UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2003

BEng Honours Degree in Computing Part III
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
PhD
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

Wednesday 14 May 2003, 14:30
Duration: 120 minutes

Answer THREE questions

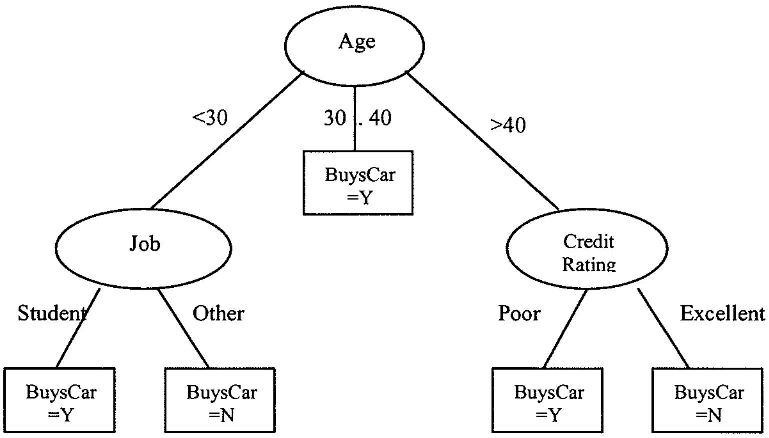
Paper contains 4 questions
Calculators required

1.

a. Compare the advantages and disadvantages of using decision trees and neural networks as classification methods.

b. Describe how decision tree shown below can be used to decide whether any of the following two individuals will buy a used car from Used_Cars_Ltd.

Mike: 25 year old male student with excellent credit rating.
Sarah: 45 year old city worker with poor credit rating.



c. Using pseudo code, describe the major steps of a typical decision tree induction algorithm. Make sure your answer describes how the order of the decision nodes is chosen using the concept of entropy.

d. The table shown below provides a set of records with their true classes (Buys Car), Using the decision tree given in part b,

Name	**Age**	**Sex**	**Credit Rating**	**Post Code**	**Job**	**Buys Car**
Mike	25	Male	Excellent	SW7	Student	N
Sarah	45	Female	Poor	SW7	Other	N
John	60	Male	Excellent	SW5	Student	N
Sam	23	Female	Poor	NW1	Student	Y
Emma	29	Female	Excellent	NW2	Other	N
Nikos	35	Male	Poor	SW5	Other	Y
Kim	18	Female	Excellent	SW8	Student	Y
George	21	Male	Poor	W14	Other	Y
Steve	38	Male	Poor	NW1	Other	N
Patrick	35	Male	Poor	SW8	Student	Y

i. Calculate the overall accuracy of the decision tree.
ii. Derive a confusion matrix for the data set and the true positive and false negative rates.

e.

i. Describe what is meant by model over-fitting in machine learning.
ii. Describe two approaches that can be used to avoid over-fitting

The five parts carry equal marks.

2.

a. Describe briefly each of the following data mining functionalities: association rule discovery, classification and clustering.

b.

i. Explain the operation of the k-means clustering algorithm using pseudo code.

ii. Given the following eight points, and assuming initial cluster centroids given by A, B, C, and that a Euclidean distance function is used for measuring distance between points, use k-means to show only the three clusters and calculate their new centroids after the second round of execution.

ID	X	Y
A	2	10
B	2	5
C	8	4
D	5	8
E	7	5
F	6	4
G	1	2
H	4	9

c.

i. Explain the meaning of support and confidence in the context of association rule discovery algorithms and explain how the a priori heuristic can be used to improve the efficiency of such algorithms.

ii. Given the transactions described below, find all rules between single items that have support > 60%. For each rule report both support and confidence.

1: (Beer)
2: (Cola, Beer)
3: (Cola, Beer)
4: (Nuts, Beer)
5: (Nuts, Cola, Beer)
6: (Nuts, Cola, Beer)
7: (Crisps, Nuts, Cola)
8: (Crisps, Nuts, Cola, Beer)
9: (Crisps, Nuts, Cola, Beer)
10:(Crisps, Nuts, Cola, Beer)

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

3.

a. The L1 metric is defined by the following equation:

$Dep(A,B) = |P(A\&B) - P(A)P(B)|$

Explain why it is a measure of dependency between two variables A and B.

b. Two variables A and B each have three states. In an experiment they are measured and the results are set out in the following table.

a0	b2
a0	b0
a1	b0
a0	b2
a2	b0
a2	b1
a0	b2
a1	b2
a2	b0
a2	b1
a0	b2
a2	b1

(i) Calculate the joint probability distribution P(A&B)
(ii) Marginalise to find the probabilities P(A) and P(B)
(iii) Find the Dependency of A and B using the L1 metric.

c. Given that two nodes of a network are connected by a link matrix P(C|D), and each has three states, give an example of the link matrix that expressed no dependency between C and D, and a link matrix that expresses maximal dependency.

d. In the MDL metric the measure for model accuracy is defined as $\log_2 P(D|B)$ where:

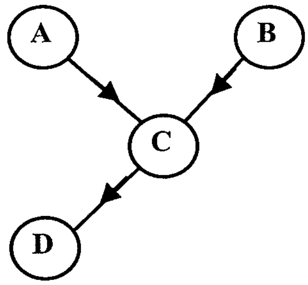
$$P(D|B) = \prod_{data} P(B)$$

and B is the joint probability of the network. Calculate the model accuracies for the following network:

e. A similar experiment to part b is carried out, but this time sixteen points are measured. The probability distribution turns out to be identical to the answer you obtained in part b. How would you expect the model accuracies of part d to change?

The five parts carry equal marks.

4 A four node network has the following structure.



A set of eight data measurements is given in the table. A,B and C have two states and D has three states.

a0	b0	c0	d0	a0	b0	c1	d1
a1	b0	c0	d1	a0	b1	c1	d0
a1	b0	c0	d2	a0	b1	c1	d0
a1	b1	c0	d0	a1	b1	c1	d1

a. Estimate the conditional probability distributions P(C|A&B), P(D|C) and the prior probabilities P(A) and P(B) from the data

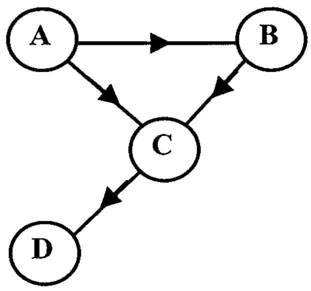
b. Explain what happens in the initialisation step of Pearl's probability propagation algorithm.

c. After initialisation, node B is instantiated to state b1. Explain what probabilities are propagated following then instantiation. Calculate the π evidence at D.

d. Node D is now instantiated to state d2. Calculate the λ evidence that reaches node C and A.

e. The value of node B is found to be incorrect. It is instantiated again to state b0, while D remains instantiated at d2 Calculate the probability distribution at node A.

f. After discovering more data the network structure is changed as follows.



For each node of the network explain briefly what will happen if it is the first node to be instantiated.

The operating equations for probability propagation are given on the next page.

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

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:

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

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 $\Sigma\, P'(ci) = 1$