Assignment 8

Written Assignment - Posterior Proabilites and Bayesian Networks

The assignment should be submitted via Blackboard.

Task 1

30 points

You are a meteorologist that places temperature sensors all of the world, and you set them up so that they automatically e-mail you, each day, the high temperature for that day. Unfortunately, you have forgotten whether you placed a certain sensor S in Maine or in the Sahara desert (but you are sure you placed it in one of those two places) . The probability that you placed sensor S in Maine is 5%. The probability of getting a daily high temperature of 80 degrees or more is 20% in Maine and 90% in Sahara. Assume that probability of a daily high for any day is conditionally independent of the daily high for the previous day, given the location of the sensor.

Part a: If the first e-mail you got from sensor S indicates a daily high over 80 degrees, what is the probability that the sensor is placed in Maine?

Part b: If the first e-mail you got from sensor S indicates a daily high over 80 degrees, what is the probability that the second e-mail also indicates a daily high over 80 degrees?

Part c: What is the probability that the first three e-mails all indicate daily highs over 80 degrees?

Task 2

10 points.

In a certain probability problem, we have 11 variables: A, B_1, B_2, ..., B_{10}.

- Variable A has 6 values.
- Each of variables B_1, ..., B_{10} have 5 possible values. Each B_i is conditionally indepedent of all other 9 $B_j$variables (with j != i) given A.

Based on these facts:

Part a: How many numbers do you need to store in the joint distribution table of these 11 variables?

Part b: What is the most space-efficient way (in terms of how many numbers you need to store) representation for the joint probability distribution of these 11 variables? How many numbers do you need to store in your solution? Your answer should work with any variables satisfying the assumptions stated above.

Task 3

10 points

George doesn't watch much TV in the evening, unless there is a baseball game on. When there is baseball on TV, George is very likely to watch. George has a cat that he feeds most evenings, although he forgets every now and then. He's much more likely to forget when he's watching TV. He's also very unlikely to feed the cat if he has run out of cat food (although sometimes he gives the cat some of his own food). Design a Bayesian network for modeling the relations between these four events:

- baseball_game_on_TV
- George_watches_TV
- out_of_cat_food
- George_feeds_cat

Your task is to connect these nodes with arrows pointing from causes to effects. No programming is needed for this part, just include an electronic document (PDF, Word file, or OpenOffice document) showing your Bayesian network design.

Task 4

20 points

For the Bayesian network of Task 3, the text file at this link contains training data from every evening of an entire year. Every line in this text file corresponds to an evening, and contains four numbers. Each number is a 0 or a 1. In more detail:

- The first number is 0 if there is no baseball game on TV, and 1 if there is a baseball game on TV.
- The second number is 0 if George does not watch TV, and 1 if George watches TV.
- The third number is 0 if George is not out of cat food, and 1 if George is out of cat food.
- The fourth number is 0 if George does not feed the cat, and 1 if George feeds the cat.

Based on the data in this file, determine the probability table for each node in the Bayesian network you have designed for Task 3. You need to include these four tables in the drawing that you produce for question 3. You also need to submit the code/script that computes these probabilities.

Task 5

30 points.

