10-601 Machine Learning, Fall 2009: Homework 2 Solutions

Due: Wednesday, September 16^{nd}, 10:30 am

Instructions There are 4 questions on this assignment worth the total of 100 points. Please hand in a hard copy at the beginning of the class. Refer to the webpage for policies regarding collaboration, due dates, and extensions.

1 Bayesian networks and factor graphs [12 pts]

1. For each of the networks given in Figure 1 (a,b,c), do the following statements hold? Please explain your reasoning.
   - $A \perp C \mid B, D$ [3 x 2 pts]
   - $B \perp D \mid A, C$ [3 x 2 pts]

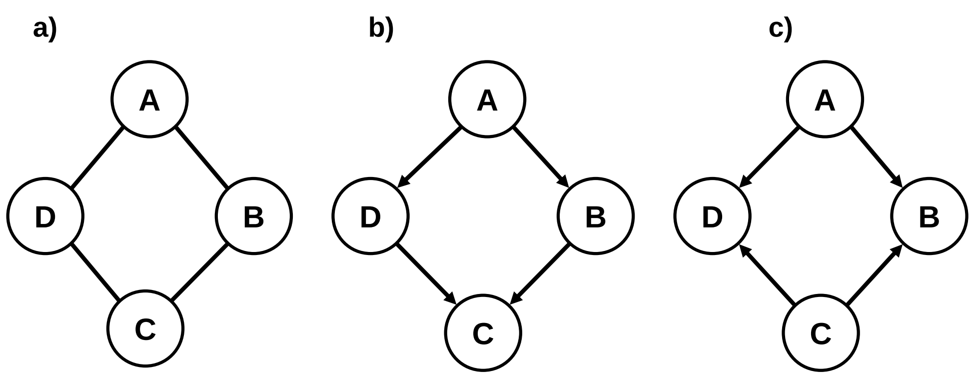

Figure 1: Factor graph (a) and Bayesian networks (b,c) for Problem 1.

★ **SOLUTION:**

1. For the factor graph shown in Figure 1 (a)
   - $A \perp C \mid B, D$
     True. All the paths between A and C are inactive once B and D are observed.
   - $B \perp D \mid A, C$
     True. Similarly all the paths between B and D are inactive when A and C are observed.
2. For the Bayesian network shown in Figure 1 (b)

- $A \perp C \mid B, D$
  True. All the paths are inactive between A and C.
- $B \perp D \mid A, C$
  False. Notice the v-structure on C; once C is observed the path between B and D is active.

3. For the Bayesian network shown in Figure 1 (c)
   - $A \perp C \mid B, D$
     False. There is an active between A and C once B and D are observed. Notice the v-structure on both B and D.
   - $B \perp D \mid A, C$
     True. All the paths between A and C are inactive.

2 Conditional probabilities [8 pts]

Prove or disprove (by providing a counter-example) each of the following properties of independence:

1. $\mathbf{P}(X, Y \mid Z) = \mathbf{P}(X \mid Y, Z)\,\mathbf{P}(Y \mid Z)$ [3 pts]

   ★ **SOLUTION:**

$$
\begin{aligned}
\mathbf{P}(X, Y \mid Z) &= \frac{\mathbf{P}(X, Y, Z)}{\mathbf{P}(Z)} \\
&= \frac{\mathbf{P}(X \mid Y, Z)\,\mathbf{P}(Y \mid Z)\,\mathbf{P}(Z)}{\mathbf{P}(Z)} \\
&= \mathbf{P}(X \mid Y, Z)\,\mathbf{P}(Y \mid Z)
\end{aligned}
$$

   First step follows from Bayes rule and chain rule is applied in the second step.

2. $(X \perp Y \mid Z)$ and $((X, Y) \perp W \mid Z)$ implies $(X \perp W \mid Z)$ (This statement means: If X is independent of Y given Z and the joint probability distribution of X and Y is independent of W given Z, then X and W are independent given Z.) [5 pts]

   ★ **SOLUTION:**

$$
\begin{aligned}
\mathbf{P}(X, W \mid Z) &= \sum_{Y=y} \mathbf{P}(X, Y = y, W \mid Z) \\
&= \sum_{Y=y} \mathbf{P}(X, Y = y, Z)\,\mathbf{P}(W \mid Z) \\
&= \sum_{Y=y} \mathbf{P}(X \mid Z)\,\mathbf{P}(Y = y \mid Z)\,\mathbf{P}(W \mid Z) \\
&= \mathbf{P}(X \mid Z)\,\mathbf{P}(W \mid Z)
\end{aligned}
$$

3 Disease and symptoms [20 pts]

A patient goes to the doctor for a medical condition, the doctor suspects three diseases as the cause of the condition. The three diseases are D_1, D_2, D_3, which are marginally independent from each other. There are four symptoms S_1, S_2, S_3, S_4 which the doctor wants to check for presence in order to find the most probable cause of the condition. The symptoms are conditionally dependent to the three diseases as follows: S_1 depends only on D_1, S_2 depends on D_1 and D_2. S_3 is depends on D_1 and D_3, whereas S_4 depends only on D_3. Assume all random variables are Boolean, they are either 'true' or 'false'.

