Deep Generative Models: Homework #1

Due on Octotber 30, 2023 at 11:59pm

Causality and Probabilistic Graphical Models

Problem 1 : (PGM) 15 points

For $i \in \{1, 2, 3\}$, X_i is a Bernoulli random variable representing the outcome of a coin toss where heads occur with probability q. If X_i's are independent and $Z_4 = X_1 \oplus X_2$ and $Z_5 = X_2 \oplus X_3$:

1. Determine the conditional distributions of (X_2, X_3) given $Z_5 = 0$ and $Z_5 = 1$, where $\oplus$ denotes the exclusive or operation.
2. Draw the directed probabilistic graphical model and find the conditional probability tables for the 5 random variables. Determine the independence relations in this graph.
3. Draw the undirected probabilistic graphical model and find the potential functions defining the model for the 5 random variables. Determine the independence relations in this graph.
4. Find conditions on q that lead to the following independencies: $Z_5 \perp X_3$ and $Z_4 \perp X_1$. Do these marginal independencies follow from the two previously introduced graphical models?

Problem 2 : (PGM) 25 points

This question will refer to the graphical models shown in Figures 1 and 2, which encode a set of independencies among the following variables: Season (S), Flu (F), Dehydration (D), Chills (C), Headache (H), Nausea (N), Dizziness (Z). Note that the two models have the same skeleton, but Figure 1 depicts a directed model (Bayesian network) whereas Figure 2 depicts an undirected model(Markov network).

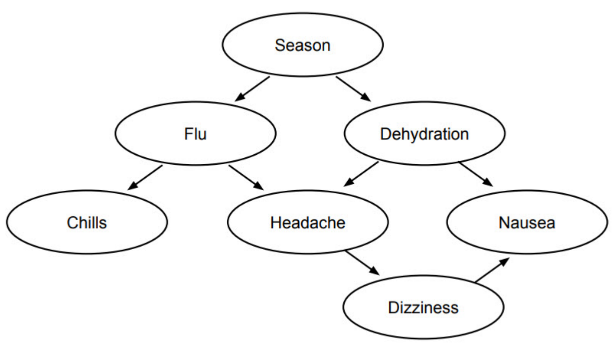

Figure 1: A Bayesian network that represents a joint distribution over the variables Season, Flu, Dehydration, Chills, Headache, Nausea, and Dizziness.

Consider the model shown in Figure 1. Indicate whether the following independence statements are true or false according to this model. Provide a very brief justification of your answer (no more than 1 sentence).

1. $Season \perp Chills$
2. $Season \perp Chills|Flu$
3. $Season \perp Headache|Flu$
4. $Season \perp Headache|Flu, Dehydration$
5. $Season \perp Nausea|Dehydration$
6. $Season \perp Nausea|Dehydration, Headache$

7. $Flu \perp Dehydration$
8. $Flu \perp Dehydration | Season, Headache$
9. $Flu \perp Dehydration | Season$
10. $Flu \perp Dehydration | Season, Nausea$
11. $Chills \perp Nausea$
12. $Chills \perp Nausea | Headache$

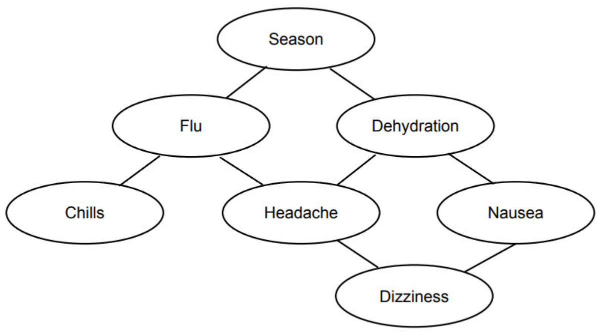

Figure 2: A Markov network that represents a joint distribution over the variables Season, Flu, Dehydration, Chills, Headache, Nausea, and Dizziness

1. Using the directed model shown in Figure 1, write down the factorized form of the joint distribution over all of the variables, P(S, F, D, C, H, N, Z).
2. (optional) Using the undirected model shown in Figure 2, write down the factorized form of the joint distribution over all of the variables, assuming the model is parameterized by one factor over each node and one over each edge in the graph.

Problem 2 continued on next page. . .

