Assignment 5

Written Assignment - Posterior Proabilites and Bayesian Networks

Max points:

- CSE 4308: 125
- CSE 5360: 125

The assignment should be submitted via Canvas

Task 1

20 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

Task 4

10 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

20 points.

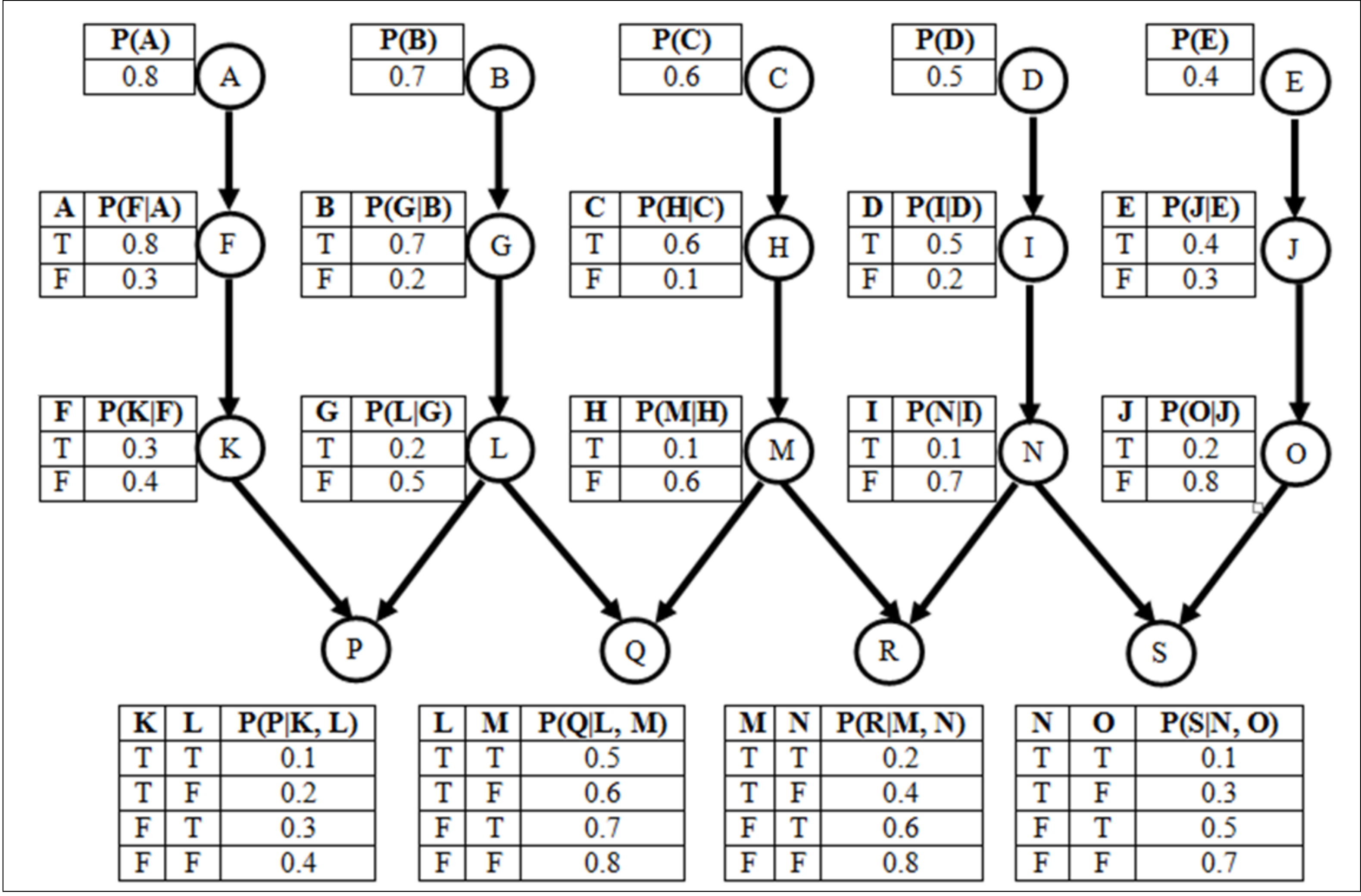

Figure 1: Yet another Bayesian Network.

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?

Task 6

25 points

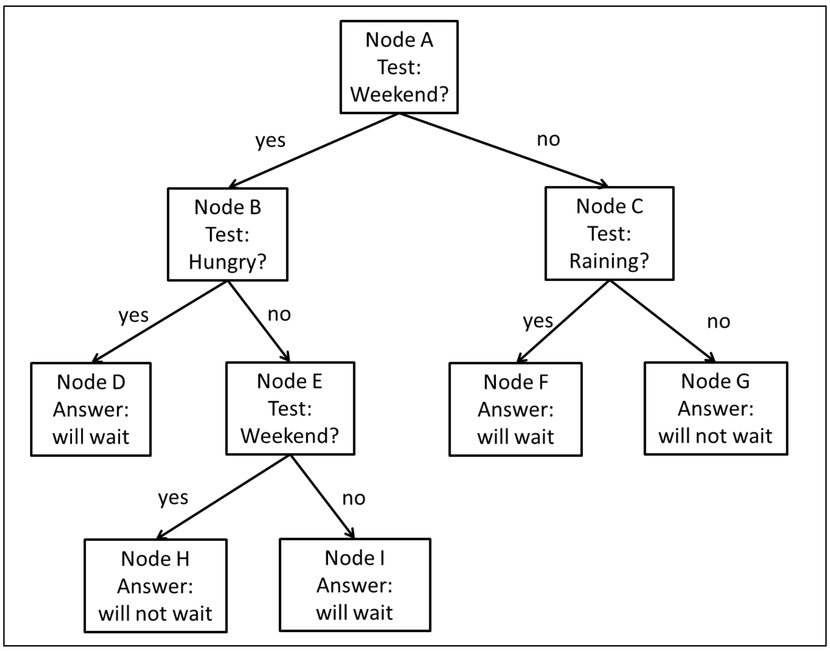

Figure 2: A decision tree for estimating whether the patron will be willing to wait for a table at a restaurant.

Part a: Suppose that, on the entire set of training samples available for constructing the decision tree of Figure 1, 80 people decided to wait, and 20 people decided not to wait. What is the initial entropy at node A (before the test is applied)?

Part b: As mentioned in the previous part, at node A 80 people decided to wait, and 20 people decided not to wait.

- Out of the cases where people decided to wait, in 20 cases it was weekend and in 60 cases it was not weekend.
- Out of the cases where people decided not to wait, in 15 cases it was weekend and in 5 cases it was not weekend.

What is the information gain for the weekend test at node A?

Part c: In the decision tree of Figure 1, node E uses the exact same test (whether it is weekend or not) as node A. What is the information gain, at node E, of using the weekend test?

Part d: We have a test case of a hungry patron who came in on a rainy Tuesday. Which leaf node does this test case end up in? What does the decision tree output for that case?

Part e: We have a test case of a not hungry patron who came in on a sunny Saturday. Which leaf node does this test case end up in? What does the decision tree output for that case?

Task 7

20 points

Class	A	B	C
X	1	2	1
X	2	1	2
X	3	2	2
X	1	3	3
X	1	2	2
Y	2	1	1
Y	3	1	1
Y	2	2	2
Y	3	3	1
Y	2	1	1

We want to build a decision tree that determines whether a certain pattern is of type X or type Y. The decision tree can only use tests that are based on attributes A, B, and C. Each attribute has 3 possible values: 1, 2, 3 (we do not apply any thresholding). We have the 10 training examples, shown on the table (each row corresponds to a training example).

What is the information gain of each attribute at the root? Which attribute achieves the highest information gain at the root?

Task 8

10 points

Suppose that, at a node N of a decision tree, we have 1000 training examples. There are four possible class labels (A, B, C, D) for each of these training examples.

Part a: What is the highest possible and lowest possible entropy value at node N?

Part b: Suppose that, at node N, we choose an attribute K. What is the highest possible and lowest possible information gain for that attribute?