1. Draw the Bayesian network for this problem. [2 pts]

★ **SOLUTION:** The Bayesian network is shown in Figure 2.

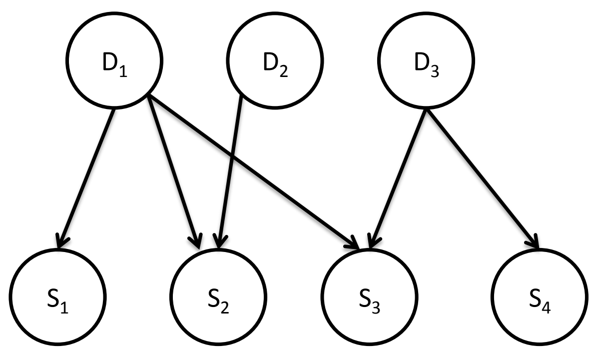

Figure 2: The Bayesian network for disease symptom problem.

2. Write down the expression for the joint probability distribution as a product of conditional probabilities. [2 pts]

★ **SOLUTION:**

$$\mathbf{P}(D_1, D_2, D_3, S_1, S_2, S_3, S_4) = \mathbf{P}(D_1)\,\mathbf{P}(D_2)\,\mathbf{P}(D_3)\,\mathbf{P}(S_1 \mid D_1)\,\mathbf{P}(S_2 \mid D_1, D_2)\,\mathbf{P}(S_3 \mid D_1, D_3)\,\mathbf{P}(S_4 \mid D_3)$$

3. What is the number of independent parameters that is required to describe this joint distribution? [3 pts]

★ **SOLUTION:** The number of independent parameters is 15. The number of independent parameters needed to describe each conditional probability distribution that is part of the joint are listed in Table 1.

4. Assume there were no conditional independence between the variables, how many independent parameters would be required then? [3 pts]

CPT	Number of independent parameters
$\mathbf{P}(D_1)$	1
$\mathbf{P}(D_2)$	1
$\mathbf{P}(D_3)$	1
$\mathbf{P}(S_1 \mid D_1)$	2
$\mathbf{P}(S_2 \mid D_1, D_2)$	4
$\mathbf{P}(S_3 \mid D_1, D_3)$	4
$\mathbf{P}(S_4 \mid D_3)$	2
Total number of independent parameters	15

Table 1: Number of independent parameters for each conditional probability distribution.

★ **SOLUTION:** Without conditional independence assumptions, the number of parameters required to specify the joint would be 127. There are 7 random variables, and each of them take 2 values, therefore $2^7 - 1 = 127$

5. What is the Markov Blanket of variable S_2? [2 pts]

★ **SOLUTION:** The Markov Blanket of S_2 is D_1 and D_2.

6. What is an example of the 'explaining away' phenomenon in the graph? [2 pts]

★ **SOLUTION:** S_3 depends on D_1 and D_3. When we observe S_3 is true and if we observe that D_1 is true, D_3 would be a less likely cause for S_3. This is an example 'explaining away' phenomenon.

7. If we observe the fourth symptom,(S_4 = true), for which diseases do we gain information? [3 pts]

★ **SOLUTION:** By observing S_4 = true we gain information about only disease D_3. Notice D_1 and D_2 are independent of S_4.

8. Suppose we observed second symptom is present in the patient (S_2 = true), what does observing the fourth symptom (S_4 = true) tell us now? [3 pts]

★ **SOLUTION:** With $S_2 = true$, observing $S_4 = true$ still gives us information only about D3. If we had observed $S_3 = true$ instead of S2, then observing $S_4 = true$ would have given us information about D_1 and D_3 due to explaining away.

