IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2011

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
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
This paper is also taken for the relevant examinations for the Associateship of the Royal College of Science

PAPER C493

INTELLIGENT DATA AND PROBABILISTIC INFERENCE

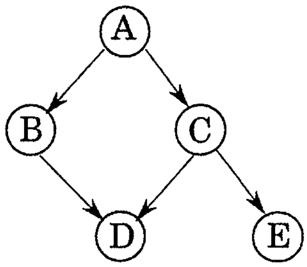
Wednesday 18 May 2011, 14:30
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators required

1 Probability Propagation.

The following Bayesian network has been proposed to model the data from metastatic cancer. The variables are all binary, with, for example, variable A having states a_1 and a_2.



The conditional probabilities are given in the form

$$P(D|B\&C) = \begin{bmatrix} P(d_1|b_1\&c_1) & P(d_1|b_1\&c_2) & P(d_1|b_2\&c_1) & P(d_1|b_2\&c_2) \\ P(d_2|b_1\&c_1) & P(d_2|b_1\&c_2) & P(d_2|b_2\&c_1) & P(d_2|b_2\&c_2) \end{bmatrix}$$

and are

$$P(A) = [0.1, 0.9]$$

$$P(D|B\&C) = \begin{bmatrix} 0.6 & 0.4 & 0.3 & 0.8 \\ 0.4 & 0.6 & 0.7 & 0.2 \end{bmatrix}$$

$$P(B|A) = \begin{bmatrix} 0.1 & 0.3 \\ 0.9 & 0.7 \end{bmatrix}$$

$$P(E|C) = \begin{bmatrix} 0.2 & 0.3 \\ 0.8 & 0.7 \end{bmatrix}$$

$$P(C|A) = \begin{bmatrix} 0.4 & 0.2 \\ 0.6 & 0.8 \end{bmatrix}$$

Pearl's operating equations are given at the end of the paper.

a Find the π evidence of nodes D and E after initialisation, but before any nodes have been instantiated.

b Node E is instantiated to state e_2. Calculate the posterior probability of node A.

c With node E instantiated as above, node D is now instantiated to state d_1. Unfortunately this causes a loop and the probability propagation algorithm will not terminate. In order to calculate the posterior distribution of variable C it is decided to use the method of cutset conditioning, using variable A as the cutset. What is the value of C found by this method?

d The dependency between B and D is small, so another approach to propagate the probability for part c is to remove the link between B and D. Does this give a different result to cutset conditioning in this case? Explain your reasoning.

The four parts carry equal marks.

2 Join Trees

The network given in Question 1 is to be transformed into a join tree to achieve exact computation of the probabilities.

a Moralise the original graph and thus find a set of cliques of the variables. List the variables that belong to each clique.

b Explain what is meant by the running intersection property, and find a tree of cliques that has this property.

c Using the conditional probability tables of question 1, find the initial potential table allocated to each clique.

d Assuming that D is instantiated to d_1, find the λ message that will be propagated from the clique in which D is in the R set of variables.

e What are the advantages and disadvantages of the join tree algorithm compared with using Pearl's message passing algorithm.

The five parts carry equal marks.

3 Dependency Measures and Causal Directions.

Dependency between pairs of variables is frequently measured using the Kullback-Liebler divergence. It is defined by the following formula:

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

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

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

b Three binary variables, A, B, C have the following data points:

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

Calculate the dependency between each pair using the Kullbach Liebler divergence, and hence find the maximum weighted spanning tree of the three variables.

c Using the method of marginal independence, investigate to see whether causal directions can be found for the spanning tree found in part b.

d What are the limitations and potential sources of error in using marginal independence to determine the causal directions in Bayesian networks.

e Explain how the method of marginal independence can be computed from the joint probability matrix $P(A\&B\&C)$ of the three variables, rather than from the data set.

The five parts carry equal marks.

4 Principal Component Analysis.

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

X	Y	Z
1	3	2
-3	1	0

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

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 principal components. Use the matrix U of part a to find the matrix $A = UU^T$ and calculate its principal eigenvector. (First calculate the eigenvalues using $det(A - \lambda I) = 0$, then solve $(A - \lambda I)\mathrm{e} = 0$ to find the eigenvectors.)

c Using the eigenvector found in part b and the mean centered data matrix of part a, find the principal eigenvector of the covariance matrix of part a. Check that this is indeed an eigenvector of the covariance matrix, and find its corresponding eigenvalue.

d PCA is often used in face recognition. In a particular application the face data base has 60 subjects. Each subject has three different 2D images. Each image is grayscale with resolution 256 by 256 pixels.

i) What is the dimension of the input data space?

ii) How many eigenvectors with real non-zero corresponding eigenvalues can be obtained from the data?

iii) Why are the eigenvectors called eigenfaces?

e When PCA is used for face recognition is is normal to discard some of the eigenfaces that are found from the covaraince matrix. Briefly describe a criterion for deciding how many of the eigenfaces should be retained.

The five parts carry equal marks.

Pearl's Operating equations for probability propagation

Equation 1: The λ message

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

and for the case of the single parent there is a simpler matrix form:

$$\lambda_\mathbf{c}(\mathbf{A}) = \lambda(\mathbf{C})\mathbf{P}(\mathbf{C}|\mathbf{A})$$

Equation 2: The π Message from A to C is given by:

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

Equation 3: The λ evidence of node C with n children $D_1, D_2, ..D_n$:

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

Equation 4: The π evidence of node C with two parents A and B:

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

This can be written in matrix form using as follows:

$$\pi(\mathbf{C}) = \mathbf{P}(\mathbf{C}|\mathbf{A}\&\mathbf{B})\pi_\mathbf{C}(\mathbf{A}\&\mathbf{B})$$

where

$$\pi_C(a_j\&b_k) = \pi_C(a_j)\pi_C(b_k)$$

Equation 5: The posterior probability of variable C:

$$P'(c_i) = \alpha\lambda(c_i)\pi(c_i)$$

where α is chosen to make $\sum_i P'(c_i) = 1$