Probabilistic Graphical Models

Part A HWK#1
Assigned Tuesday, Jan. 14, 2020
Due (both Parts A and B) : Thursday, Feb. 6, 2020

1 Fundamentals

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).

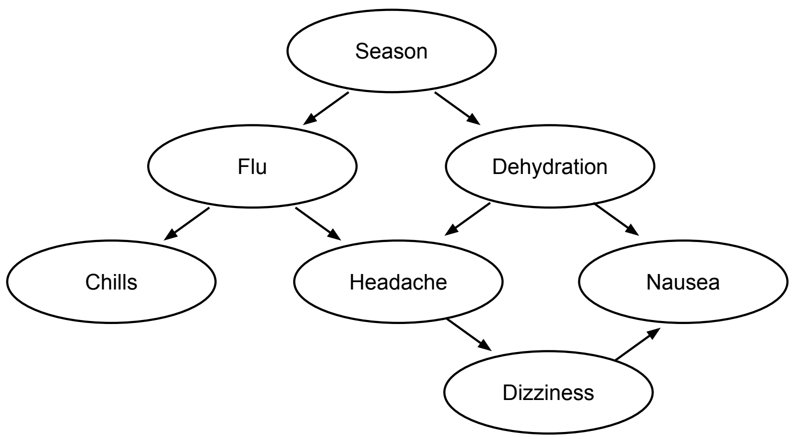

Figure 1: A Bayesian network that represents a joint distribution over the variables Season, Flu, Dehydration, Chills, Headache, Nausea, and Dizziness.

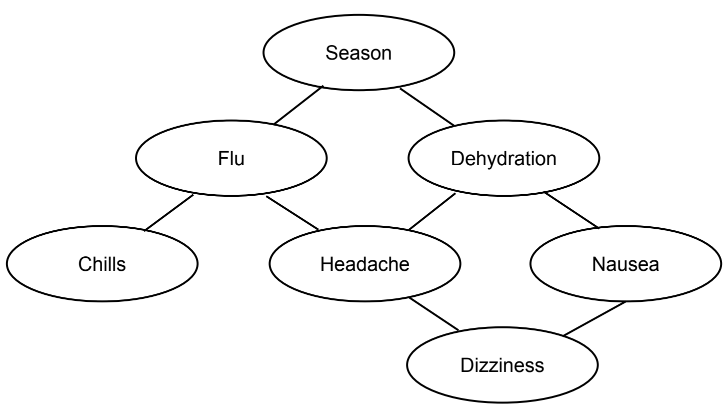

Figure 2: A Markov network that represents a joint distribution over the variables Season, Flu, Dehydration, Chills, Headache, Nausea, and Dizziness.

$P(S = \text{winter})$	$P(S = \text{summer})$
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 1.1

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

Problem 1.2

1. Using the <u>directed model</u> 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. Using the <u>undirected model</u> 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 1.3

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 1.4

Now consider the undirected model shown in Figure 2.

1. Are there any differences between the set of marginal independencies encoded by the directed and undirected versions of this model? If not, state the full set of marginal independencies encoded by both models. If so, give one example of a difference.
2. Are there any differences between the set of conditional independencies encoded by the directed and undirected versions of this model? If so, give one example of a difference.

2 Bayesian Networks

Problem 2.1

In this problem you will construct your own Bayesian network (BN) for a few different modeling scenarios described as word problems. By standard convention, we will use shaded circles to represent observed quantities, clear circles to represent random variables, and uncircled symbols to represent distribution parameters.

In order to do this problem, you will first need to understand plate notation, which is a useful tool for drawing large BNs with many variables. Plates can be used to denote repeated sets of random variables. For example, suppose we have the following generative process:

- Draw $Y \sim \text{Normal}(\mu, \Sigma)$
- For $m = 1, \dots, M$:

  Draw $X_m \sim \text{Normal}(Y, \Sigma)$

This BN contains $M + 1$ random variables, which includes M repeated variables $X_1, \dots, X_M$ that all have Y as a parent. In the BN, we draw the repeated variables by placing a box around a single node, with an index in the box describing the number of copies; we've drawn this in Figure 3.

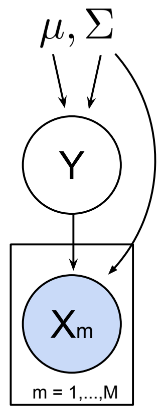

Figure 3: An example of a Bayesian network drawn with plate notation.

For each of the modeling scenarios described below, draw a corresponding BN. Make sure to label your nodes using the variable names given below, and use plate notation if necessary.

1. (Gaussian Mixture Model). Suppose you want to model a set of clusters within a population of N entities, $X_1, \dots, X_N$. We assume there are K clusters $\theta_1, \dots, \theta_K$, and that each cluster represents a vector and a matrix, $\theta_k = \{\mu_k, \Sigma_k\}$. We also assume that each entity X_n "belongs" to one cluster, and its membership is given by an assignment variable $Z_n \in \{1, \dots, K\}$.

   Here's how the variables in the model relate. Each entity X_n is drawn from a so-called "mixture distribution," which in this case is a Gaussian distribution, based on its individual cluster assignment and the entire set of clusters, written $X_n \sim \text{Normal}(\mu_{Z_n}, \Sigma_{Z_n})$. Each cluster assignment Z_n has a prior, given by $Z_n \sim \text{Categorical}(\beta)$. Finally, each cluster θ_k also has a prior, given by $\theta_k \sim \text{Normal-invWishart}(\mu_0, \lambda, \Phi, \nu) = \text{Normal}(\mu_0, \frac{1}{\lambda}\Sigma) \cdot \text{invWishart}(\Phi, \nu)$.

2. (Bayesian Logistic Regression). Suppose you want to model the underlying relationship between a set of N input vectors $X_1, \dots, X_N$ and a corresponding set of N binary outcomes $Y_1, \dots, Y_N$. We assume there is a single vector β which dictates the relationship between each input vector and its associated output variable.

   In this model, each output is drawn with $Y_n \sim \text{Bernoulli}(\text{invLogit}(X_n\beta))$. Additionally, the vector β has a prior, given by $\beta \sim \text{Normal}(\mu, \Sigma)$.

Problem 2.2

Prove the following theorem.

Theorem 1. *Let $\mathcal{G}$ be a Bayesian Network structure over a set of random variables $\mathcal{X}$ and let P be a joint distribution over $\mathcal{X}$. If P factorizes according to $\mathcal{G}$, then $\mathcal{G}$ is an I-map for P.*

Problem 2.3

Prove the following theorem:

Let $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ be three disjoint sets of nodes in a Bayesian network $\mathcal{G}$. Let $\boldsymbol{U} = \boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z}$, and let $\mathcal{G}' = \mathcal{G}^{+}[\boldsymbol{U}]$ be the induced Bayesian network over $\boldsymbol{U} \cup \mathrm{Ancestors}_{\boldsymbol{U}}$. Let $\mathcal{H}$ be the moralized graph $\mathcal{M}[\mathcal{G}']$. Then $\text{d-sep}_{\mathcal{G}}(\boldsymbol{X}; \boldsymbol{Y} \mid \boldsymbol{Z})$ if and only if $\mathrm{sep}_{\mathcal{H}}(\boldsymbol{X}; \boldsymbol{Y} \mid \boldsymbol{Z})$.

Following two Problems are from the textbook K&F.

Problem 2.4

Exercise 3.11 (part a only) from the book, at page 98.

Problem 2.5

Exercise 4.10 from the book, at page 154.