Naman Jain Vimal Kumar
1001670153

Problem 1) 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.

Let:

M – Maine
S – Sahara
T – Temperature higher than 80 degrees

Given:

$$p(M) = 0.05$$
$$p(T \mid M) = 0.20$$
$$p(T \mid S) = 0.90$$

Some Calculations:

$$p(S) = 1 - p(M)$$
$$p(S) = 1 - 0.05$$
$$\boldsymbol{p(S) = 0.95}$$

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?

To find:

$$p\ (M \mid T) = \frac{p(M) * p(T \mid M)}{p(T)}$$

$$p\ (M \mid T) = \frac{p(M) * p(T \mid M)}{(p(S) * p(T| S) + (p(M) * p(T \mid M))}$$

$$p\ (M \mid T) = \frac{0.05 * 0.20}{(0.95 * 0.90) + (0.05 * 0.20)}$$

$$\boldsymbol{p(M \mid T) = 0.0115}$$

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?

To Find:

$$p(T_1 \mid T_0) = \frac{p(T_1, T_0)}{P(T_0)}$$

$$p(T_0) = (p(T \mid M) * p(M)) + (p(T \mid S) * p(S))$$
$$p(T_0) = (0.20 * 0.05) + (0.90 * 0.95)$$
$$p(T_0) = (0.01 + 0.855)$$
$$p(T_0) = 0.865$$

$$p(T_1, T_0) = (p(T_1, T_0 \mid M) * p(M)) + (p(T_1, T_0 \mid S) * p(S))$$
$$p(T_1, T_0) = (p(T_1 \mid M) * p(T_0 \mid M) * p(M)) + (p(T_1 \mid S) * p(T_0 \mid S) * p(S))$$
$$p(T_1, T_0) = (0.2 * 0.2 * 0.5) + (0.8 * 0.8 * 0.95)$$
$$p(T_1, T_0) = (0.02 + 0.608)$$
$$p(T_1, T_0) = 0.628$$

From, $p(T_1 \mid T_0) = \frac{p(T_1, T_0)}{P(T_0)}$

$$p(T_1 \mid T_0) = \frac{0.628}{0.865}$$
$$\boldsymbol{p(T_1 \mid T_0) = 0.726}$$

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

We know that,

$$p(T_0 \cap T_1 \cap T_2) = (p(T_0, T_1, T_2 \mid M) * p(M)) + (p(T_0, T_1, T_2 \mid S) * p(S))$$
$$p(T_0 \cap T_1 \cap T_2) = (p(T_0 \mid M) * p(T_0 \mid M) * p(T_0 \mid M) * p(M)) + (p(T_0 \mid S) * p(T_0 \mid S) * p(T_0 \mid S) * p(S))$$
$$p(T_0 \cap T_1 \cap T_2) = 0.2 * 0.2 * 0.2 * 0.05 + 0.9 * 0.9 * 0.9 * 0.95$$
$$p(T_0 \cap T_1 \cap T_2) = 0.0004 + 0.69255$$
$$\boldsymbol{p(T_0 \cap T_1 \cap T_2) = 0.69295}$$

Problem 2) In a certain probability problem, we have 11 variables: A, B1, B2, ..., B10.

- ***Variable A has 6 values.***
- ***Each of variables B1, ..., B10 have 5 possible values. Each Bi is conditionally independent of all other 9 B_j variables (with j != i) given A.***

Based on these facts:

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

Possible values of $A = 6$
Possible values of B_1 to $B_{10} = 5^{10}$

There, numbers to be stored in joint distribution table = $\boldsymbol{6 * 5^{10}}$

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.***

Each $p(B_i \mid A)$ $will\ have$ $6 * (5 - 1) = 24$ $values$
Hence $p(B_1, .., B_{10}) = 24 * 10 = 240$ $values$
For $p(A), (6 - 1) = 5$ $values$

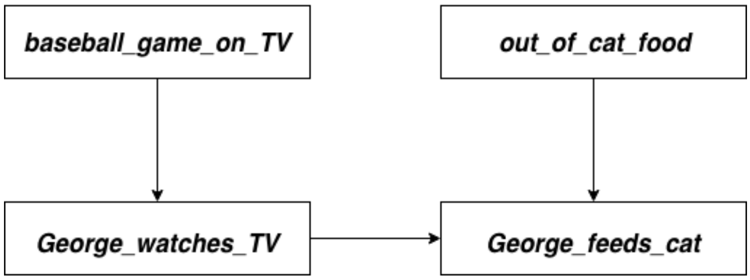
Therefore, total space = 240 + 5 = **245 values**

Problem 3) 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.

