2,y_2) is defined by the formula:

$$\sqrt{((x_2-x_1,y_2-y_1)\, \Sigma^{-1}\, (x_2-x_1,y_2-y_1)^T)}$$

where Σ is the covariance matrix.

Explain briefly why the Mahalanobis distance might be used in preference to the Euclidian distance in classification problems.

The five parts carry equal marks