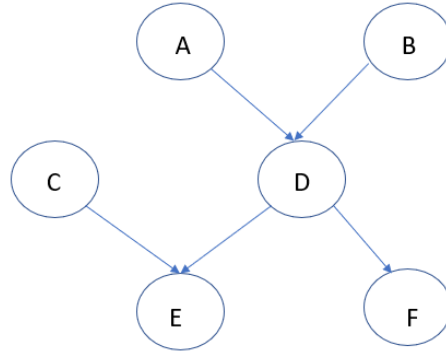


Coursework 1: Bayesian Network

Problem 1. Consider the following Bayesian network with 6 nodes.



1. Use the above structure to factor the joint distribution of $p(A, B, C, D, E, F)$ into different conditional probabilities. A naive factorization is

$$p(A, B, C, D, E, F) = p(A)p(B|A)p(C|A, B)p(D|A, B, C)p(E|A, B, C, D)p(F|A, B, C, D, E).$$

However, this is not the correct answer because it does not use the structure of the Bayesian network. You need to write down the simplest factorization by using the conditional independency induced from the above structure (2 marks).

2. Suppose each random variable takes three states, i.e., A can take values in $\{a_0, a_1, a_2\}$, B can take values in $\{b_0, b_1, b_2\}$ and so on. What is the minimum number of parameters to exactly express the above model, i.e., to express the joint distributions $p(A, B, C, D, E, F)$ (2.5 marks)? Please also justify your answer (2.5 marks). (In total 5 marks)
3. Check whether the following statements are true. If you think the statements are true, then type T , otherwise type F . Please just type five characters for the five questions in order. For example, if you think the first four are true and the last is false, then simply type $TTTTF$. We also need you to justify your answer in the last bullet.
 - A and B are independent (0.5 mark)
 - A and B are conditionally independent given D (0.5 mark)
 - A and E are conditionally independent given D (0.5 mark)
 - E and F are independent (0.5 mark)
 - E and F are conditionally independent given D (0.5 mark). Use the expression of joint distribution to justify your answer, i.e., either prove the independency or show it does not hold (2.5 marks). Hint: there is no need to consider all random variables in the Bayesian network.
4. Suppose each random variable only takes two states, i.e., A can take values in $\{a_0, a_1\}$, B can take values in $\{b_0, b_1\}$ and so on. Suppose the (conditional) probability tables of this model are as follows.

a_0	a_1	b_0	b_1	d_0	d_1	c_0	c_1	e_0	e_1	f_0	f_1
0.1	0.9	0.4	0.6	0.9	0.1	0.3	0.7	0.9	0.1	0.8	0.2
				0.7	0.3			0.8	0.2		
				0.6	0.4			0.7	0.3		
				0.1	0.9			0.1	0.9		
$p(A)$		$p(B)$		$p(D A, B)$		$p(C)$		$p(E C, D)$		$p(F D)$	

Use pomegranate to write a program for this Bayesian network. Use this program to compute the following probabilities. Please keep four digits after the decimal point. (6 marks, 2 marks for the last question)

- $p(E = e_0)$
- $p(A = a_0, B = b_1, C = c_0, D = d_1, E = e_0, F = f_1)$
- $p(A = a_0 | F = f_0)$
- $p(E = e_1 | A = a_0, B = b_0)$
- $p(A = a_0, B = b_0 | E = e_1)$

Problem 2. In this problem, we want you to build a Bayesian network from a mushroom dataset and use it to predict whether a mushroom is poisonous or edible given some physical characteristics (cap-shape, stalk-shape, bruises, gill-size and so on). The training dataset consists of 1120 training examples. Each training example is a vector in $\{0, 1\}^{12}$. For example, a training example can be

$$(1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0).$$

Each training example can be considered as a sampling of a random vector $\mathbf{X} = (X^1, X^2, \dots, X^{12})$, where X^i is a random variable taking values in $\{0, 1\}$ and X^{12} represents whether or not the mushroom is poisonous. We have provided you some code including the loading of packages and the loading of the dataset (code.py).

- Use pomegranate to build a Bayesian network to model the training examples. The Bayesian network has 12 nodes, each of which represents a $X^i, i \in \{1, 2, \dots, 12\}$. Draw the structure of the network, e.g., a graph similar to that in Problem 1 (I use PPT to draw the graph in Problem 1. You can use any drawing software. You can also draw a picture with your pen, take a photo and then insert the photo in your answer sheet). (2 marks)
- What is the minimum number of parameters to represent the joint distribution of this Bayesian network? Please justify your answer. (2 marks)
- Use the constructed Bayesian network to do binary classification. That is, we give you 6 training examples where only the first 11 attributes are given. You need to output the value of the last random variable. In more details, you need to build a classifier $f : (X^1, \dots, X^{11}) \mapsto X^{12}$. (3 marks)
 - $(X^1, \dots, X^{11}) = (0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1)$ (0.5 mark)
 - $(X^1, \dots, X^{11}) = (0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1)$ (0.5 mark)
 - $(X^1, \dots, X^{11}) = (0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1)$ (0.5 mark)
 - $(X^1, \dots, X^{11}) = (1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0)$ (0.5 mark)
 - $(X^1, \dots, X^{11}) = (1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1)$ (0.5 mark)
 - $(X^1, \dots, X^{11}) = (0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1)$ (0.5 mark)

Please explicitly write your predictions (0 or 1) in your answers.

Note:

- Please submit a single zip file containing a PDF file and a python code. The PDF file should contain your answers to Problems 1 and 2, and may contain scanned images. You should use code to solve Problem 1(4) and Problem 2. The code will be considered in the final marks.

- The PDF file should be named as SIS_ID.PDF and the python code should be named as SIS_ID.py. SIS_ID is your student ID.
- Please also write down your SIS_ID in the PDF file.