$P(S = \text{winter})$	$P(S = \text{summer})$
0.5	0.5

	$P(F = \text{true} \mid S)$	$P(F = \text{false} \mid S)$
$S = \text{winter}$	0.4	0.6
$S = \text{summer}$	0.1	0.9

	$P(D = \text{true} \mid S)$	$P(D = \text{false} \mid S)$
$S = \text{winter}$	0.1	0.9
$S = \text{summer}$	0.3	0.7

	$P(C = \text{true} \mid F)$	$P(C = \text{false} \mid F)$
$F = \text{true}$	0.8	0.2
$F = \text{false}$	0.1	0.9

	$P(H = \text{true} \mid F, D)$	$P(H = \text{false} \mid F, D)$
$F = \text{true}, D = \text{true}$	0.9	0.1
$F = \text{true}, D = \text{false}$	0.8	0.2
$F = \text{false}, D = \text{true}$	0.8	0.2
$F = \text{false}, D = \text{false}$	0.3	0.7

	$P(Z = \text{true} \mid H)$	$P(Z = \text{false} \mid H)$
$H = \text{true}$	0.8	0.2
$H = \text{false}$	0.2	0.8

	$P(N = \text{true} \mid D, Z)$	$P(N = \text{false} \mid D, Z)$
$D = \text{true}, Z = \text{true}$	0.9	0.1
$D = \text{true}, Z = \text{false}$	0.8	0.2
$D = \text{false}, Z = \text{true}$	0.6	0.4
$D = \text{false}, Z = \text{false}$	0.2	0.8

Table 1: : Conditional probability tables for the Bayesian network shown in Figure 1

Problem 2 continued on next page. . .

Assume you are given the conditional probability tables listed in Table 1 for the model shown in Figure 1. Evaluate each of the probabilities queries listed below, and show your calculations.

1. What is the probability that you have the flu, when no prior information is known?
2. What is the probability that you have the flu, given that it is winter?
3. What is the probability that you have the flu, given that it is winter and that you have a headache?
4. What is the probability that you have the flu, given that it is winter, you have a headache, and you know that you are dehydrated?
5. Does knowing you are dehydrated increase or decrease your likelihood of having the flu? Intuitively, does this make sense?

Problem 3 : (Variational Inference) 35 points

1. Assume G is a Bayesian network with observed nodes of $O = \{O_1, \ldots, O_N\}$ and hidden nodes of $H = \{H_1, \ldots, H_M\}$. Show that in mean-field variational inference if we assume the variational distribution factors completely over $H_1, \ldots, H_M$ nodes, then:

$$\ln q_m^*(H_m) = E_{H_{-m}}[\ln p(H_m | H_{-m}, O_{1:N})] + \text{const}$$

   in which H_{-m} is the set of all of the H_s but H_m.

2. (optional) Assume for $1 \leq m \leq M$, $p(H_m | H_{-m}, O_{1:N})$ follows an exponential distribution with the natural parameter of $\eta_m(H_{-m}, O_{1:N})$. Show that:
   (a) $p(H_m | H_{-m}, O_{1:N})$ and $q_m^*(H_m)$ both belong to the same family of distributions.
   (b) The natural parameter of the distribution of $q_m^*(H_m)$ is $E_{H_{-m}}[\ln p(H_m | H_{-m}, O_{1:N})]$.

3. In this question, the goal is to learn the parameters of a Gaussian Mixture Model using the Bayesian approach. The random variables' distributions are as follows:

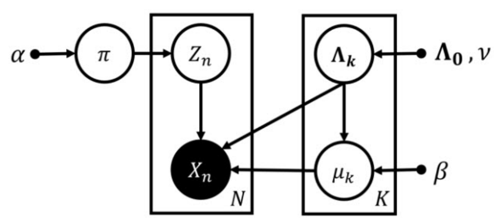

Figure 3: Bayesian Network related to distributions at (1)

$$\begin{aligned}
p(\pi) &= \text{Dirichlet}(\alpha) \\
p(\Lambda_k) &= \text{Wishart}(\Lambda_0, \nu) \\
p(\mu_k | \Lambda_k) &= \text{Normal}(0, (\beta\Lambda_k)^{-1}) \\
p(Z_n | \pi) &= \text{Categorical}(\pi) \\
p(X_n | Z_{nk} = 1, \Lambda_k, \mu_k) &= \text{Normal}(\mu_k, \Lambda_k^{-1})
\end{aligned} \tag{1}$$

In the above equations, the Z_n vector has a one-of-K coding style (one-hot vector of length K). Additionally, all of the complete conditional distributions belong to the exponential family. Therefore, the results of (2) may be useful for the mean-field variational inference.

