UNIVERSITY OF SASKATCHEWAN

Department of Computer Science
176 Thorvaldson Building
110 Science Place, Saskatoon, SK, S7N 5C9, Canada
Telephine: (306) 966-4886, Facimile: (306) 966-4884

Assignment 3

Bayesian Networks

Date Due: March 16, 2020, 5pm **Total Marks: 57**

General Instructions

- **This assignment is individual work.** You may discuss questions and problems with anyone, but the work you hand in for this assignment must be your own work.
- Each question indicates what to hand in.
- **Assignments must be submitted to Moodle.**
- Assignments will be accepted until 11pm without penalty.
- Use of Python and Jupyter Notebooks is highly encouraged, even for questions 1-5.

Version History

- **6/03/2020**: released to students

Notation

The language of probability is mathematical, and notational conventions vary. The following conventions are useful and practical for some concepts, but they may be unfamiliar.

- Capital letters will represent random variables, e.g., "Let X denote the result of a coin toss, $X{=}0$ for tails, $X{=}1$ for heads." One might call X a variable, but that's imprecise, because it is not varying.

  Specific outcomes will be represented by assertions of the form $X{=}0$. Typically, discrete outcomes are represented by integers, but sometimes short strings are used, e.g., $X{=}'\mathrm{tails}'$.

- We will use the probability notation $P()$ in two ways. The most familiar is $P(X{=}1)$. This is a probability, a non-negative value between 0 and 1.

  We will also write $P(X)$, by which we mean a probability distribution over the outcome space of X. It is short hand notation for $\forall x\, P(X{=}x)$. Similarly, $P(XY)$ is a joint probability distribution over the joint outcome space of X and Y, and $P(X|Y)$ is a conditional probability distribution over the joint outcome space of X and Y.

  We prefer to write $P(X|Y)$ because $\forall x \forall y\, P(X{=}x|Y{=}y)$ is cluttered and contributes nothing to any derivations.

- One of the important operations is summation or integration over a probability distribution. To cut down on clutter, we'll write formulae like this:

$$P(X) = \sum_{Y} P(X|Y)P(Y)$$

  instead of the more precise but cluttered version:

$$\begin{aligned} P(X) &= \sum_{y \in Y} P(X|Y{=}y)P(Y{=}y) \\ &= P(X|Y{=}0)P(Y{=}0) + P(X|Y{=}1)P(Y{=}1) \end{aligned}$$

  where $y \in Y$ means "All the outcomes for Y" (here, just 0 and 1). The same informal notation works for integration of continuous probability mass functions.

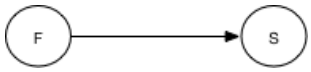
- When lots of outcome labels are involved, we will use adjacency to imply a joint outcome. Sometimes a comma will be used to help clarify. For example, $P(XY) = P(X, Y)$.

Question 1 (6 points):

Purpose: Simple calculations as a warm-up.

Competency Level: Basic

The following diagram represents a simple, 2 node Bayesian Network.



The nodes represent outcomes F and S. The Conditional Probability Distributions (CPDs) for these nodes are as follows:

			$P(S{=}0\|F{=}0)$	=	0.82
$P(F{=}0)$	=	0.13	$P(S{=}1\|F{=}0)$	=	0.18
$P(F{=}1)$	=	0.87	$P(S{=}0\|F{=}1)$	=	0.07
			$P(S{=}1\|F{=}1)$	=	0.93

Using this information, answer the following questions.

(a) Calculate $P(S{=}0|F{=}0)$. Hint: Look up the value in the CPD.

(b) Calculate $\sum_S P(S|F{=}0)$.

(c) Calculate $P(F{=}0)$. Hint: Look up the value.

(d) Calculate $\sum_S P(S|F{=}0)P(F{=}0)$. Use your calculations from previous steps.

> **Note: Take care in the algebra here. There is a correct way and an incorrect way to get the same answer. Remember that multiplication has higher precedence than addition!**

This calculation illustrates the meaning of irrelevance. The rule for irrelevance in a Bayesian network allows us to leave out variables that just will not affect the answer. If you leave irrelevant variables in your calculations, you get the same answer, but you do more work.

(e) Calculate $P(F{=}0|S{=}0)$. Use Bayes Rule. Pay attention specifically to the denominator.

(f) Calculate $P(F{=}1|S{=}0)$. Use Bayes Rule. Pay attention specifically to the denominator.

Notice the similarity in the formulae of this answer and the previous answer. This is why we prefer to derive a symbolic formula, e.g. $P(F|S)$: the same formula can be applied to any outcome for F or S.

What to Hand In

- Six small calculations.

Add these answers to a document for questions 1-5, named `A3.PDF`. You can use Jupyter notebooks, or if you prefer, a LaTeX document.

Evaluation

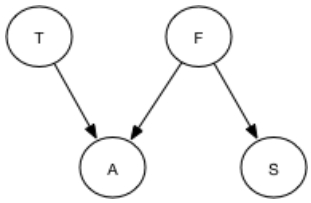
- 1 mark each. The numeric answer has to be correct to get the mark.

Question 2 (6 points):

Purpose: A few more complicated calculations, to make the formulae concrete.

Competency Level: Basic

The following diagram represents a simple, 4 node Bayesian Network.



