Dieter Fox

Homework 5

Due on Friday, Dec 8, 2017
(No late submissions will be accepted)

1. (8 points) **Probabilities**

Consider the *conditional probability distributions* below.

A	**B**	**P(A|B)**
true	true	0.5
false	true	0.5
true	false	0.1
false	false	0.9

B	**C**	**P(B|C)**
true	true	0.05
false	true	0.95
true	false	0.6
false	false	0.4

C	**P(C)**
true	0.1
false	0.9

(a) (2 points) Compute the joint distribution $P(B, C)$. Write all values needed to describe this distribution.

B	**C**	**P(B, C)**
true	true	0.005
false	true	0.095
true	false	0.54
false	false	0.36

(b) (2 points) Compute $P(B)$ by summing out the necessary variables. Write all values needed to describe this distribution.

B	**P(B)**
true	$0.005 + 0.54 = 0.545$
false	$0.095 + 0.36 = 0.455$

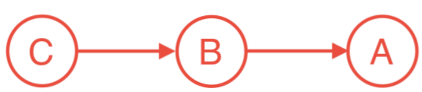
(c) (2 points) Use the chain rule to factor the joint probability distribution $P(A, B, C)$. What specific conditional independence assumption(s) must be made between the three variables for the joint distribution to be fully described by the conditional distributions $P(A|B)$, $P(B|C)$, and $P(C)$?

By the chain rule, $P(A, B, C) = P(C)P(B|C)P(A|B, C)$ with no independence assumptions. For $P(A|B, C) = P(A|B)$, we have to assume that A is independent of C given B.

(d) (1 points) Assuming the independence assumptions in (c), what is $P(A = false \wedge B = true \wedge C = false)$? Use the probabilities provided above.

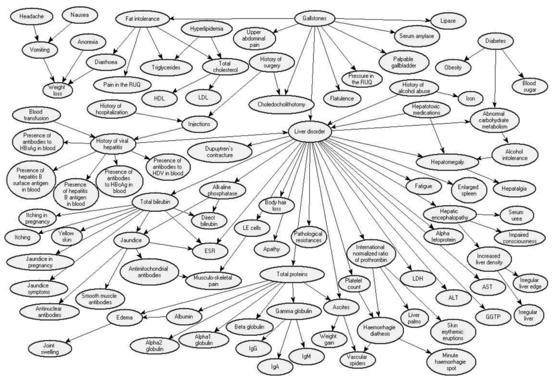
$0.5 * 0.6 * 0.9$

(e) (1 points) Draw a Bayesian Network that encodes the independence assumptions that allows the joint distribution P(A,B,C) to be fully described by the three conditional distributions above.

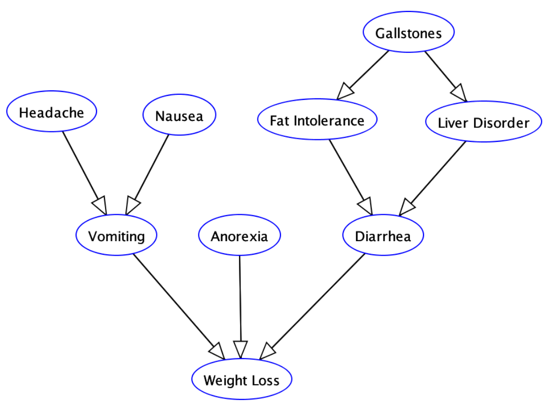


2. (8 points) **Bayes Net: Independence**

   In 1999, Professor Agnieszka Onisko published a research paper modeling liver disorder diagnosis via a Bayesian Network model which featured an "astounding" 94 variables, shown below.



   Let us interpret some independence properties of a small subset of the featured model, using only the "Headache", "Nausea", "Vomiting", "Anorexia", "Fat Intolerance", "Gallstones", "Diarrhea", "Weight Loss" and "Liver Disorder" features, shown below.



   The following (Y/N) questions are worth 1 point each with a negative point for incorrect answers (don't guess randomly). By independent we mean whether they are independent for any setting of the CPTs.

   (a) Is Headache independent of Anorexia? Y

   (b) Is Gallstones conditionally independent of Nausea given Anorexia? Y

   (c) Is Gallstones conditionally independent of Fat Intolerance given Liver Disorder? N

   (d) Is Gallstones conditionally independent of Diarrhea given Fat Intolerance? N

   (e) Is Gallstones conditionally independent of Weight Loss given Fat Intolerance? N

   (f) Is Weight Loss conditionally independent of Fat Intolerance given Diarrhea? Y

