COL 776 - Practice Questions

November 21, 2015

Notes:

1. Suppose you are the system administrator for a server room running safety critical applications. It is important for the temperature in the room to be below a certain threshold for the servers to function properly and you have installed a system to monintor the temperature on a continuous basis. When the temperature rises above the desired threshold, the system sends you a text message. The temperature in the room could rise above the desired threshold under the following two circumstances a) Someone leaves the door open for a long time b) the AC in the server room stops working. There is a security alarm in the room which goes off when the door is kept open for a long time. We will model this problem using a Bayesian network. Use the following (binary) variables:
   T: Temperature in the server room rises above the desired threshold
   O: Someone leaves the door open for a long time
   A: Air conditioner stops working
   M: A text message is received on the phone
   S: The security alarm goes off.

   (a) Draw a Bayesian network which precisely encodes the conditional independences implied by the statements above. Also, write the expression for the joint probability distribution.

   (b) For each of the following conditional independence statements, state whether they are true or false based on the network structure you drew in the part above. Justify your answers briefly.

      i. O and A are independent
      ii. O and A are independent given M
      iii. S and M are independent
      iv. S and M are independent given O

2. Consider the following Markov network structure:

```
A-----B
|     |
|     |
D-----C
```

   Let the only clique potentials be those defined over edges in the network. Let the potentials be given as below (this is inspired by the misconception example covered in the book).
   $P(A, B, C, D) = \frac{1}{Z}\phi_1(A, B)\phi_2(B, C)\phi_3(C, D)\phi_4(D, A)$

A	B	phi1(A,B)
0	0	100
0	1	1
1	0	1
1	1	100

B	C	phi2(B,C)
0	0	1
0	1	100
1	0	100
1	1	1

C	D	phi3(C,D)
0	0	1
0	1	100
1	0	100
1	1	1

D	A	phi4(D,A)
0	0	100
0	1	1
1	0	1
1	1	100

   Calculate the probability of $P(A = 1|B = 1, C = 1)$. You can produce the final answer as a simplified fraction.

3. Suppose you like to go out for a movie whenever a) there is a new movie in the town or b) you just finished the exams. For both of these cases (and their combinations thereof), you have certain probability of going out to watch a movie.

   - Design an appropriate Bayesian network (along with the CPDs) to model this problem. Use the variables G (going out for a movie), O (exams are over) and N (new movie in the town). Come up with reasonable numbers for your conditional distributions. Is $G \perp O \mid N$?
   - Suppose now that you necessarily go out for a movie whenever there is a new movie in the town. Is it now the case that $G \perp O \mid N = 1$. Argue. These kind of independences are called context specific independences.

4. Consider two events α and β such that $P(\alpha) = p_\alpha$ and $P(\beta) = p_\beta$.

   - Suppose you do not have any additional knowledge about how α and β are related. What are the minimum and maximum possible values for $P(\alpha \cup \beta)$ and $P(\alpha \cap \beta)$?
   - What can you say if you know that α and β are independent?

5. Consider a distribution specified by a Markov network.

   - Show that multiplying all the entries in any one of the factor tables by a constant $k > 0$ does not change the original distribution.
   - Now, suppose we multiply some of the entries of a factor table by a constant $k_1 > 0$ and the remaining entries by a constant $k_2 > 0$ such that $k_1 \neq k_2$. Assume that all the other factors remain the same as before. Show that this necessarily results in a distribution which is different from the original distribution.

6. Consider a distribution $P(X_1, \ldots, X_n)$ $(n > 2)$ which satisfies conditional independencies (CIs) of the form: $X_k \perp X_l | X_1, X_2, Z \quad \forall k > 2, \forall l > 2, k \neq l, \forall Z \subseteq \mathbf{X} - \{X_1, X_2, X_k, X_l\}$, and no other CI. Here $\mathbf{X} = \{X_1, X_2, \ldots X_n\}$ denotes the set of all the variables.

   (a) Draw a Markov network H to represent $P(\mathbf{X})$ as faithfully as possible. In other words, H should satisfy the maximum number of independences in P while ensuring $I(H) \subseteq I(P)$. Justify your construction.