The nodes represent outcomes F, S, T, A. The Conditional Probability Distributions (CPDs) for these nodes are as follows:

$P(T{=}0)$	=	0.99		$P(F{=}0)$	= 0.13
$P(T{=}1)$	=	0.01		$P(F{=}1)$	= 0.87
$P(A{=}0\|F{=}0,T{=}0)$	=	0.93			
$P(A{=}1\|F{=}0,T{=}0)$	=	0.07			
$P(A{=}0\|F{=}0,T{=}1)$	=	0.02	$P(S{=}0\|F{=}0)$	=	0.82
$P(A{=}1\|F{=}0,T{=}1)$	=	0.98	$P(S{=}1\|F{=}0)$	=	0.18
$P(A{=}0\|F{=}1,T{=}0)$	=	0.006	$P(S{=}0\|F{=}1)$	=	0.07
$P(A{=}1\|F{=}1,T{=}0)$	=	0.994	$P(S{=}1\|F{=}1)$	=	0.93
$P(A{=}0\|F{=}1,T{=}1)$	=	0.25			
$P(A{=}1\|F{=}1,T{=}1)$	=	0.75			

Using this information, answer the following questions.

(a) Calculate $P(A{=}1|F{=}1)$. Hint: Which nodes are not relevant to the query?

(b) Calculate $P(F{=}0|S{=}0)$. Hint: Which nodes are not relevant to the query?

(c) Calculate $P(A{=}1|S{=}1)$. Hint: Which nodes are not relevant to the query?

Hopefully you have the idea that numbers can make calculations seem less abstract, but they don't actually buy us anything, and they're a pain in the neck. From now on, we'll just do the algebra.

What to Hand In

- Three calculations, made odious by keeping track of arithmetic.

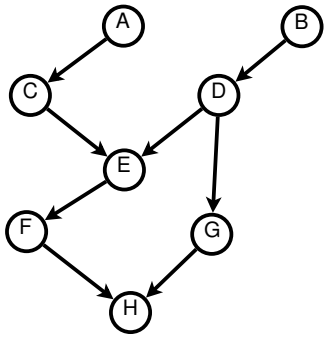
Add these answers to a document for questions 1-5, named `A3.PDF`. You can use Jupyter notebooks, or if you prefer, a LaTeX document.

Evaluation

- 2 marks each. The arithmetic is not important, as long as the values from the CPDs are substituted into a formula derived by using the Variable Elimination procedure. Give two marks if the method is right, even if the calculations or substitutions had errors. No part marks.

Bayesian network structure for Q3-6

Consider the Bayesian network given in the following diagram. Notice that there are families defined by the graph, but I have not given you the numbers for the tables. That's because doing mindless arithmetic is for computers, not people.



This network will form the basis for questions 3-6 in this assignment.

Question 3 (10 points):

Purpose: To practice reading conditional independence from a Bayesian Network.

Competency Level: Basic

Consider the Bayesian network given above. Answer the following statements about conditional independence. For each one, show the paths, and say whether the path is active or blocked.

1. (2 marks) Given $\{\}$ (i.e., the empty set), is A conditionally independent of D? (Because the set is empty, we say "unconditionally independent," or just "independent.")
2. (2 marks) Given H, is C conditionally independent of B?
3. (2 marks) Given F, is H conditionally independent of A?
4. (2 marks) Given C, is F conditionally independent of B?
5. (2 marks) Given F and D, is C conditionally independent of G?

What to Hand In

- Five statements about conditional independence, accompanied by path analysis for justification.

Add these answers to a document for questions 1-5, named `A3.PDF`. You can use Jupyter notebooks, or if you prefer, a LaTeX document.

Evaluation

- 2 marks each: one for the right answer, and one for a correct justification.

Question 4 (5 points):

Purpose: To practice reading ancestors in a Bayesian network.

Competency Level: Basic

Consider the Bayesian network given above. For each of the following queries, which nodes are relevant, according to the rule for relevance given in class?

1. $P(C|G)$
2. $P(D|E, A)$
3. $P(F)$ (i.e., no evidence)
4. $P(B|D, F)$
5. $P(C|F, G)$

Include in your answer the query and evidence variables (just to avoid ambiguity).

What to Hand In

- Five lists of variables (or nodes).

Add these answers to a document for questions 1-5, named `A3.PDF`. You can use Jupyter notebooks, or if you prefer, a LaTeX document.

Evaluation

- 1 mark each.

Question 5 (16 points):

Purpose: To practice deriving formulae for any query on a Bayesian Network.

Consider the Bayesian network given above. Write out a formula for each of the following queries, using technique called variable elimination shown in class. You cannot do any arithmetic, so leave your solution in terms of the CPDs implied in the Bayesian network. Choose any order for the marginalization; there is no need to determine the "optimal" order. Be sure to consider your answers from the previous questions, as some of them are either relevant directly, or indirectly.

1. $P(C)$ (i.e., no evidence)
2. $P(B|C)$
3. $P(A|H, C)$
4. $P(C|D, F)$

What to Hand In

- Four derivations, showing the application of Variable Elimination, making use of JPD factorization, conditional independence and marginalization of the relevant variables.

Add these answers to a document for questions 1-5, named `A3.PDF`. You can use Jupyter notebooks, or if you prefer, a LaTeX document.

Evaluation

- 4 marks each. Full marks if the answers are correct and use the graph structure to inform the algebra as in questions 3 and 4.

Question 6 (14 points):

Purpose: To work by hand with a simple example of inferring Bayesian Network structure, and model selection.

Competency Level: Intermediate.

On the Assignment 2 page, you'll find a Jupyter Notebook named `A2Q6.ipynb`, which walks through the process of inferring Bayesian Network structure from data. The document has the following steps:

1. Load a categorical dataset (`diamond.csv`)
2. Create 2 (related) Bayesian network models, fitting their parameters to the data.
3. Compare the 2 structures, using log-likelihood.

What to Hand In

- Your Jupyter notebook, exported and named `A3Q6.pdf`

Evaluation

- 5 marks: Step 1: Your notebook calculates the 5 tables correctly, and presents them neatly.
- 4 marks: Step 2: Your calculations are correct.
- 1 mark: Conclusions: Your conclusions are correct.