ECE 285 - Deep Generative Models - Assignment 1

Saqib Azim (A59010162) | sazim@ucsd.edu

January 23, 2022

1 Question 1 (25 Points):

For the following Bayesian network, judge whether the following statements are true or false. And give a brief explanation for each of your answer. (I: Intelligent; H: Hardworking; T: good Test taker; U: Understands material; E: high Exam score)

Solution:

Active Trail: In a directional trail $X \rightleftharpoons Z \rightleftharpoons Y$, if influence can flow from X to Y via Z, then the trail is called an active trail, otherwise an inactive trail.

General case: In a long trail $X_1 \rightleftharpoons X_2 ... \rightleftharpoons X_n$, for influence to flow from X_1 to X_n, it needs to flow through every single node on the trail. That is, X_1 can influence X_n if every 2-edge trail $X_{i-1} \rightleftharpoons X_i \rightleftharpoons X_{i+1}$ along the trail is ACTIVE and therefore allows the influence to flow. In order for $X_1 \perp X_n$ (X_1 cannot influence X_n), there should be NO path or trail from X_1 to X_n that has all active trails.

1. T and U are independent - FALSE

- Trail $T \rightarrow E \leftarrow U$ is active if E or one of its descendants is observed/given $\implies$ Inactive
- Trail $T \leftarrow I \rightarrow U$ is active if I is not observed $\implies$ Active trail $\implies$ T can influence U $\implies$ T and U are not independent

2. T and U are conditionally independent given I, E, and H - FALSE

- Trail $T \leftarrow I \rightarrow U$ is active if I is not observed but I is observed $\implies$ Inactive trail
- Trail $T \rightarrow E \leftarrow U$ is active if E or one of its descendants is observed/given. Since E is observed $\implies$ Active trail $\implies$ T and U can influence each other $\implies$ T and U are not conditionally independent given $\{I, E, H\}$

3. T and U are conditionally independent given I and H - TRUE

- Trail $T \leftarrow I \rightarrow U$ is active if I is not observed but I is observed $\implies$ Inactive trail
- Trail $T \rightarrow E \leftarrow U$ is active if E or one of its descendants is observed/given. Since E is not observed $\implies$ Inactive trail

- No active path/trail exist from $T \rightleftharpoons U \implies$ T and U cannot influence each other $\implies$ T is conditionally independent of U given $\{I, H\}$

4. E and H are conditionally independent given U - FALSE

- Trail $E \leftarrow U \leftarrow H$ is active if U is not observed but U is observed $\implies$ Inactive trail
- Trail $E \leftarrow T \leftarrow I$ is active if T is not observed. Since T is not given $\implies$ Active trail
- Trail $T \leftarrow I \rightarrow U$ is active if I is not observed. Since I is not observed $\implies$ Active trail
- Trail $I \rightarrow U \leftarrow H$ is active if U or one of its descendants is observed. Since U is observed $\implies$ Active trail
- There exists an active trail from $E \leftarrow T \leftarrow I \rightarrow U \leftarrow H \implies$ E and H are not conditionally independent given U

5. E and H are conditionally independent given U, I and T - TRUE

- Trail $E \leftarrow U \leftarrow H$ is active if U is not observed but U is observed $\implies$ Inactive trail
- Trail $E \leftarrow T \leftarrow I$ is active if T is not observed. But T is observed $\implies$ Inactive trail
- Trail $T \leftarrow I \rightarrow U$ is active if I is not observed. But I is observed $\implies$ Inactive trail
- Trail $I \rightarrow U \leftarrow H$ is active if U or one of its descendants is observed. Since U is observed $\implies$ Active trail
- There exists no active trail/path from $E \rightleftharpoons H \implies$ E and H are conditionally independent given U, I and T

6. I and H are conditionally independent given E - FALSE

- Trail $I \rightarrow U \leftarrow H$ is active if U or one of its descendants is observed. Since E (one of descendant of U) is observed $\implies$ Active trail
- There exists an active trail/path from $I \rightarrow U \leftarrow H \implies$ I and H are not conditionally independent given E