(a) Show that the complete conditional distributions are as follows:

$$p(\pi|\text{the other random variables}) = \text{Dirichlet}(\alpha_1 + \sum_{n=1}^{N} Z_{n1}, \ldots, \alpha_K + \sum_{n=1}^{N} Z_{nK}) \tag{2}$$

$$\begin{aligned} p(\Lambda|\text{the other random variables}) = \text{Wishart}((&\Lambda_0^{-1} + \beta\mu_k\mu_k^T \\ &+ \sum_{n=1}^{N}(X_n - \mu_k)(X_n - \mu_k)^T Z_{nk})^{-1}, \nu + 1 + \sum_{n=1}^{N} Z_{nk}) \end{aligned} \tag{3}$$

$$p(\mu_k|\text{the other random variables}) \propto \exp\left[-\frac{1}{2}\mu_k^T(\beta\Lambda_k + \Lambda_k\sum_{n=1}^{N} Z_{nk})\mu_k + \mu_k^T(\Lambda_k\sum_{n=1}^{N} X_n Z_{nk})\right] \tag{4}$$

$$p(Z_{nk} = 1|\text{the other random variables}) = \text{Categorical}\left(\frac{\xi_1}{\sum_{k=1}^{K}\xi_k}, \ldots, \frac{\xi_K}{\sum_{k=1}^{K}\xi_k}\right) \tag{5}$$

$$\xi_k = \pi_k \times \text{Normal}(X_n|\mu_k, \Lambda_k^{-1})$$

(b) Using the equations of the previous part, calculate the parameters of the complete conditional distributions and the relations required for the mean-field method. You may use the existing equations for $E[\log(x)]$ when $X \sim \text{Dirichlet}(\gamma)$ without proving them.

Problem 4 : (Causal Inference) 15 points

Suppose the following data are acquired by some health researchers in order to examine the effect of two different medicines on the recovery of patients with fatty liver grade 1 and grade 4.

	medicine A	*medicine B*
Fatty liver grade one	81/87	234/270
Fatty liver grade four	192/263	55/80

for each number in the tables, the numerator and denominator state the number of healed patients and the total number of patients respectively with the corresponding grade of disease and the medicine. Then You are asked to use your knowledge from causal inference and statistics to decide which medicine should be preferred. choose the one that you think is better and explain why.

Problem 5 : (Causal Discovery) 20 points

Consider you have seen some data from either the SCM M1 :

$$X1 = \epsilon_1$$
$$X2 = \beta_{12}X1 + \epsilon_2$$

where $\epsilon_1 \perp\!\!\!\perp \epsilon_2$, or the SCM M2:

$$X2 = \epsilon_2$$
$$X1 = \beta_{21}X2 + \epsilon_1$$

where $\epsilon_1 \perp\!\!\!\perp \epsilon_2$. Samples of $(X1, X2)$ are in the file `ngaussian.csv`, where the first column contains samples of $X1$, and the second column contains samples of $X2$.

(a) Let $\hat{\beta}_{12}$ be the linear regression coefficient when regressing X_2 onto X_1 (do not consider intercept term). Report $\hat{\beta}_{12}$ and plot $(X_1, X_2 - \hat{\beta}_{12}X_1)$.

(b) Let $\hat{\beta}_{21}$ be the linear regression coefficient when regressing X_1 onto X_2 (do not consider intercept term). Report $\hat{\beta}_{21}$ and plot $(X_2, X_1 - \hat{\beta}_{21}X_2)$.

(c) Which one of these models is the model that generated the dataset? explain.

(hint: in this question, you are supposed to use non-gaussianity for causal discovery. you can know more about the problem on the paper "A Linear Non-Gaussian Acyclic Model for Causal Discovery")

Please note it is necessary for each student to answer the questions independently. While cooperation is encouraged, students with exactly the same answers will not receive any credit.

If you have any questions, please send them to **mhdhshri@gmail.com** .

Good luck.