   (b) Draw the junction tree (with minimum possible size of the largest clique node) corresponding to the network above.

7. For the distribution defined in the previous question

   (a) Draw a Bayesian netwrok B to represent $P(\mathbf{X})$ as faithfully as possible (use the same definition of faithful as above). Justify your construction.

   (b) What can you say about the relationship between $I(B)$ and $I(H)$? Justify your answer.

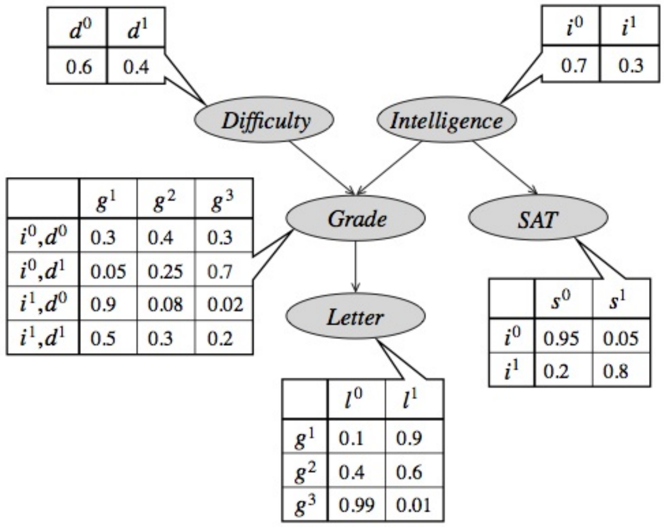
8. Construct an example of a Bayesian network where the corresponding moralized Markov network graph is not chordal.

9. Let X, Y and Z be binary valued random variables. Let z^1 denote the event $Z = 1$. Does $(X \perp Y \mid z^1)$ imply $(X \perp Y \mid Z)$? Prove or provide a counter example.

10. Consider the Bayesian network given in Figure 1 defined over the variables Difficulty(D), Intelligence(I), Grade(G), SAT(S) and Letter (L). Recall that the table associated with each variable node in the network represents its conditional distribution given the values of its parent nodes. Note that the variable Grade in the example considered can take 3 possible values. Calculate the probability $P(I = 0|G = 3)$. You can use the Bayesian network independence property that a node is indepedent of its non-descendants given its parents. Feel free to use a calculator to get the final answer.

11. Let $\mathcal{G}$ denote a Bayesian network structure. Let $\mathcal{G}$ be a perfect I-map for a set of independencies $\mathcal{I}$, i.e. $I(\mathcal{G}) = \mathcal{I}$. Let $\mathcal{G}'$ be a graph obtained by removing an edge in $\mathcal{G}$. Show that $\mathcal{G}'$ can not be an I-map for $\mathcal{I}$.

12. Consider a Bayesian Network graph $\mathcal{G}$ constructed from a Markov Network graph $\mathcal{H}$ using the procedure discussed in the class such that $\mathcal{G}$ is an I-map for $\mathcal{H}$ i.e., $I(\mathcal{G}) \subseteq I(\mathcal{H})$. If $\mathcal{G}$ is also a perfect I-Map for $\mathcal{H}$ then show that the underlying undirected graph for $\mathcal{G}$ is same as $\mathcal{H}$.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Figure 1: Source: Probabilistic Graphical Models. Daphne Koller and Nir Friedman (2009)