7. I and H are conditionally independent given T - TRUE

- Trail $I \rightarrow U \leftarrow H$ is active if U or one of its descendants is observed. Since U is not observed $\implies$ Inactive trail
- Trail $I \rightarrow T \rightarrow E$ is active if T is not observed. But T is observed $\implies$ Inactive trail
- Trail $T \rightarrow E \leftarrow U$ is active if E or one of its descendants is observed. Since E is not observed $\implies$ Inactive trail
- Trail $E \leftarrow U \leftarrow H$ is active if U is not observed. Since U is not observed $\implies$ Active trail
- There exists no active trail/path from $I \rightleftharpoons H \implies$ I and H are conditionally independent given T

8. T and H are independent - TRUE

- Trail $T \rightarrow E \leftarrow U$ is active if E or one of its descendants is observed. Since E is not observed $\implies$ Inactive trail
- Trail $E \leftarrow U \leftarrow H$ is active if U is not observed. Since U is not observed $\implies$ Active trail
- Trail $T \leftarrow I \rightarrow U$ is active if I is not observed. Since I is not observed $\implies$ Active trail
- Trail $I \rightarrow U \leftarrow H$ is active if U or one of its descendants is observed. Since U is not observed $\implies$ Inactive trail
- There exists no active trail from $T \rightleftharpoons H \implies$ T and H are independent

9. T and H are conditionally independent given E - FALSE

- Trail $T \rightarrow E \leftarrow U$ is active if E or one of its descendants is observed. Since E is observed $\implies$ Active trail
- Trail $E \leftarrow U \leftarrow H$ is active if U is not observed. Since U is not observed $\implies$ Active trail
- There exists an active trail from $T \rightarrow E \leftarrow U \leftarrow H \implies$ T and H are not conditionally independent given E

10. T and H are conditionally independent given E and U - FALSE

- Trail $T \rightarrow E \leftarrow U$ is active if E or one of its descendants is observed. Since E is observed $\implies$ Active trail
- Trail $E \leftarrow U \leftarrow H$ is active if U is not observed. Since U is observed $\implies$ Inactive trail
- Trail $T \leftarrow I \rightarrow U$ is active if I is not observed. Since I is not observed $\implies$ Active trail
- Trail $I \rightarrow U \leftarrow H$ is active if U or one of its descendants is observed. Since U is observed $\implies$ Active trail
- There exists an active trail from $T \leftarrow I \rightarrow U \leftarrow H \implies$ T and H are not conditionally independent given E and U

2 Question 2 (15 Points):

For the above Bayesian network, construct local conditional probability tables. Assume all variables are binary (1 for true and 0 for false). For example, P(E=1 | T=1, U=1) = 0.8. And give a brief explanation for your specified probabilities. For example, if a student is a good test taker and understands the material well, the student is very likely to have a high exam score.

Solution:
In a general population, hard-work (H) and intelligence (I) are less-observed characteristics. These properties can be specified independently as they have no parents in the Bayesian Network (BN). Thus, I have given probability 0.25 that a student is intelligent while 0.3 is the probability that a student is hardworking.

H=0	H=1
0.7	0.3

I=0	I=1
0.75	0.25

Based on the BN graph structure, if we know whether a student is intelligent or not, then the belief about whether that student is a good test taker or not does not depend on other parameters. Hence, we can specify a local probability model conditioned on intelligence variable. Intuitively, if a student is intelligent, then there is a strong chance (80%) that the student would also be a good test taker. On the contrary, if the student is not intelligent, then it decreases the belief that the student is a good test taker and this probability dips to 10%.

	T=0	T=1
I=0	0.9	0.1
I=1	0.2	0.8

Similarly, if the information about whether a student is hardworking (H) and intelligent (I) is given, then the belief about student's understanding of the material (U) is independent of exam scores (E) and whether he/she is a good test taker (T). Intuitively, if a student is both hardworking and intelligent, then most certainly the student understands the material (90%). On the other hand, if a student is neither intelligent nor hardworking, then there is a high chance (95%) that the student does not understands the material. If the student is either hardworking or intelligent, there is still a reasonable chance (70% and 60% resp.) that the student understands the material.