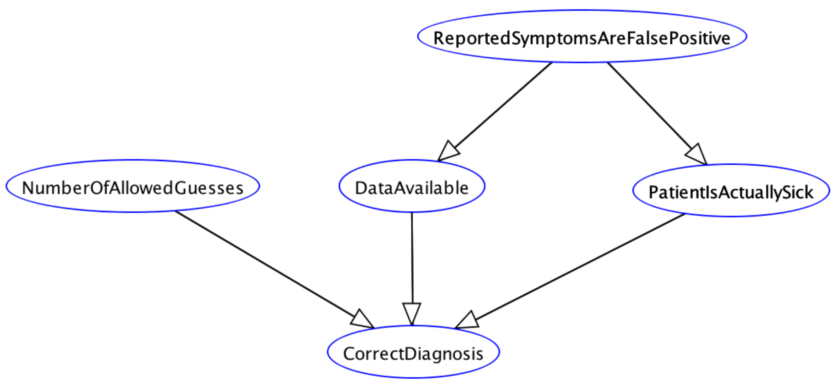
   (g) Is Headache conditionally independent of Anorexia given Weight Loss and Vomiting? Y

(h) Is Headache conditionally independent of Diarrhea given Anorexia and Vomiting?
Y

3. (8 points) **Bayes Net: Inference**

Background Context: In her paper, Professor Onisko modeled 16 different diseases related to liver disorder. Claiming a 67% "Top-4" classification accuracy (meaning that a patient's correct liver disorder given symptoms was among the four highest probability guesses generated by the network), while achieving 34% "Top-1" accuracy.

Let us consider the conditional independence for the following Bayesian Network model.



(a) (1 points) What are the conditional probability distribution tables needed to completely represent the above Bayesian Network? No need to provide values.

- Let $Symp$ be shorthand for $ReportedSymptomsAreFalsePositive$
- Let $Sick$ be shorthand for $PatientIsActuallySick$
- Let $Data$ be shorthand for $DataAvailable$
- Let Num be shorthand for $NumberOfAllowedGuesses$
- Let $Correct$ be shorthand for $CorrectDiagnosis$

$P(Symp)$, $P(Num)$, $P(Data|Symp)$, $P(Sick|Symp)$, $P(Correct|Data, Sick, Num)$

(b) (7 points) Using Variable Elimination, show each step of the general method to compute $P(Correct|Symp = True, Num = 1)$. You can assume that all variables are discrete. Please use the following notation if you want to sum over the values of a variable, for example, $X : \sum_x P(X = x)$. Define any new factors such as $f_2(X, y) = \sum_z f_1(X, y, Z = z)P(X|Z = z)$. Eliminate variables in alphabetical order.

(a) Apply evidence observation
$P(Symp = True)$, $P(Num = 1)$, $P(Data|Symp = True)$, $P(Sick|Symp = True)$, $P(Correct|Data, Sick, Num = 1)$

(b) Eliminate Variables in alphabetical order: *Data*

i. Join relevant distribution tables, with "intermediate factor" f_1'

$$f_1'(Symp = True, Num = 1, Data, Correct, Sick) \tag{1}$$
$$= P(Data|Symp = True)P(Correct|Data, Num = 1, Sick) \tag{2}$$
$$= P(Data, Correct|Symp = True, Num = 1, Sick) \tag{3}$$

ii. Sum out, and eliminate *Data* in place with a new factor f_1

$$f_1(Symp = True, Num = 1, Correct, Sick) \tag{4}$$
$$= \sum_{d \in Data} P(Data = d|Symp = True)P(Correct|Data = dNum = 1, Sick) \tag{5}$$
$$= \sum_{d \in Data} f_1'(Symp = True, Num = 1, Data = d, Correct, Sick) \tag{6}$$
$$= P(Correct|Symp = True, Num = 1, Sick) \tag{7}$$

(c) Eliminate Variables in alphabetical order: *Sick*

i. Join relevant distribution tables, with "intermediate factor" f_2'

$$f_2'(Symp = True, Num = 1, Correct, Sick) \tag{8}$$
$$= f_1(Symp = True, Num = 1, Correct, Sick)P(Sick|Symp = True) \tag{9}$$
$$\tag{10}$$

ii. Sum out, and eliminate *Sick* in place with a new factor f_2

$$f_2(Symp = True, Num = 1, Correct) \tag{11}$$
$$= \sum_{s \in Sick} f_1(Symp = True, Num = 1, Correct, Sick = s)P(Sick = s|Symp = True) \tag{12}$$
$$= \sum_{s \in Sick} f_2'(Symp = True, Num = 1, Correct, Sick = s) \tag{13}$$
$$= P(Correct|Symp = True, Num = 1) \tag{14}$$

(d) Join remaining factors and Normalize over the Query Variable $Correct$

i. Only factor: $P(Correct|Symp = True, Num = 1)$

ii. Simply normalize across values of $c \in Correct$ to compute the probability

4. (10 points) **Hidden Markov Models**

At time t, Kenny is in some state X_t. The two states Kenny alternates between are saving the world (denote this as S) and being a student in CSE (denote this as C). Let the evidence E_t be whether or not Kenny is seen in the CSE labs at time t.

The transition probabilities are provided in the following table, where the row corresponds to X_{t-1} and the column to X_t.

	S	C
S	0.8	0.2
C	0.4	0.6

For example, $P(X_t = S|X_{t-1} = C) = 0.4$.

