IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2012

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

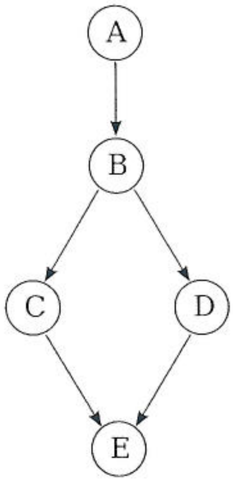
Monday 14 May 2012, 10:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators required

1 Probability Propagation.

In the following Bayesian network the variables are all binary, with, for example, variable A having states a_0 and a_1.



The conditional probability tables are written in the form:

$$P(E|C\&D) = \begin{bmatrix} P(e_0|c_0\&d_0) & P(e_0|c_0\&d_1) & P(e_0|c_1\&d_0) & P(e_0|c_1\&d_1) \\ P(e_1|c_0\&d_0) & P(e_1|c_0\&d_1) & P(e_1|c_1\&d_0) & P(e_1|c_1\&d_1) \end{bmatrix}$$

and are

$$P(A) = [0.4, 0.6]$$

$$P(E|C\&D) = \begin{bmatrix} 0.5 & 0.4 & 0.3 & 0.8 \\ 0.5 & 0.6 & 0.7 & 0.2 \end{bmatrix}$$

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

$$P(D|B) = \begin{bmatrix} 0.1 & 0.8 \\ 0.9 & 0.2 \end{bmatrix}$$

$$P(C|B) = \begin{bmatrix} 0.1 & 0.2 \\ 0.9 & 0.8 \end{bmatrix}$$

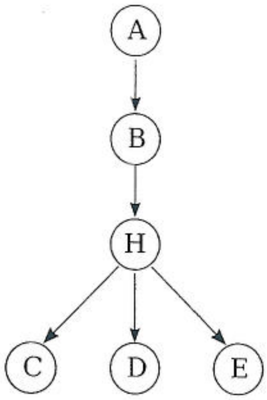
Pearl's operating equations are given at the end of the paper.

a Find the π evidence for node E after initialisation, but before any nodes have been instantiated.

b If node E is instantiated to e_1 and node C is instantiated to c_0 calculate the posterior probability over node D.

c With node E instantiated to state e_1 and node C uninstantiated the probability propagation algorithm will not terminate. One method to get an approximate solution is to delete one arc from the network, as would be done in the case of the maximum weighted spanning tree algorithm. Given that the arc BD is removed from the network, and that D is assumed to have equal priors ($P(D) = (0.5, 0.5)$), calculate the posterior probability of node C.

d An alternative method to get an approximate solution is to make use of hidden nodes. A possible alternative network is given in the figure below. Briefly explain how the new conditional probability matrices $P(H|B)$, $P(C|H)$, $P(D|H)$ and $P(E|H)$ could be found using the original data set.



The four parts carry equal marks.

2 Joining Nodes

Given the Bayesian network of question 1 it is possible to get exact solutions for the probabilities of the unknown variables by joining them into groups. The figure shows a network which can be used as a replacement for the original.

a
 i) Calculate the conditional probability matrix $P(CD|B)$.

 ii) Given that node B is instantiated to state b_0 and node E is instantiated to e_0 calculate the probability distribution over node C.

b Find a join tree for the network of question 1 by using the following steps:

 i) Find the moral graph.

 ii) Find a tree of cliques of the moral graph that obeys the running intersection property.

 iii) Define a potential table for each clique in terms of the original conditional probability tables.

 iv) Define the R and S sets for each clique found

c In the case where the join tree is being initialised before any instantiation occurs, the λ messages propagated will contain no evidence. Explain briefly why this is so.

d To complete the initialisation of the join tree, after propogating the λ messages upwards, π messages are propagated down the tree. Explain briefly how the π messages are computed.

e Once the initialisation of the join tree is completed the probability distribution over E can be calculated by marginalisation. Would you expect to get the same result if you calculated the distribution over E using Pearl's message passing algorithm on the original Bayesian network (following the method described in the first part of question 1)? Explain your answer.

The five parts carry equal marks.

3 Sampling and re-sampling

a Sampling provides a practical approach to estimating probability distributions in many cases where exact computation may be computationally infeasible. Explain what it means for a sample chain to be:

i) Ergodic

ii) Balanced

b Define what is meant by the Markov blanket of a node in a Bayesian network. What is the Markov blanket of node D in the network defined in question 1?

c For the network of question 1 explain how a Monte Carlo Markov chain (MCMC) can be used to estimate the probabilities of the nodes B, C and D in the case where nodes A and E have been instantiated.

d Briefly explain how the Metropolis-Hastings algorithm could be used for selecting samples in the MCMC chain, and what the advantages of using it is.

e Explain what is meant by a bootstrap data set and why they are used.

The five parts carry equal marks.

4 Samll sample size problems

a A small sample size problem is one where the number of data points is smaller than the number of class variables. A very simple problem has three variables and two data points. There are two classes, each defined by a data matrix D where the rows are the data points and the columns are the variables.

$$D_1 = \begin{bmatrix} 2 & 3 & 3 \\ 5 & 1 & 0 \end{bmatrix}$$

$$D_2 = \begin{bmatrix} -2 & -2 & -1 \\ -4 & -6 & -5 \end{bmatrix}$$

Calculate the 3×3 covariance matrix $U^T U$ for each class.

b Explain why it would not be possible to use a parametric Bayesian classifier with a Gaussian 'plug in' kernel to calculate the probability of a given point belonging to each of the above classes.

c Fisher defined a pooled covariance for a multi class problem using the following formula:

$$\Sigma_p = \frac{1}{N-g} \sum_{i=1}^{g} (N_i - 1)\Sigma_i$$

Using this method calculate a pooled covariance for the above two class problem.

d With the aid of a suitable diagram explain how a non-parametric classifier, such as the Parzen window, makes an estimate of the probability that a point $\mathbf{x}$ belongs to a class i.

e For the two classes above calculate a suitable covariance estimate that could be used in a Parzen classifier as follows:

i) For class 1 use the estimate proposed by van Ness.

ii) For class 2 shrink the covariance towards the pooled estimate with a parameter of 1/2.

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 $\sum_i P'(c_i) = 1$