	U=0	U=1
I=0, H=0	0.95	0.05
I=0, H=1	0.3	0.7
I=1, H=0	0.4	0.6
I=1, H=1	0.1	0.9

If the information regarding whether the student is a good test taker and whether he/she understands the material is known, then the belief regarding student's high exam score is independent of student's intelligence and hardworking capabilities. Considering that, if the student understands the material as well as a good test taker, then it is highly likely (80%) that he will receive high exam score. But if he is neither a good test taker nor understands the material, then most probably (90%) he will receive a low exam score. If student understands material well but not a good test taker, then also there exists a reasonable chance (60%) that he will get high score, whereas if he does not understand material, then despite being a good test taker, there is a high belief (70%) that he will receive a low exam score.

	E=0	E=1
T=0, U=0	0.9	0.1
T=0, U=1	0.4	0.6
T=1, U=0	0.7	0.3
T=1, U=1	0.2	0.8

3 Question 3 (10 Points):

For the above Bayesian network, write down the joint distribution of all variables.

Solution:
Let G be a bayesian network graph over $X_1, X_2, ..., X_n$. A distribution P over the same space factorizes according to G if

$P(X_1, ..., X_n) = \prod_{i=1}^{N} P(X_i \mid \text{Pa}_{X_i}^{G})$ where $\text{Pa}_{X_i}^{G}$ represents parents of X_i in G

P(I, H, T, U, E) = P(I) * P(H) * P(T | I) * P(U | I, H) * P(E | T, U)

Below table shows the joint probability distribution for few entries (out of possible 32 entries)

I	H	T	U	E	P(I, H, T, U, E)
0	0	0	0	0	0.4039875
0	0	0	0	1	0.0448875
0	0	0	1	0	0.00945
0	0	0	1	1	0.014175

4 Question 4 (15 Points):

Calculate P(E=1 | Hardworking=1)

Solution:
P(E = 1 | H = 1)
= P(E = 1 | U = 1, T = 1, H = 1) * P(U = 1, T = 1 | H = 1) +
P(E = 1 | U = 1, T = 0, H = 1) * P(U = 1, T = 0 | H = 1) +
P(E = 1 | U = 0, T = 1, H = 1) * P(U = 0, T = 1 | H = 1) +
P(E = 1 | U = 0, T = 0, H = 1) * P(U = 0, T = 0 | H = 1)

Since, if the variables U and T are observed, then there exists no active trail/path from E $\rightleftharpoons$ H $\implies$ E is conditionally independent of H given {U, T}

Therefore, P(E = e | U = u, T = t, H = 1) = P(E = e | U = u, T = t)

P(E = 1 | H = 1)
= P(E = 1 | U = 1, T = 1) * P(U = 1, T = 1 | H = 1) +
P(E = 1 | U = 1, T = 0) * P(U = 1, T = 0 | H = 1) +
P(E = 1 | U = 0, T = 1) * P(U = 0, T = 1 | H = 1) +
P(E = 1 | U = 0, T = 0) * P(U = 0, T = 0 | H = 1)

P(U = u, T = t | H = 1) (where u and t can take values 0, 1)
= P(U = u, T = t | I = 1, H = 1) * P(I = 1 | H = 1) +

P(U = u, T = t | I = 0, H = 1) * P(I = 0 | H = 1)

Since U and T are conditionally independent given I and H (shown in Q1.3). In addition, I and H are independent (inferred from BN structure) $\implies$ P(I = i | H = 1) = P(I = i)

P(U = u, T = t | H = 1)
= P(U = u | I = 1, H = 1) * P(T = t | I = 1, H = 1) * P(I = 1) +
P(U = u | I = 0, H = 1) * P(T = t | I = 0, H = 1) * P(I = 0)