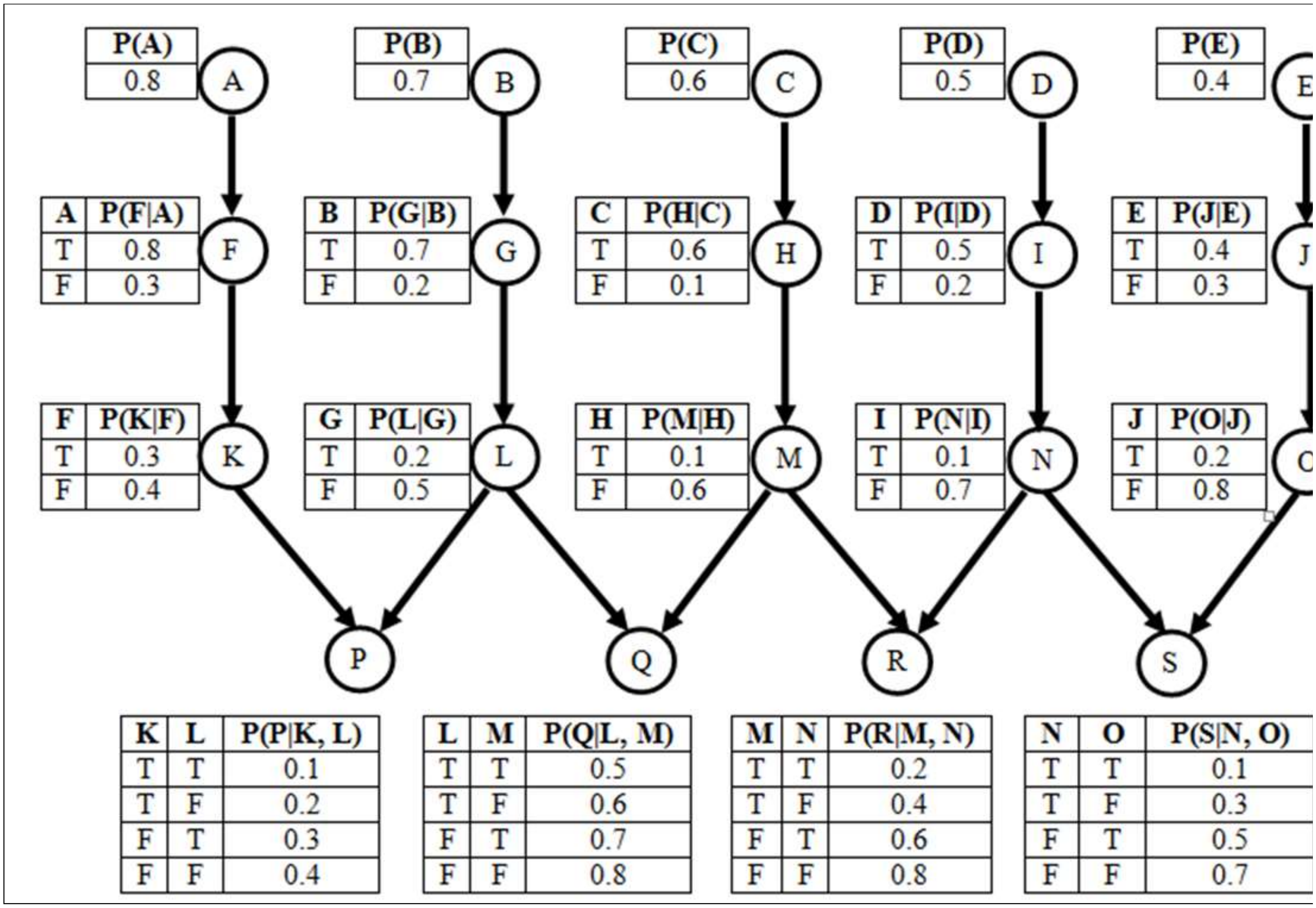

Figure 1: Yet another Bayesian Network.

Part a: On the network shown in Figure 2, what is the Markovian blanket of node N?

Part b: On the network shown in Figure 2, what is P(I, D)? How is it derived?

Part d: On the network shown in Figure 2, what is P(M, not(C) | H)? How is it derived?

Other Instructions

- The answers for the tasks can be typed as a document or handwritten and scanned.
  - Accepted document formats are (.pdf, .doc or .docx). Please do not submit .txt files. If you are using OpenOffice or LibreOffice, make sure to save as .pdf or .doc
  - If you are scanning handwritten documents make sure to scan it at a minimum of 600dpi and save as a .pdf or .png file.
- For Task 4 also submit whatever method you used to calculate the probabilities. (C, Python or Java Code, Matlab or shell script, Excel spreadsheet etc). Don't bother making sure it runs on OMEGA
  - The probabilty values can also just be written along with answers to Task 3.
- Zip all the files together into assignment8_<netid>.zip. Submit on Blackboard.