IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2008

BEng Honours Degree in Computing Part III
MEng Honours Degree in Electrical Engineering Part IV
MEng Honours Degree in Information Systems Engineering Part IV
MEng Honours Degrees in Computing Part IV
MSc in Advanced Computing
MSc in Computing Science (Specialist)
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

PAPER C493=I4.48=E4.41

INTELLIGENT DATA AND PROBABILISTIC INFERENCE

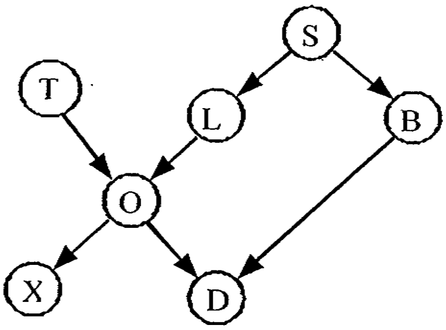
Wednesday 14 May 2008, 10:00
Duration: 120 minutes

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:



S	Smoker
T	Tuberculosis
L	Lung Cancer
B	Bronchitis
O	Reduced Lung Capacity
X	Positive X Ray
D	Dypsnea

All the nodes are binary, with the the convention that the lower index state indicates absence of the condition. For example, B=(b_1, b_2) where b_1 means the patient is clear of bronchitis and b_2 means he is suffering from bronchitis. 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:

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

for two parents:

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

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

a Calculate the π evidence for the node X before any measurements are made.

b A new patient arrives and is diagnosed to be suffering from tuberculosis. Given just this evidence, calculate the probability of his suffering from Lung cancer (ie calculate the probability distribution over L).

c An X ray is taken but unfortunately it is not very conclusive. The radiologist estimates that it is more likely to be positive. Thus node X is instantiated with virtual evidence (0.3,0.7). Re-calculate the probability that the patient has lung cancer.

d The patient is now examined further and found to be suffering from Dyspnea (D is instantiated to state d_1). Explain why it is no longer possible to compute a probability of his suffering from lung cancer. Which other nodes need to be instantiated before the probability distribution over L can be found?

e Explain briefly how cutset conditioning could be used to calculate the probability distribution over L, given just the instantiations outlined in parts a b and c.

The five parts carry equal marks

2. Dependency Measures and Causal Directions

a The L1 dependency metric is defined as follows:

$$Dep(A,B) = \sum_{i \times j} |P(a_i \& b_j) - P(a_i)P(b_j)|$$

where the sum is taken over all the states of variables A and B.

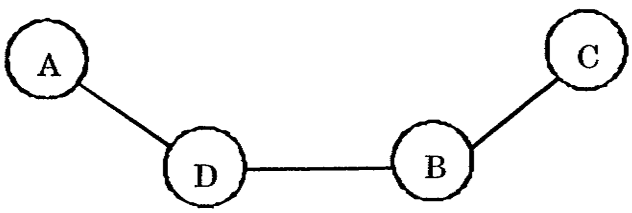
Explain how and why it is used as a dependency measure between two discrete variables in a data set.

b The following data set has four variables each with 3 states.

A	B	C	D
a_1	b_1	c_2	d_3
a_1	b_2	c_3	d_1
a_1	b_3	c_1	d_3
a_2	b_2	c_3	d_2
a_2	b_1	c_2	d_2
a_3	b_3	c_1	d_1
a_3	b_1	c_3	d_3
a_3	b_2	c_3	d_1

Find the dependencies of the variable pairs AB and CD using the L1 dependency metric

c Given that the most dependent arcs are AD, BC and BD the spanning tree will be:



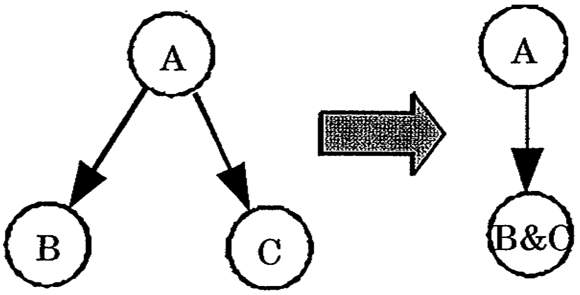
Use the L1 metric to determine the marginal independence and hence the possible directions of the arrows at each triplet in the spanning tree. Hence deduce what you can about the arrow directions on the tree.

d List the possible sources of error in determining arc direction using marginal independence.

The four parts carry, respectively, 20%, 20%, 40% and 20% 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:



a_1	b_1	c_1
a_1	b_1	c_1
a_1	b_2	c_2
a_1	b_2	c_1
a_2	b_1	c_1
a_2	b_1	c_2
a_2	b_2	c_2
a_2	b_2	c_2

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 node C is instantiated to state c1, but node B is not instantiated. Calculate the λ evidence at node A using
  i) the original network
  ii) the joined network

c An alternative method to deal with unaccounted dependencies would be to delete one of the dependent nodes, B or C in the above example. Describe the advantages and disadvantages of node deletion compare to joining nodes.

d Given the network shown in question 1 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. Classification

A classifier is to be designed to separate two classes for which the following data points are known:

Class 1: (2,1), (1,2), (2,4), (2,3)
Class 2: (2,2), (3,3), (4,3), (5,3)

a Find the two individual class means and, using all the data, the grand mean and the pooled scatter matrix S_p. (The scatter matrix is the un-normalised co-variance matrix.)

b Calculate the "between class" covariance matrix S_b from the two class means found in part a above.

c Find the direction of the most discriminant LDA feature by finding the principal eigenvector of the matrix $L = S_p^{-1} S_b$. Hint: First calculate the eigenvalues using $\det(L-\lambda I)=0$, then solve $(L-\lambda I)\,\mathbf{e} = 0$ to find the eigenvectors.

d Using the pooled scatter matrix find the directions of the principal modes of variation of the whole data set, that is, the directions of the PCA axis system.

e Make an accurate sketch of the data plotted on the *xy*-Cartesian plane. On your sketch show:
(i) The most discriminant LDA feature direction
(ii) The principal PCA direction.

The five parts carry equal marks