Now, T is conditionally independent of H given I $\implies$ P(T = t | I = i, H = 1) = P(T = t | I = i)

P(U = u, T = t | H = 1)
= P(U = u | I = 1, H = 1) * P(T = t | I = 1) * P(I = 1) +
P(U = u | I = 0, H = 1) * P(T = t | I = 0) * P(I = 0)

P(U = 1, T = 1 | H = 1)
= P(U = 1 | I = 1, H = 1) * P(T = 1 | I = 1) * P(I = 1) +
P(U = 1 | I = 0, H = 1) * P(T = 1 | I = 0) * P(I = 0)
= 0.9 * 0.8 * 0.25 + 0.7 * 0.1 * 0.75
= 0.2325

P(U = 1, T = 0 | H = 1)
= P(U = 1 | I = 1, H = 1) * P(T = 0 | I = 1) * P(I = 1) +
P(U = 1 | I = 0, H = 1) * P(T = 0 | I = 0) * P(I = 0)
= 0.9 * 0.2 * 0.25 + 0.7 * 0.9 * 0.75
= 0.5175

P(U = 0, T = 1 | H = 1)
= P(U = 0 | I = 1, H = 1) * P(T = 1 | I = 1) * P(I = 1) +
P(U = 0 | I = 0, H = 1) * P(T = 1 | I = 0) * P(I = 0)
= 0.1 * 0.8 * 0.25 + 0.3 * 0.1 * 0.75
= 0.0425

P(U = 0, T = 0 | H = 1)
= P(U = 0 | I = 1, H = 1) * P(T = 0 | I = 1) * P(I = 1) +
P(U = 0 | I = 0, H = 1) * P(T = 0 | I = 0) * P(I = 0)
= 0.1 * 0.2 * 0.25 + 0.3 * 0.9 * 0.75
= 0.2075

Finally computing the desired probability

P(E = 1 | H = 1)
= P(E = 1 | U = 1, T = 1) * P(U = 1, T = 1 | H = 1) +

P(E = 1 | U = 1, T = 0) * P(U = 1, T = 0 | H = 1) +
P(E = 1 | U = 0, T = 1) * P(U = 0, T = 1 | H = 1) +
P(E = 1 | U = 0, T = 0) * P(U = 0, T = 0 | H = 1)

= 0.8 * 0.2325 + 0.6 * 0.5175 + 0.3 * 0.0425 + 0.1 * 0.2075
= 0.53

P(E = 1 | H = 1) = 0.53

5 Question 5 (35 points):

Consider a task: based on the infected cases in the past week, predict the number of infected COVID-19 cases for all cities in San Diego County for tomorrow. Design a Markov random field model to perform this task. In the model, capture the correlation between cities: nearby cities have similar number of infected cases

Solution:
Let there be 6 cities in the San Diego county denoted by $C_1, C_2, ..., C_6$. Let the distance between city C_i and C_j be denoted by D_{ij}.

$D_{ii} = 0$ and $D_{ij} = D_{ji}$ for all i,j = 1, ..., 6

Let X_i denote the number of infected COVID-19 cases in city C_i, where i = 1,...,6. In

