Machine Learning, 2024 Fall
Homework 2

Notice

Due 23:59 (CST), Nov 19, 2024
Plagiarizer will get 0 points.
LaTeXis highly recommended. Otherwise you should write as legibly as possible.

A Bayesian Network

(24pt) The figure below shows a Bayesian network, illustrating the conditional dependencies among the following variables: Season (S), Flu (F), Dehydration (D), Chills (C), Headache (H), Nausea (N), and Dizziness (Z).

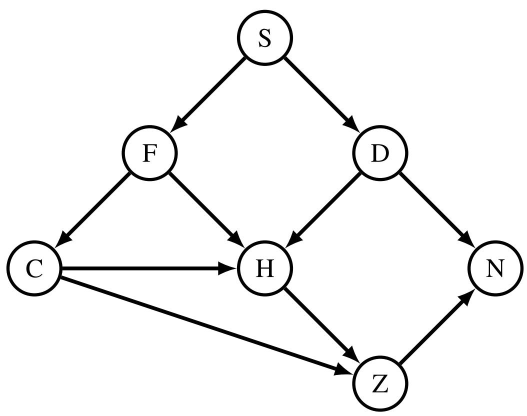

Figure 1: A Bayesian network that represents the conditional dependencies among the variables: Season, Flu, Dehydration, Chills, Headache, Nausea, and Dizziness.

A.1 Independence in Bayesian Network

(12pt) Justify the following independence statements are true or false and give a brief explanation.

- $S \perp N|D$
- $S \perp Z|H$
- $F \perp N|Z, D$
- $F \perp Z|C, H, N$
- $C \perp N|Z$
- $C \perp D|S$

A.2 Evaluating Probability Queries

(12pt) Given the conditional probability tables for the Bayesian network in the following table, calculate each of the queried probabilities. **Note:** You have to write down the calculation steps if necessary.

(1) Use conditional independency properties, write down the factorized form of the conditional independence over all of the variables, $P(S, F, D, C, H, N, Z)$

(2) What is the probability that you have the flu, given that it is summer?

(3) What is the probability that you are feeling chill,when no prior information is known?

(4) What is the probability that you have the flu, given that it is summer and that you have a headache, and you know that you are dehydrated?

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

	$P(F = \text{true}\|S)$	$P(F = \text{false}\|S)$
$S = \text{winter}$	0.4	0.6
$S = \text{summer}$	0.2	0.8

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

	$P(C = \text{true}\|F)$	$P(C = \text{false}\|F)$
$F = \text{true}$	0.7	0.3
$F = \text{false}$	0.8	0.2

	$P(H = \text{true}\|C, F, D)$	$P(H = \text{false}\|C, F, D)$
$C = \text{true}, F = \text{true}, D = \text{true}$	0.7	0.3
$C = \text{true}, F = \text{true}, D = \text{false}$	0.6	0.4
$C = \text{true}, F = \text{false}, D = \text{true}$	0.5	0.5
$C = \text{true}, F = \text{false}, D = \text{false}$	0.4	0.6
$C = \text{false}, F = \text{true}, D = \text{true}$	0.3	0.7
$C = \text{false}, F = \text{true}, D = \text{false}$	0.2	0.8
$C = \text{false}, F = \text{false}, D = \text{true}$	0.1	0.9
$C = \text{false}, F = \text{false}, D = \text{false}$	0.8	0.2

	$P(N = \text{true}\|D, Z)$	$P(N = \text{false}\|D, Z)$
$D = \text{true}, Z = \text{true}$	0.7	0.3
$D = \text{true}, Z = \text{false}$	0.8	0.2
$D = \text{false}, Z = \text{true}$	0.2	0.8
$D = \text{false}, Z = \text{false}$	0.5	0.5

	$P(Z = \text{true}\|C, H)$	$P(Z = \text{false}\|C, H)$
$C = \text{true}, H = \text{true}$	0.1	0.9
$C = \text{true}, H = \text{false}$	0.3	0.7
$C = \text{false}, H = \text{true}$	0.4	0.6
$C = \text{false}, H = \text{false}$	0.8	0.2

B Variable Elimination

(26pt) Given a Bayesian network in the following figure, which consists of binary variables. We will use variable elimination. The chosen variable elimination ordering is A, C, E, G to compute the query $P(B, D, H|f = 1)$.

