UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2004

MEng Honours Degree in Electrical Engineering Part IV
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

Tuesday 27 April 2004, 14:30
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators required

1.

a. Describe the structure of the Decision Tree algorithm and explain why the algorithm is guaranteed to terminate.

b. The table shown below provides a set of records with their true classes (Buys Car):

Name	**Age**	**Sex**	**Credit Rating**	**Post Code**	**Job**	**Buys Car**
Mike	Young	Male	Excellent	SW7	Student	N
Sarah	Middle	Female	Poor	SW7	Other	N
John	Old	Male	Excellent	SW5	Student	N
Sam	Young	Female	Poor	NW1	Student	Y
Emma	Middle	Female	Excellent	NW2	Other	N
Nikos	Middle	Male	Poor	SW5	Other	Y
Kim	Young	Female	Excellent	SW8	Student	Y
George	Young	Male	Poor	W14	Other	Y
Steve	Old	Male	Poor	NW1	Other	N
Patrick	Middle	Male	Poor	SW8	Student	Y

i. Which attribute will be selected by the decision tree algorithm as the root node? And why?

ii. Which attribute/attributes can be removed? And why?

iii. If the number of data items in the table are increased, does the accuracy of the tree model learned from the decision tree algorithm increase as well?

The two parts carry, respectively, 30% and 70% (20%, 30%, 20%) of the marks.

2.

a. Describe briefly each of the following data-mining functionalities: naïve Bayes classification and hierarchical clustering.

b. What is the basic feature of the *apriori* association rule learning algorithm that makes it qualitatively different from an exhaustive rule search?

c. Explain the main advantage of KNN (K nearest neighbors) classification algorithm

d. Using the Table in Question 1, classify a new instance

Lee	Young	Female	Excellent	SW7	Student

using Naïve Bayes, make sure you identify how you deal with the value= "Lee" for attribute Name.

The four parts carry equal marks.

3. Dependency Measures and Causal Directions

a. The mutual entropy measure, or Kullback-Liebler divergence, is defined by the following formula:

$$H(A,B) = \sum_{i \times j} P(a_i\&b_j) \log_2 (P(a_i\&b_j) / P(a_i) P(b_j))$$

Where the sum is taken of all the states of variables A and B. It is frequently used to determine the "distance" or "divergence" between two probability distributions.

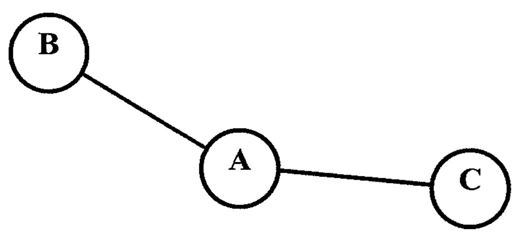
Explain how it can be used as a dependency measure between two discrete variables in a data set.

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

A	B	C
a_1	b_1	c_2
a_2	b_2	c_3
a_3	b_3	c_1
a_2	b_2	c_3
a_1	b_1	c_2
a_3	b_3	c_1
a_1	b_1	c_2
a_2	b_2	c_3

Find the mutual entropy between each pair of variables.

c. A Bayesian network incorporating variables A, B and C has the following structure:

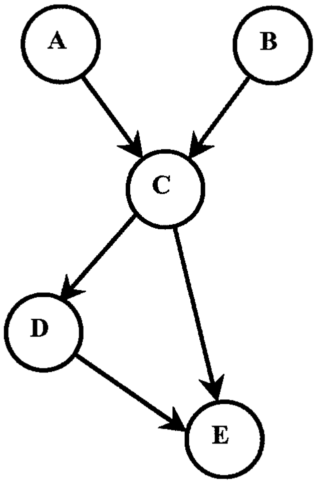


Given the data of part b determine the possible directions of the arrows.

The three parts carry, respectively, 30%, 35% and 35% of the marks.

4. Probability Propagation

A Bayesian network has the following structure, prior probabilities and conditional probability matrices:



P(A)	a1	a2
	0.5	0.5

P(B)	b1	b2
	0.25	0.75

P(C|A&B)		a1 b1	a1 b2	a2 b1	a2 b2
	c1	0.5	0	1	0.25
	c2	0.5	1	0	0.75

P(D|C)		c1	c2
	d1	0.25	0.75
	d2	0.75	0.25

P(E|C&D)		c1 d1	c1 d2	c2 d1	c2 d2
	e1	0	1	0.25	0.5
	e2	1	0	0.75	0.5

a Find the joint probability of the network for the data point {a1,b2,c2,d2,e1}

b The network is to be used for inference. Describe what happens during the initialisation stage, and calculate the π evidence for variable C after instantiation.

c The following three instantiations are made: B=b1, C=c2, D=d1. Calculate the *posterior* probabilities of variables A and E.

d Identify one possible set of instantiated nodes in which it is not possible to propagate probabilities using Pearl's operating equations. Briefly describe the messages that would be sent if your instantiation occurred.

e Briefly describe the method of cutset conditioning for propagating probabilities in networks with loops, indicating which nodes in the above example could be used to form cut sets.

f An alternative to using cutsets is the introduction of a hidden node. Re-draw the above network to include a hidden node which will make it singly connected.

Pearl's operating equations for probability propagation are provided on a supplementary sheet.

The six parts carry, respectively, 15%,20%, 20%15%,15%,15% of the marks.

Supplementary Sheet: Pearl's Operating Equations for Probability Propagation

Operating Equation 1: λ message

The lamda message from C to A is given by

For one parent only

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

For two parents

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

Operating Equation 2: The π Message

If C is a child of A, the π message from A to C is given by:

$$\pi_c(a_j) = \begin{array}{ll} 1 & \text{if A is instantiated for aj} \\ 0 & \text{if A is instantiated but not for aj} \\ P'(a_j)/\lambda_c(a_j) & \text{if A is not instantiated} \end{array}$$

Operating Equation 3: The λ evidence

If C is a node with n children D1, D2, . . Dn, then the λ evidence for C is:

$$\lambda(cj) = \begin{cases} 1 & \text{if C is instantiated for cj} \\ 0 & \text{if C is instantiated but not for cj} \\ \prod_i \lambda_{Di}(c_j) & \text{if C is not instantiated} \end{cases}$$

Operating Equation 4: The π evidence

If C is a child of two parents A and B the π evidence for C is given by:

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

Operating Equation 5: the posterior probability

If C is a variable the (posterior) probability of C based on the evidence received is written as:

$P'(ci) = \alpha\, \lambda(ci)\, \pi(ci)$

where α is chosen to make $\Sigma\, P'(ci) = 1$