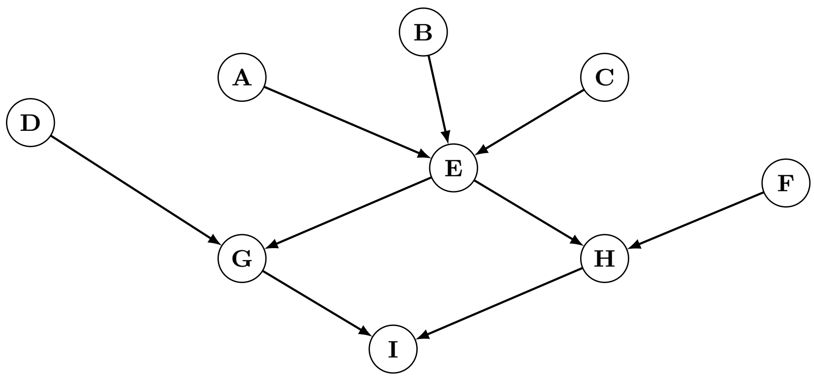
50.007 Machine Learning, Fall 2020
Homework 4

Due Sunday 13 Dec 2020, 11.59pm

Sample Solutions

In this homework, we would like to look at the Bayesian Networks. You are given a Bayesian network as below. All nodes can take 2 different values: $\{1, 2\}$.



1. (10 pts) How many effective parameters are needed to for this Bayesian network? What would be the number of effective parameters for the same network if node **D** and **F** can take 4 different values: $\{1, 2, 3, 4\}$, and all other nodes can only take 2 different values: $\{1, 2\}$?

   Q1 - each question is worth 5 pts.

   Answer The number of parameters correspond to the number of entries in the probability table of each node in the Bayesian network. Assume the number of values for node k to take is r_k. For a node i with parents pa_i, the number of rows is $\prod_{j \in pa_i} r_j$. The number of columns is r_i. However the values in the last column can be uniquely determined from the other columns since the values of each row sum to 1. This means for the node i there are $(r_i - 1)\prod_{j \in pa_i} r_j$ free/independet/effective parameters involved.

   Therefore in the initial Bayesian network, the number of free parameters is:

   $1(A)+1(B)+1(C)+1(D)+2\times2\times2\times1(E)+1(F)+2\times2\times1(G)+2\times2\times1(H)+2\times2\times1(I) = 25$

   If node D and F can take 4 different values: 1, 2, 3, 4, and all other nodes can only take 2 different values: 1, 2, the number of free parameters is:

   $1(A)+1(B)+1(C)+3(D)+2\times2\times2\times1(E)+3(F)+2\times4\times1(G)+2\times4\times1(H)+2\times2\times1(I) = 37$

2. (10 pts) Without knowing the actual value of any node, are node **A** and **F** independent of each other? What if we know the value of node **C** and **I**?

   Q2 - each question is worth 5 pts.

Answer Without knowing the actual value of any node, node A and F are independent of each other. This is because there does not exist any path from A to F that is open. Based on the Bayes' ball algorithm, A and F are independent of each other.

If we know the value of node C and I, then the two variables A and F become dependent. This is because there exist a path connecting A and F that is open: $A - E - H - I - H - F$ or $A - E - G - I - H - F$. Based on the Bayes' ball algorithm, A and F are dependent.

3. (10 pts) If we have the following probability tables for the nodes. Compute the following probability. Clearly write down all the necessary steps.

$$P(\mathbf{E} = 1 | \mathbf{C} = 2)$$

	A 1	**A** 2
	0.2	0.8

	B 1	**B** 2
	0.5	0.5

	C 1	**C** 2
	0.2	0.8

	D 1	**D** 2
	0.1	0.9

A	**B**	**C**	**E** 1	**E** 2
1	1	1	0.1	0.9
1	1	2	0.3	0.7
1	2	1	0.5	0.5
1	2	2	0.0	1.0
2	1	1	0.9	0.1
2	1	2	0.6	0.4
2	2	1	0.4	0.6
2	2	2	0.5	0.5

	F 1	**F** 2
	0.3	0.7

D	**E**	**G** 1	**G** 2
1	1	0.1	0.9
1	2	0.4	0.6
2	1	0.5	0.5
2	2	0.5	0.5

E	**F**	**H** 1	**H** 2
1	1	0.1	0.9
1	2	0.4	0.6
2	1	0.5	0.5
2	2	0.5	0.5

G	**H**	**I** 1	**I** 2
1	1	0.1	0.9
1	2	0.9	0.1
2	1	0.1	0.9
2	2	0.9	0.1

Q3 - derivation of $P(C, E)$ deserves 4 pts, derivation of $P(E|C)$ deserves 2 pts, calculation of $P(E = 1|C = 2)$ deserves 2 pts, and the final result is worth 2 pts.

