Probabilistic Graphical Models
Assignment 2

Issued on: February 5, 2020 | Due by: February 12, 2020 (6pm)

Guidelines for submission

Theory Problems:

- Solutions should preferably be submitted in hard copy (written solutions on A4 sheets). A submission box will be placed before deadline.
- Alternatively, a solution can be prepared in doc/latex as well. For that please export it in .pdf format (as Theory.pdf).

Programming Problems:

- You can use python/matlab for programming problems.
- Along with the main code file, please submit all required dependencies.
- Also add a report (as Code.pdf) with a brief summary of your solution.

Submit a A2_RollNo.zip file on backpack with all required files.

---

1. (5 points) Consider a Gibbs distribution as :

$$P(x_1, x_2, x_3, x_4, x_5) \propto \phi_{12}(x_1, x_2)\phi_{13}(x_1, x_3)\phi_{14}(x_1, x_4)\phi_{23}(x_2, x_3)\phi_{25}(x_2, x_5)\phi_{45}(x_4, x_5) \quad (1)$$

   (a) Show it in form of an undirected graph.
   (b) List neighbours of all nodes in the graph.
   (c) $(x_3 \perp x_4 | x_1, x_2)$, Is this relation True or False. Justify briefly.
   (d) Find markov blanket of all nodes.
   (e) Find set of variables Z, such that $(x_1 \perp x_5 | Z)$

2. (4 Points) Given a network available in Figure-1, with P being the probability distribution associated with the graph:

   (a) Write the Gibbs distribution with all potential factors.
   (b) Can we say that $P(x_3|x_2, x_4) = P(x_3|x_4)$? If yes, can we also say that $P(x_3|x_2) = P(x_3)$? If no, how can we use global markov property to make x_3 and x_2 independent?
   (c) Justify why $(x_1 \perp x_5 | x_2, x_3, x_4, x_6)$ holds.
   (d) Using local markov property, simplify $P(x_3|x_1, x_2, x_4, x_5, x_6)$.

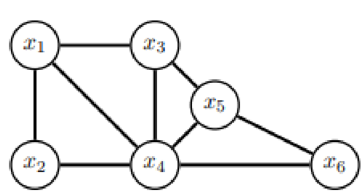


Figure 1: Network for Problem-2

3. (4 Points) We have a bayesian network as shown in Figure-2.

   (a) What will be the standard factorization for the graph? List nodes with no parents/ancestors.

   (b) Considering that there is no conditioning, for which pairs (j,k) do $X_j \perp X_k$ satisfies?

   (c) Let we have a set $A = \{X_2, X_9\}$ (shown with shaded nodes) that satisfies $(X_1 \perp S|A)$. Find the largest possible set S for which the condition holds.

   (d) Also find the set R for which $(X_8 \perp R|A)$ satisfies.

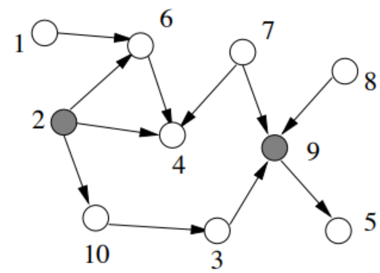


Figure 2: Network for Problem-3

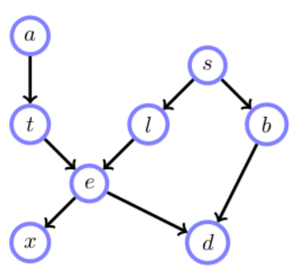
4. (2 Points) The Chest Clinic network concerns the diagnosis of lung disease (tuberculosis, lung cancer, or both, or neither), see Figure 3. In this model a visit to Asia is assumed to increase the probability of tuberculosis. State if the following conditional independence relationships are true or false. Also Provide a very brief justification of your answer.

   (a) Tuberculosis ⊥ Smoking | Shortness of breath

   (b) Lung cancer ⊥ Bronchitis | Smoking

   (c) Visit to Asia ⊥ Smoking | Lung cancer

   (d) Visit to Asia ⊥ Smoking | Lung cancer, Shortness of breath



x = Positive X-ray
d = Dyspnea (Shortness of breath)
e = Either Tuberculosis or Lung Cancer
t = Tuberculosis
l = Lung Cancer
b = Bronchitis
a = Visited Asia
s = Smoker

Figure 3: Belief network structure for the Chest Clinic example

5. (1 Points) Imagine that you live in a town of 10,000 people out of which 1% of the population has a deadly disease. After a health checkup your doctor told you that your test came positive for this disease. Doctor also mentioned that this test is 99% accurate. What are the chances that you actually have this deadly disease?

Probabilistic Graphical Models
Assignment 2

Issued on: February 5, 2020 | Due by: February 12, 2020 (6pm)

Programming Problem:
(4 Points) Given a Bayesian Network, and several queries in the form of $X\ Y \mid Z$ where X, Y are two query nodes and Z is a set of observed nodes, write a program which checks whether X and Y are d-separated given Z and prints True or False. Input and output descriptions are given below.

Input:
3 2 2
A B
B C
A C | B
C A | 0

Output:
True
False

Input Description:
1. First line should take input as $V\ E\ Q$ where V and E are the number of nodes and edges in the Bayesian Network respectively and Q is the number of d-separation queries that will follow.
2. Next E lines denotes a directed edge in the graph. In above example $A\ B$ denotes a directed edge $A \rightarrow B$.
3. Next Q lines denotes queries, in above example queries are:
(1) Are A and C d-separated given B? (True)
(2) Are C and A d-separated? Here 0 means empty set Φ (False)

Hint: You can approach this by checking if an active trail exists between nodes.

Note: Please implement d-separation algorithm from scratch. Use of existing probabilistic programming libraries like pymc3 is not permitted.