The model for evidence E_t is provided in the following table, where the row corresponds to X_t and the column to E_t.

	$true$	$false$
S	0.1	0.9
C	0.7	0.3

For example, $P(E_t = false|X_t = S) = 0.9$.

(a) (2 point) Let the initial beliefs be $P(X_0 = S) = P(X_0 = C) = 0.5$. Fill in the following table for $t = 1$, first computing for passage of time, and then for observation of evidence $E_1 = true$. Finally, normalize the values from the observation column to get the beliefs. Round to three decimal places.

	passage of time	**observation**	$B(X_1)$
S	$0.8 \cdot 0.5 + 0.4 \cdot 0.5 = 0.6$	$0.1 \cdot 0.6 = 0.06$	0.176
C	$0.2 \cdot 0.5 + 0.6 \cdot 0.5 = 0.4$	$0.7 \cdot 0.4 = 0.28$	0.824

(b) (2 point) Repeat for $t = 2$, with the observation of evidence $E_2 = false$. When using any previous value for computations, use their rounded value. Round to three decimal places.

	passage of time	**observation**	$B(X_2)$
S	$0.8 \cdot 0.176 + 0.4 \cdot 0.824 = 0.470$	$0.9 \cdot 0.470 = 0.423$	0.727
C	$0.2 \cdot 0.176 + 0.6 \cdot 0.824 = 0.530$	$0.3 \cdot 0.530 = 0.159$	0.273

(c) (2 point) Recall from lecture that $m_t[x_t] = \max_{x_{1:t-1}} P(x_{1:t-1}, x_t, e_{1:t})$, which is the probability of the most likely path that ends at x_t, considering the path up to t and the evidence up to t.

Use Viterbi's algorithm to compute $m_1[S_1], m_1[C_1], m_2[S_2]$, and $m_2[C_2]$ for the sequence of evidence $\{E_1 = true,\ E_2 = false\}$. Define $m_0[S] = m_0[C] = 0.5$. Use exact numbers in your calculations and answers. What was Kenny most likely doing at time $t = 1$ and at time $t = 2$?

$$\begin{aligned} m_1[S_1] &= P(E_1 = true|X_1 = S) \max_{X_0} P(X_1 = S|X_0) m_0[X_0] \\ &= 0.1 \cdot \max(0.8 \cdot 0.5, 0.4 \cdot 0.5) \\ &= 0.04 \end{aligned}$$

$$\begin{aligned} m_1[C_1] &= P(E_1 = true|X_1 = C) \max_{X_0} P(X_1 = C|X_0) m_0[X_0] \\ &= 0.7 \cdot \max(0.2 \cdot 0.5, 0.6 \cdot 0.5) \\ &= 0.21 \end{aligned}$$

$$\begin{aligned} m_2[S_2] &= P(E_2 = false|X_2 = S) \max_{X_1} P(X_2 = S|X_1) m_1[X_1] \\ &= 0.9 \cdot \max(0.8 \cdot 0.04, 0.4 \cdot 0.21) \\ &= 0.0756 \end{aligned}$$

$$\begin{aligned} m_2[C] &= P(E_2 = false|X_2 = C) \max_{X_1} P(X_2 = C|X_1) m_1[X_1] \\ &= 0.3 \cdot \max(0.2 \cdot 0.04, 0.6 \cdot 0.21) \\ &= 0.0378 \end{aligned}$$

At $t = 1$, Kenny was a CSE student. At $t = 2$, Kenny was saving the world.

(d) (1 point) Assume now we are using a particle filter with 3 particles to approximate our belief instead of using exact inference. Imagine we have just applied transition model sampling (elapse-time) from state X_0 to X_1, and now have the set of particles $\{S, S, C\}$. What is our belief about X_1 before considering noisy evidence?

X_1	$B(X_1)$ (elapse time)
S	$2/3$
C	$1/3$

(e) (2 points) Now assume we receive sensor evidence $E_1 = true$. What is the weight for each particle, and what is our belief now about X_1 (before weighted resampling)?

Particle	Weight
S	0.1
S	0.1
C	0.7

X_1	$B(X_1)$ (after observation)
S	0.222
C	0.778

(f) (1 point) Will performing weighted resampling on these weighted particles to obtain our final three particle representation for X_1 cause our belief to change? **Briefly** explain why or why not.

Yes, because there will be three unweighted particles which can't represent this belief

5. **(Optional, not graded) Create Bayes Net**

Create a Bayes net with exactly four states {A,B,C,D}, that follows all of the independence constraints below.

(a) $A \perp\!\!\!\perp B$

(b) $A \not\perp\!\!\!\perp D|B$

(c) $A \perp\!\!\!\perp D|C$

(d) $A \not\perp\!\!\!\perp C$

(e) $B \not\perp\!\!\!\perp C$

(f) $A \not\perp\!\!\!\perp B|D$

(g) $B \perp\!\!\!\perp D|A, C$

Answer: