IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2014

MEng Honours Degree in Electronic and Information 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
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

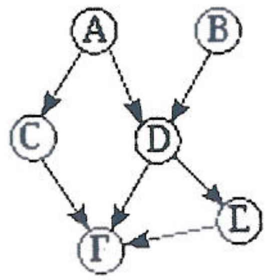
Tuesday 25 March 2014, 10:00
Duration: 120 minutes

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_1 and a_2.



The conditional probabilities are given consistently in the form:

$$P(D|A\&B) = \begin{bmatrix} P(d_1|a_1\&b_1) & P(d_1|a_1\&b_2) & P(d_1|a_2\&b_1) & P(d_1|a_2\&b_2) \\ P(d_2|a_1\&b_1) & P(d_2|a_1\&b_2) & P(d_2|a_2\&b_1) & P(d_2|a_2\&b_2) \end{bmatrix}$$

and are

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

$$P(D|A\&B) = \begin{bmatrix} 0.6 & 0.4 & 0 & 0.1 \\ 0.4 & 0.6 & 1 & 0.9 \end{bmatrix}$$

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

$$P(C|A) = \begin{bmatrix} 0 & 0.3 \\ 1 & 0.7 \end{bmatrix} \qquad P(E|D) = \begin{bmatrix} 0.4 & 1 \\ 0.6 & 0 \end{bmatrix}$$

$$P(F|C\&D\&E) = \begin{bmatrix} 1 & 0.3 & 0.2 & 0.1 & 1 & 0.6 & 0 & 0 \\ 0 & 0.7 & 0.8 & 0.9 & 0 & 0.4 & 1 & 1 \end{bmatrix}$$

The equations for propagating probabilities in Bayesian networks are:

The λ message from child C to parents A and B is given by:

$$\lambda_C(a_i) = \sum_{j=1}^{m} \pi_C(b_j) \sum_{k=1}^{n} P(c_k|a_i\&b_j)\lambda(c_k)$$

In the case where we have a single parent (A) this reduces to:

$$\lambda_C(a_i) = \sum_{k=1}^{n} P(c_k|a_i)\lambda(c_k)$$

and for the case of the single parent we can use the simpler matrix form:

$$\boldsymbol{\lambda_C(A) = \lambda(C)P(C|A)}$$

The matrix form for multiple parents relates to the joint states of the parents.

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

It is necessary to separate the λ evidence for the individual parents with a scalar equation of the form:

$$\lambda_C(a_i) = \Sigma_j \pi_C(b_j)\lambda_C(a_i\&b_j)$$

The π evidence to child node C from two parents A and B is given by:

$$\pi(c_k) = \sum_{i=1}^{l}\sum_{j=1}^{m} P(c_k|a_i\&b_j)\pi_C(a_i)\pi_C(b_j)$$

This can be written in matrix form as follows:

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

where

$$\pi_C(a_i\&b_j) = \pi_C(a_i)\pi_C(b_j)$$

The single parent matrix equation is:

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

a Find the values of the π evidence for node D after initialisation and before any nodes are instantiated.

b Node D is instantiated to state d_1 and all other nodes remain uninstantiated. Calculate the posterior probability of node A when propagation terminates.

c Node E is now instantiated to state e_1. Calculate the π evidence for node F when propagationn finishes.

d i) For what instantiations would the message passing algorithm fail to terminate.

ii) Which single node would form a cutset ensuring propagation always terminates.

e i) Explain what is meant by the Markov blanket of a node.

ii) What is the Markov blanket of node D?

The five parts carry equal marks.

2 Join Tree Algorithm

a For the Bayesian network defined in question1 find the moral graph and its cliques.

b Fill up a table in the following format for each clique found in part a:

W	R	S	Ψ

where W, R and S are sets of varibles used in the join tree algorithm with the usual naming conventions and Ψ is the potential function of the clique.

c Explain what is meant by the running intersection property. Find all join trees that display the running intersection property.

d Calculate the initial potential table for one three node clique of a join tree found in part c.

e Explain why the moralisation step is necessary in the join tree algorithm.

The five parts carry equal marks.

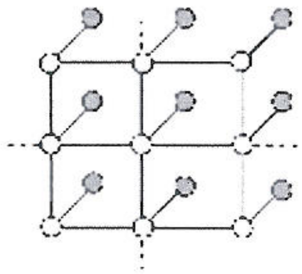
3 Graphical Models and Mixture Models

A Markov random field is defined as a factorisation of a probability distribution with the following equation:

$$P(X,Y) = \frac{1}{Z}\prod_{i,j}\Psi(X_i,X_j)\prod_{i}\Phi(X_i,Y_i) \qquad (1)$$

a Explain the purpose of the functions Ψ and Φ.

b The diagram shows a small part of a pairwise Markov random field used for medical image segmentation. Explain briefly how image segmentation is achieved using this model.



c i) Draw the factor graph corresponding to the Markov random field of part b, labeling typical nodes and factors using the standard labeling conventions.

ii) Explain why it is often not possible to calculate the normalisation constant Z in equation 1.

d A mixture model is one in which a probability distribution is represented as a weighted sum of Gaussian distributions. It is written as:

$$p(x|\theta) = \sum_{j=1}^{M}\alpha_j p_j(x|\theta_j)$$

where there are M probability distributions in the mixture and α_i are the mixing weights that have the property $\sum_{j=1}^{M}\alpha_j = 1$, and the symbol θ_j denotes the vector of unknown parameters that defines each individual distribution.

Explain, with the aid of a suitable diagram, why mixture models may be prone to overfitting.

e Write down an expression for the log likelihood of the mixture model defined in part d for a data set with N variables. Explain why the log liklihood is preferred to the likelihood.

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	0	3
5	2	1

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 When PCA is used for face recognition it 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.

e Explain briefly how a 3D active appearance model of the human face can be implemented, indicating what its input variables are and how PCA is used in its construction.

The five parts carry equal marks.