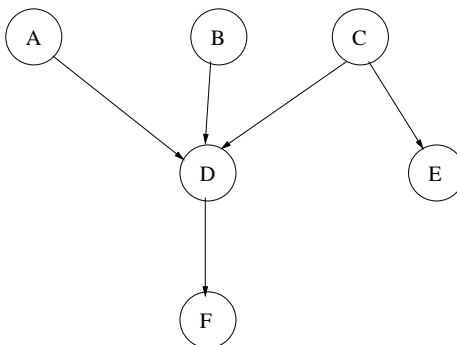


Problem assignment 7

Due: Wednesday, March 28, 2007

Problem 1. Bayesian belief networks

Assume the Bayesian belief network in the figure below. Assume that every variable in the network is binary representing T,F values, except variable D that can take on three possible values T,F,X (X stands for undecided).



The belief network encodes the full joint distribution over random variables (represented by nodes) by exploiting conditional independences that hold among variables.

Part a. Give examples of at least five independences that hold among variables in the networks.

Part b. What is the number of parameters needed to define the full joint distribution over variables in the problem domain without the belief network representation?

Part c. Show how to compute the joint probability $P(A=T, B=T, C=F, D=F, E=F, F=T)$ using the parameters of the belief network model.

Part d. What is the number of parameters needed to define the belief network in the figure?

Part e. Assume you want to compute $P(B = T, E = T)$. Show how would you compute the expression more efficiently. You do not have to provide counts of operations, it is sufficient to illustrate what you would do to improve the efficiency of the inference.

Problem 2. Pneumonia diagnosis.

Assume a Bayesian belief network with the Naive-Bayes structure for a simplified version of the pneumonia problem. The variable Pneumonia is at the root of the Naive Bayes structure and features (attributes) Fever, Paleness, Cough and HighWBCcount are conditionally independent given Pneumonia. Assume that random variables in our model are discrete with the following set of values:

- **Pneumonia:** True, False
- **Fever:** True, False
- **Paleness:** True, False
- **Cough:** True, False
- **HighWBCcount:** True, False

Part a. Write a program `main7learning.m` that takes the data in the file 'pneumonia.tex' and learns the ML estimates of parameters of the BBN network. The pneumonia file includes 500 examples (in rows). The features are in columns, in the following order: Fever, Paleness, Cough, HighWBCcount. Last column represents Pneumonia variable. The data are represented such that True corresponds to 1 and False to 0. Report the parameters of the network. You do not have to submit the program for this part.

Part b. Assume that the patient comes with the following set symptoms: Fever and Cough are true; Paleness and HighWBCcount are false. What is the probability $P(Pneumonia = T | Fever = T, Paleness = F, Cough = T, HighWBCcount = F)$, that is, the probability that the patient suffers from Pneumonia, given the symptoms? Simplify the expression as much as possible before plugging in the values.

Part c. Assume that the patient reports Cough and a Fever (they are true); and values of Paleness and HighWBCcount are not known. Simplify the expression as much as possible. Compute the probability $P(Pneumonia = T | Fever = T, Cough = T)$ of the patient suffering from the pneumonia, given the symptoms?

Part d. Write and submit a Matlab program `main7inference` that reads in a combination of patient symptoms (the values of Fever, Paleness, Cough, HighWBCcount) from the file `example.txt` and computes the probability of $P(Pneumonia = True | CurrentSymptoms)$. The values of symptoms for the current patient case are given in the following order:

Fever, Paleness, Cough, HighWBCcount.

The values are encoded as follows:

- 0 for False,
- 1 for True, and
- -1 for not given (not known).

For example, the symptoms of the patient case in part b are encoded as a vector of 4 values:
(1 0 1 0).
and symptoms in part c as:
(1 -1 1 -1).

Note: Please note that your program should work on an arbitrary combination of input values in example.txt file.