

## COMP-424: Artificial intelligence

### Homework 3

Due on *myCourses* on March 22<sup>nd</sup>, 2018 at 11:59pm.

#### General instructions.

- This is an individual assignment. You can discuss solutions with your classmates, but should only exchange information orally, or else if in writing through the discussion board on *myCourses*. All other forms of written exchange are prohibited.
- Unless otherwise mentioned, the only sources you should need to answer these questions are your course notes, the textbook, and the links provided. Any other source used should be acknowledged with proper referencing style in your submitted solution.
- Submit a single pdf document containing all your pages of your written solution on your McGill's *myCourses* account. You can scan-in hand-written pages. If necessary, learn how to combine many pdf files into one.
- You may solve the questions by hand or by writing a program, but if you write a program, you must not rely on existing implementations, and must do it from scratch (and must submit your code along with the pdf).

#### Question 1: Designing a Bayesian Network

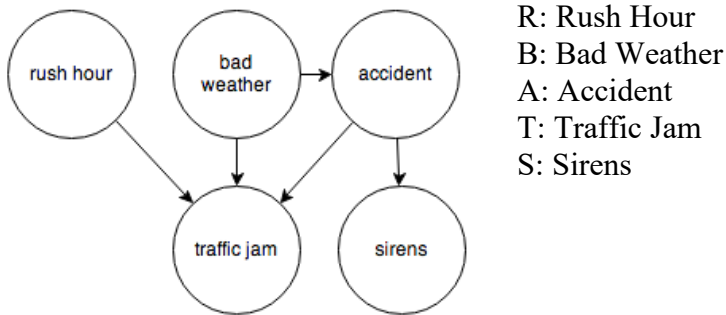
(Reproduced from Russell & Norvig, Exercise 14.11, pg. 561)

In your local nuclear power station, there is an alarm that senses when a temperature gauge exceeds a given threshold. The gauge measures the temperature of the core. Consider the Boolean variables:  $A$  (alarm sounds),  $F_A$  (alarm is faulty), and  $F_G$  (gauge is faulty), and the multivalued nodes  $G$  (gauge reading) and  $T$  (actual core temperature).

- Draw a Bayesian network for this domain, given that the gauge is more likely to fail when the core temperature gets too high.
- Is your network a polytree? Why or why not?
- Suppose there are just two possible actual and measured temperatures, normal and high; the probability that the gauge gives the correct temperature is  $x$  when it is working, but  $y$  when it is faulty. Give the conditional probability table associated with  $G$ .
- Suppose the alarm works correctly unless it is faulty, in which case it never sounds. Give the conditional probability table associated with  $A$ .
- Suppose the alarm and gauge are working and the alarm sounds. Calculate an expression for the probability that the temperature of the core is too high, in terms of the various conditional probabilities in the network.

## Question 2: Inference in Bayesian Networks

Consider the following Bayesian Network



We will denote random variables with capital letters (e.g., R), and the binary outcomes with lowercase letters (e.g., r, and  $\neg r$ ).

The network has the following parameters:

$$P(b)=0.2$$

$$P(r)=0.15$$

$$P(t|r, b, a) = 0.95$$

$$P(t|r, \neg b, a) = 0.92$$

$$P(t|r, b, \neg a) = 0.90$$

$$P(t|r, \neg b, \neg a) = 0.85$$

$$P(t|\neg r, b, a) = 0.35$$

$$P(t|\neg r, b, \neg a) = 0.4$$

$$P(t|\neg r, \neg b, a) = 0.6$$

$$P(t|\neg r, \neg b, \neg a) = 0.05$$

$$P(s|a) = 0.8$$

$$P(s|\neg a) = 0.2$$

$$P(a|b) = 0.6$$

$$P(a|\neg b) = 0.4$$

Compute the following terms using basic axioms of probability and the conditional independency properties encoded in the above graph. You can use Bayes Ball properties to simplify the computation, if applicable.

- $P(s, r)$
- $P(a, \neg t)$
- $P(t|s)$

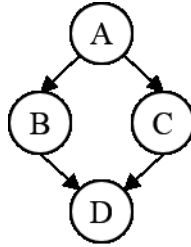
## Question 3: Variable Elimination

For the graph above, compute the MAP result of querying  $P(T|a)$  using variable elimination with the following order: R,S,B,A,T.

Clearly explain each step. For each of the intermediate factors created, explain what probabilistic function it represents.

## Question 4: Learning with Bayesian Networks

Consider the following Bayesian network. Assume that the variables are distributed according to Bernoulli distributions.



- a. We are given the following dataset with 146 samples, from which we will estimate the parameters of the model.

| A | B | C | D | # Instances |
|---|---|---|---|-------------|
| 0 | 0 | 0 | 0 | 1           |
| 0 | 0 | 0 | 1 | 4           |
| 0 | 0 | 1 | 0 | 20          |
| 0 | 0 | 1 | 1 | 4           |
| 0 | 1 | 0 | 0 | 34          |
| 0 | 1 | 0 | 1 | 2           |
| 0 | 1 | 1 | 0 | 32          |
| 0 | 1 | 1 | 1 | 0           |
| 1 | 0 | 0 | 0 | 1           |
| 1 | 0 | 0 | 1 | 0           |
| 1 | 0 | 1 | 0 | 1           |
| 1 | 0 | 1 | 1 | 4           |
| 1 | 1 | 0 | 0 | 10          |
| 1 | 1 | 0 | 1 | 19          |
| 1 | 1 | 1 | 0 | 14          |
| 1 | 1 | 1 | 1 | 0           |

- Enumerate the parameters that must be learned. Specify the parameter name and the probability that it represents (i.e., for each parameter, write something in the form,  $\theta_X = \Pr(X)$ ).
- Give the maximum likelihood estimate for each parameter.
- Give the maximum a posterior estimate for each parameter after applying Laplace smoothing.

- b. Assume that in addition to the data in the table above, you are given the following incomplete data instances:

|    | <b>A</b> | <b>B</b> | <b>C</b> | <b>D</b> |
|----|----------|----------|----------|----------|
| S1 | 1        | 0        | 1        | ?        |
| S2 | 1        | 1        | ?        | 0        |

We will apply the (soft) EM algorithm on these instances. Initialize the model using your parameter estimates from part a subpart ii (i.e., use the MLE).

- i. Show the computation of the first E-step, providing the weights for each possible assignment of the incomplete data for each sample.
- ii. What are the parameters obtained for the first M-step? Weight each of the samples from the original dataset and the two new samples equally (i.e., you now have 148 samples).
- iii. Show the computation of the second E-step.