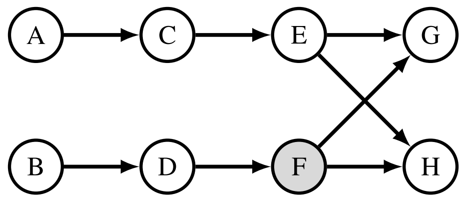

Figure 2: A Bayesian network.

(1) (3pt) What is the corresponding moral graph?
(2) (3pt) Write down all initial factors after inserting evidence $f = 1$.
(3) (15pt) Run variable elimination

1. When eliminating A, please write all factors including the new generated factor f_1.
2. When eliminating C, please write all factors including the new generated factor f_2.
3. When eliminating E, please write all factors including the new generated factor f_3.
4. When eliminating G, please write all factors including the new generated factor f_4.
5. Compute $P(B, D, H|f = 1)$ from the factors left in 4

(4) (5pt) Among the factors, f_1, f_2, f_3, f_4, which is the most largest factor generated, and what is the size of this factor? In this context, we assume that all variables have binary domains, and we determine the factor size by counting the number of rows in the table that represents the factor.

C Learning in Naive Bayesian Networks

Assume a training dataset $\mathcal{D} = \{\mathbf{x}, y\}$, where $\mathbf{x} \in R^d$ represents different features and y represents categories, and let $\mathcal{G}$ be a network over these variables.

(1) If $\mathcal{G}$ is a Naive Bayesian network, consider a apple quality dataset that describes the features of good and bad apples. The training data is as follows:

Color	**Skin Smoothness**	**Size**	**Aroma**	**Flesh Firmness**	**Category**
Red	Smooth	Medium	Strong	Hard	Good Apple
Green	Rough	Large	Faint	Medium	Bad Apple
Red	Smooth	Medium	Faint	Soft	Bad Apple
Red	Rough	Large	Strong	Hard	Good Apple
Green	Smooth	Small	Strong	Hard	Good Apple
Red	Smooth	Medium	Strong	Medium	Good Apple
Yellow	Rough	Medium	Faint	Soft	Bad Apple
Red	Smooth	Large	Strong	Hard	Good Apple
Green	Rough	Small	Faint	Soft	Bad Apple
Red	Smooth	Medium	Strong	Hard	Good Apple
Yellow	Smooth	Medium	Faint	Medium	Bad Apple
Red	Rough	Small	Strong	Soft	Good Apple

(a) (5pts) Please draw this Naive Bayesian network

(b) (5pts) Please write down the CPTs with add-1 Laplace smoothing.

(c) (5pts) Given an apple with the features *Red, Rough, Small, Faint, and Soft*, Is this apple a good apple?

(2)(10pts) For each local node in $\mathcal{G}$, the parameters of CPT can be written as $\boldsymbol{\theta}_i$. If we independently maximize the local likelihood and combine each local optimal solution $\boldsymbol{\theta}_i^*$ into a result denoted $\boldsymbol{\theta}^P = [\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \cdots \boldsymbol{\theta}_n^*]$. Is the result $\boldsymbol{\theta}^P$ equal to the optimal maximum likelihood estimate for the global parameter $\boldsymbol{\theta}^* = \arg\max L(\boldsymbol{\theta})$? If so, please prove it; if not, please explain why.

D Latent Variable Analysis

A latent variable model for T data vectors $\mathbf{x}_0, \cdots, \mathbf{x}_T$ is

$$P(\mathbf{x}_0, \cdots, \mathbf{x}_T) = P(\mathbf{x}_0) \prod_{t=1}^{T} P(\mathbf{x}_t | \mathbf{x}_0)$$

where $P(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\sqrt{\alpha_t}\mathbf{x}_0, (1-\alpha_t)\mathbf{I}), \mathcal{N}(\cdot)$ is a Normal distribution, $\alpha_t > 0$ is a scalar parameter, and $\mathbf{I}$ is a identity matrix.

(1)(5pts) Drawing a graphical model to depict the generative process. (Note that each variable $\mathbf{x}_i$ is influenced by its own parameter α_t)

(2)(10pts) If $P(\mathbf{x}_{t-1}|\mathbf{x}_t, x_0)$ follows a Gaussian distribution with variance σ_t^2, show this distribution.

(3)(10pts) If the condition in (2) is relaxed to a, b, where $0 < a < b < T$, show the distribution of $P(\mathbf{x}_a|\mathbf{x}_b, \mathbf{x}_0)$