13. Consider the star network as discussed in the class. Recall that the network consists of a hub node X_0 which is connected to rest of the nodes in the network i.e. $X_1, \ldots X_n$. There are no other edges in the network. Assume that all the X_i's are binary. Consider the Markov network represented by the star network with potentials defined over singletons and edges in the graph. Let the singleton potentials be uniform i.e. $\phi^s(X_i = 0) = \phi^s(X_i = 1) = 1, \forall i$. Let the edge potentials be given as $\phi^e_i(X_0, X_i) = 3$ if $X_0 = X_i$ and 1 otherwise, $\forall i \in \{1, \ldots n\}$
    - Construct a clique tree for a star network with $n = 4$, i.e., the variables in the network are $X_0, X_1, \ldots X_4$. You should choose a clique tree which minimizes the size of the largest clique node in the tree.
    - Perform message passing over the clique tree as constructed above to calculate $P(X_0 = 1)$.
14. Show that the complexity of (marginal) inference, i.e., the complexity of calculating the marginal probability for each of the hidden nodes, in a Hidden Markov Model is linear in the number of hidden nodes. You can assume that all the hidden nodes as well as the observed nodes are binary valued.
15. Show that the clique tree generated by VE computation satisfies the running intersection property.
16. Recall the asynchronous two way message passing algorithm to compute the bi-directional messages in a clique tree. Show that the messages computed by this algorithm are equivalent to the ones computed by the following procedure:
    - Designate (any) node in the clique tree as root C_r.
    - Calculate the upward messages going towards the root (as described in class).
    - Pass the appropriate messages downward from the root all the way to the leaves.

    For the last point above, derive the exact expression for a downward message $\delta_{j \to i}$, where $C_j - C_i$ is an edge and C_j is the immediate upward neighbor of C_i. You should clearly explain the dependence of $\delta_{j \to i}$ on the downward message coming down from the side of the root to C_j, i.e., $\delta_{k \to j}$ where C_k is the immediate upward neighbor of C_j.

```
                        /
                       /
                      /
                    C_k         .
                   / | \       /|\
                  /  |  \       |  root
  .                 C_j         |
  | downward      /  | \
```

```
  | messages      /   |  \
 \|/                 C_i
                    /  | \
                   /   |  \
```

17. Consider the student network as given in Figure 2. Use variable elimination (with elimination ordering L, I) to calculate the prbability of $P(S = 0|G = 2, D = 1)$. Simplify your expression as much as you can (use of calculators is not allowed).

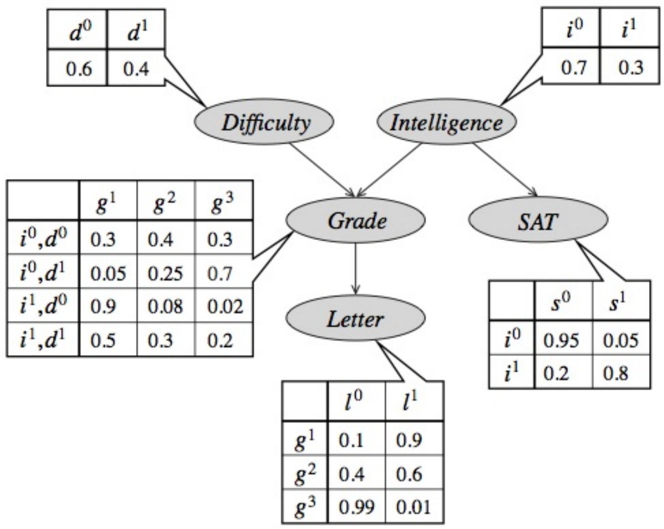


Figure 2: Source: Probabilistic Graphical Models. Daphne Koller and Nir Friedman

18. Consider the Bayesian network shown in Figure 3. Write the expression for the joint distribution specified by this network. Suppose you have a training set composed of the examples given in Table 1, with "?" indicating a missing value. Show the first iteration of the EM algorithm (initial parameters, E-step, M-step), assuming the parameters are initliazed ignoring missing values.

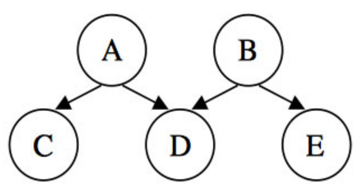


Figure 3: Source: Course Taught by Pedro Domingos.

A	B	C	D	E
0	0	1	1	1
0	0	0	?	1
1	1	0	0	?
1	1	0	0	0
0	1	?	0	1
0	1	1	1	1
1	0	1	0	0
?	1	1	0	1

Table 1: Data with Missing Values