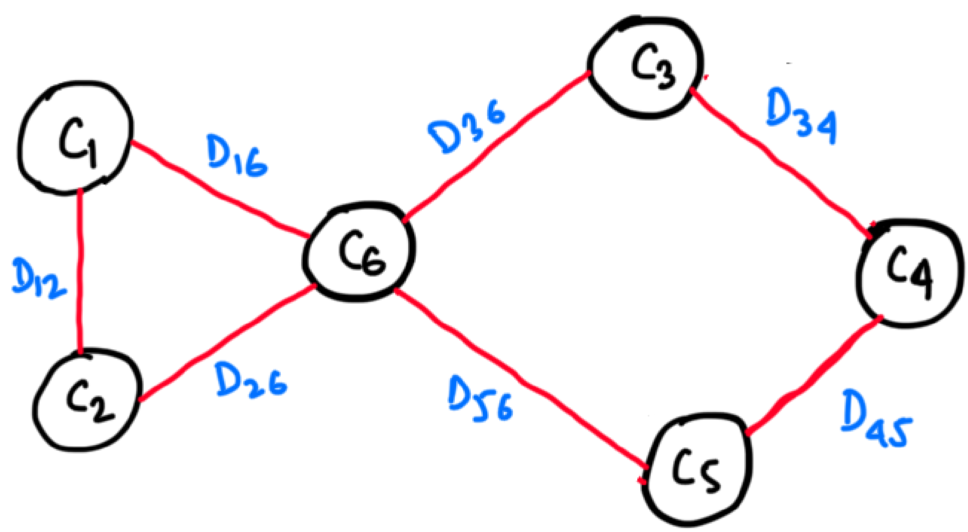

Figure 1: Markov Random Field Dummy Model for the San Diego county. Here C_i represents i^{th} city in the county

order to make future number of cases predictions, we can classify the number of infected cases in a city into three categories - very low (0), medium but significant (1), very high (2). Let $Y_1, Y_2, ..., Y_6$ denote the category class in corresponding city where each $Y_i = \{0, 1, 2\}$ depending on whether the number of infected cases are low, medium or high.

We define city C_i and C_j to be neighbors if the direct line-joining distance between their centers D_{ij} is less than some threshold D_0 (say 50 miles). The neighboring cities are expected to have similar number of infected cases and therefore, there is high probability that cities connected in the MRF graph belong to same infected class category {0,1,2}.

Based on the MRF, there are 5 maximal cliques in the undirected graph - {1,2,6}, {3,6}, {3,4}, {4,5}, {5,6}. The joint probability distribution (also called Gibbs distribution) can be written as -

$P(Y_1, Y_2, ..., Y_6) = \frac{1}{Z} * \phi_{126}(Y_1, Y_2, Y_6) * \phi_{36}(Y_3, Y_6) * \phi_{34}(Y_3, Y_4) * \phi_{45}(Y_4, Y_5) * \phi_{56}(Y_5, Y_6)$ where ϕ is the potential function and Z denotes the partition function used for normalizing the joint distribution.

$Z = \sum_{Y_1,...,Y_6} \phi_{126}(Y_1, Y_2, Y_6) * \phi_{36}(Y_3, Y_6) * \phi_{34}(Y_3, Y_4) * \phi_{45}(Y_4, Y_5) * \phi_{56}(Y_5, Y_6)$

The potential function $\phi(.) \in \mathbb{R}^+ \cup \{0\}$ captures the correlation between the number of infected cases among neighboring cities. The below tables 1 and 2 shows example of potential values for two of the maximal cliques - {1,2,6} and {3,6} - in the graph. The remaining clique potentials are not shown for brevity. Based on the clique potentials, the joint probability can be estimated for all possible configuration of Y_i which will help us to predict approximate infected number of cases in each city.

Y_1	Y_2	Y_6	$\phi_{126}(Y_1, Y_2, Y_6)$
0	0	0	10
0	0	1	100
0	0	2	30
0	1	0	20
0	1	1	40
0	1	2	60
0	2	0	15
0	2	1	5
0	2	2	20
1	0	0	20
1	0	1	40
1	0	2	60
1	1	0	30
1	1	1	150
1	1	2	200
1	2	0	4
1	2	1	60
1	2	2	120

Table 1: Potential values for various configurations of clique {1,2,6}. Neighboring cities are expected to have similar number of infected cases and therefore, highly likely that they belong to same class category (low-0, medium-1, high-2). NOTE: The table is not filled entirely to keep it short.

Y_3	Y_6	$\phi_{36}(Y_3, Y_6)$
0	0	100
0	1	70
0	2	20
1	0	60
1	1	150
1	2	50
2	0	30
2	1	90
2	2	150

Table 2: Potential values for different configuration values of clique {3,6}