Answer One standard approach is to start by computing the following marginal probability:

$$P(C, E) = \sum_{A,B,D,F,G,H,I} P(A)P(B)P(C)P(D)P(E|A, B, C)P(F)P(G|D, E)P(H|E, F)P(I|G, H)$$

Simplify the above expression, and next compute $P(C = 2, E = 1)$ and $P(C = 2, E = 2)$ respectively, and then compute $P(C = 2) = P(C = 2, E = 1) + P(C = 2, E = 2)$. The conditional probability $P(E = 1|C = 2) = P(C = 2, E = 1)/P(C = 2)$.

Here we describe an alternative approach based on some observations about the independence properties of the graph.

Note that given E, the variables A, B, C and D, F, G, H, I are conditionally independent. Mathematically, this means:

$$P(D, F, G, H, I|A, B, C, E) = P(D, F, G, H, I|E)$$

Now, mathematically, we always have the following:

$$P(A, B, C, D, E, F, G, H, I) = P(A, B, C, E)P(D, F, G, H, I|A, B, C, E)$$

Based on the earlier equation, we have:

$$P(A, B, C, D, E, F, G, H, I) = P(A, B, C, E)P(D, F, G, H, I|E)$$

This yeids:

$$\begin{aligned}
P(C, E) &= \sum_{ABDFGHI} P(A, B, C, D, E, F, G, H, I) \\
&= \sum_{ABDFGHI} P(A, B, C, E)P(D, F, G, H, I|E) \\
&= \sum_{AB} P(A, B, C, E) \sum_{DFGHI} P(D, F, G, H, I|E) \\
&= \sum_{AB} P(A, B, C, E) \\
&= \sum_{AB} P(A)P(B)P(C)P(E|A, B, C) \\
&= P(C) \sum_{AB} P(A)P(B)P(E|A, B, C)
\end{aligned}$$

$$P(E|C) = \frac{P(C, E)}{P(C)} = \sum_{AB} P(A)P(B)P(E|A, B, C)$$

$$\begin{aligned}
P(E = 1|C = 2) = P(A = 1)P(B = 1)P(E = 1|A = 1, B = 1, C = 2) \\
+P(A = 1)P(B = 2)P(E = 1|A = 1, B = 2, C = 2) \\
+P(A = 2)P(B = 1)P(E = 1|A = 2, B = 1, C = 2) \\
+P(A = 2)P(B = 2)P(E = 1|A = 2, B = 2, C = 2) \\
= 0.2 \times 0.5 \times 0.3 + 0.2 \times 0.5 \times 0.0 + 0.8 \times 0.5 \times 0.6 + 0.8 \times 0.5 \times 0.5 \\
= 0.03 + 0.0 + 0.24 + 0.2 = 0.47
\end{aligned}$$

4. (10 pts) Now, assume we do not have any knowledge about the probability table for the nodes in the network, but we have the following 12 observations. Find a way to estimate the probability table associated with the nodes **A** and **H**.

A	B	C	D	E	F	G	H	I
1	1	2	2	1	2	1	1	1
1	2	1	1	2	1	1	1	2
2	2	2	1	2	2	1	2	1
1	1	2	1	2	1	1	2	2
1	2	1	1	1	1	2	1	1
2	2	1	2	1	2	2	1	2
2	1	2	2	1	2	2	2	1
1	2	2	1	2	1	2	2	2
1	1	1	1	2	2	1	1	1
1	1	1	1	2	1	1	1	2
1	2	1	2	2	1	2	1	2
2	1	1	2	1	2	2	1	1

Q4 - each result in Eq. (1) deserves 1 pt.

Answer We can use the maximum likelihood estimation to find the optimal model parameters.

$$\begin{aligned}
\theta_A(1) &= \frac{\text{Count}(A=1)}{\text{Count}(A)} = 8/12 \\
\theta_A(2) &= \frac{\text{Count}(A=2)}{\text{Count}(A)} = 4/12 \\
\theta_H(1) &= \frac{\text{Count}(E=1,F=1,H=1)}{\text{Count}(E=1,F=1)} = 1 \\
\theta_H(2) &= \frac{\text{Count}(E=1,F=1,H=2)}{\text{Count}(E=1,F=1)} = 0 \\
\theta_H(1) &= \frac{\text{Count}(E=1,F=2,H=1)}{\text{Count}(E=1,F=2)} = 3/4 \\
\theta_H(2) &= \frac{\text{Count}(E=1,F=2,H=2)}{\text{Count}(E=1,F=2)} = 1/4 \\
\theta_H(1) &= \frac{\text{Count}(E=2,F=1,H=1)}{\text{Count}(E=2,F=1)} = 3/5 \\
\theta_H(2) &= \frac{\text{Count}(E=2,F=1,H=2)}{\text{Count}(E=2,F=1)} = 2/5 \\
\theta_H(1) &= \frac{\text{Count}(E=2,F=2,H=1)}{\text{Count}(E=2,F=2)} = 1/2 \\
\theta_H(2) &= \frac{\text{Count}(E=2,F=2,H=2)}{\text{Count}(E=2,F=2)} = 1/2
\end{aligned} \tag{1}$$

The resulting probability tables for A and H are:

	A	
	1	2
	8/12	4/12

E	**F**	**H** 1	2
1	1	1	0
1	2	3/4	1/4
2	1	3/5	2/5
2	2	1/2	1/2