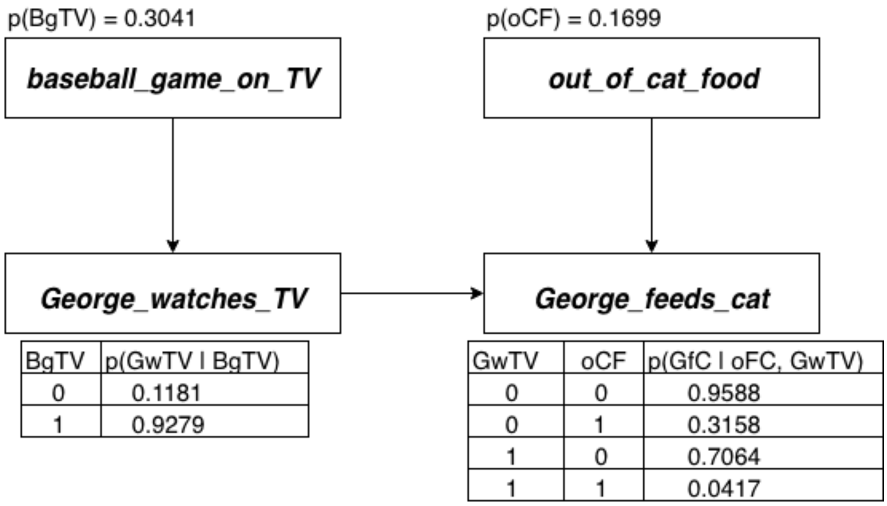


Problem 4) For the Bayesian network of Task 3, the text file at this link contains training data from every evening of an entire year. Every line in this text file corresponds to an evening, and contains four numbers. Each number is a 0 or a 1. In more detail:

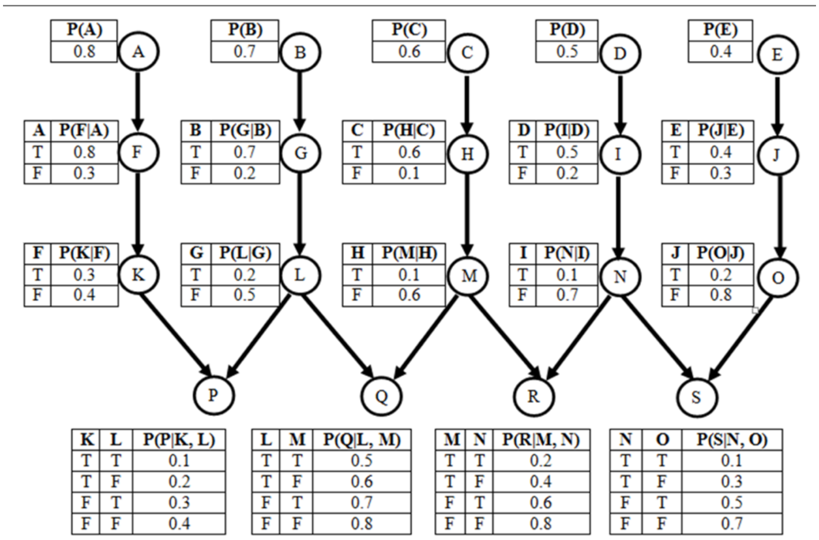
- ***The first number is 0 if there is no baseball game on TV, and 1 if there is a baseball game on TV.***
- ***The second number is 0 if George does not watch TV, and 1 if George watches TV.***
- ***The third number is 0 if George is not out of cat food, and 1 if George is out of cat food.***
- ***The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

Naman Jain Vimal Kumar
1001670153



Problem 5)



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

Following belongs Markov Blanket of 'N':

- Parents of the node: 'I'
- Children of the node: 'R', 'S'
- Other parents of the children: 'M', 'O'

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

$$p(I, D) = p(I) * p(D)$$
$$p(I, D) = p(I \mid D) * p(D)$$
$$p(I, D) = 0.5 * 0.5$$
$$\boldsymbol{p(I, D) = 0.25}$$

c. ***On the network shown in Figure 2, what is P(M, not(C) | H)? How is it derived?***

Numerator:

$$p(M, not\ C \mid H) = p(M) * p(not\ C) * p(H)$$
$$p(M, not\ C \mid H) = p(M \mid H) * p(not\ C) * p(H \mid not(C))$$

$$p(M, not\ C \mid H) = 0.1 * (1 - 0.6) * \ 0.1$$
$$p(M, not\ C \mid H) = 0.004$$

Denominator:

$$p(H) = p(H \mid C) + p(H \mid not\ C)$$
$$p(H) = 0.6 + 0.1$$
$$p(H) = 0.7$$

Therefore:

$$p(M, not\ C \mid H) = {}^{0.004}/_{0.7}$$
$$\boldsymbol{p(M, not\ C \mid H) = 0.005714}$$