4 Bayesian about Happiness [60 points]

As part of a comprehensive study of the role of 10-601 on people's happiness we have been collecting important data from graduating students. In an entirely optional survey that all students are required to complete, we ask the following highly objective questions:

- Do you party frequently [`Party`: Yes/No]?
- Are you wicked smart [`Smart`: Yes/No]?
- Are you creative [`Creative`: Yes/No]? (Please only answer Yes or No)
- Did you do well on all your homework assignments? [`HW`: Yes/No]
- Do you use a Mac? [`Mac`: Yes/No]

- Did your 10-601 project succeed? [Project: Yes/No]
- Did you succeed in your most important class (which is 10-601)? [Success: Yes/No]
- Are you currently Happy? [Happy: Yes/No]

You can obtain the comma-separated survey results from http://www.cs.cmu.edu/~ggordon/10601/hws/hw2/students.csv.zip. Each row in students.csv corresponds to the responses of a separate student. The columns in students.csv correspond to each question (random variable) in the order Party, Smart, Creative, HW, Mac, Project, Success, and Happy. The entries are either zero, corresponding to No response, or one, corresponding to a Yes response. After consulting a behavioral psychologist we obtained the following complete set of conditional relationships:

- HW depends only on Party and Smart
- Mac depends only on Smart and Creative
- Project depends only on Smart and Creative
- Success depends only on HW and Project
- Happy depends only on Party, Mac, and Success

4.1 Understanding The Model [12 Points]

1. Draw the Bayesian network.

★ **SOLUTION:** The Bayesian network is shown in Fig. 1. It is obtained by setting each variables parents on which it depends.

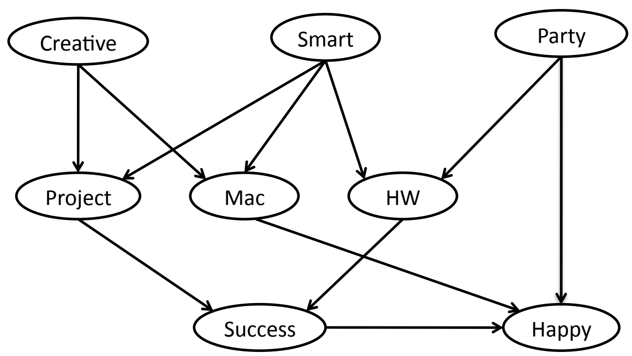

Figure 3: Bayesian Network for Happiness in Life.

2. Write joint distribution as a product of conditional probabilities.

★ **SOLUTION:** The joint probability expression can be read directly from the graph:

$$\mathbf{P}(\text{creative}, \text{smart}, \text{party}, \text{project}, \text{mac}, \text{hw}, \text{success}, \text{happy}) = \\ \mathbf{P}(\text{creative})\,\mathbf{P}(\text{smart})\,\mathbf{P}(\text{party})\,\mathbf{P}(\text{project} \mid \text{creative}, \text{smart})\,\mathbf{P}(\text{mac} \mid \text{creative}, \text{smart}) \\ \mathbf{P}(\text{hw} \mid \text{smart}, \text{party})\,\mathbf{P}(\text{success} \mid \text{project}, \text{hw})\,\mathbf{P}(\text{happy} \mid \text{success}, \text{mac}, \text{party})$$

3. What is the number of independent parameters needed for each conditional probability table?

Factors	Parameters
$\mathbf{P}(\text{creative})$	1
$\mathbf{P}(\text{smart})$	1
$\mathbf{P}(\text{party})$	1
$\mathbf{P}(\text{project} \mid \text{creative}, \text{smart})$	4
$\mathbf{P}(\text{mac} \mid \text{creative}, \text{smart})$	4
$\mathbf{P}(\text{hw} \mid \text{smart}, \text{party})$	4
$\mathbf{P}(\text{success} \mid \text{project}, \text{hw})$	4
$\mathbf{P}(\text{happy} \mid \text{success}, \text{mac}, \text{party})$	8

Table 2: Parameter counts table factor.

★ **SOLUTION:** Since each variable is binary only one parameter is needed for each possible assignment to the conditional. The totals are given in Table 2.

4. What is the total number of independent parameters?

★ **SOLUTION:** This is the sum of the parameter counts from the previous question: 27.

4.2 D-Separation [8 Points]

Using only the Bayesian network structure from part 4.1, answer the following True/False questions and provide a brief explanation:

1. `Party` is independent of `Success` given `HW`.

★ **SOLUTION:** False. There is an active path:

$$\text{party} \rightarrow (\text{hw}) \leftarrow \text{smart} \rightarrow \text{project} \rightarrow \text{success}$$

because of the v-structure at hw is observed.

2. `Party` is independent of `Smart` given `Success`.

★ **SOLUTION:** False. There is an active path between Party and Smart through the v-structure at HW which is activated by Success.

3. `Party` is independent of `Creative` given `Happy`.

★ **SOLUTION:** False. There is an active path between Party and Creative through the v-structure at Happy. There are actually several active paths can you find them?

4.3 Confounded Intelligence [10 Points]

1. Using only the data in `students.csv` and Matlab calculate the correlation between success on the homework `HW` and success on the project `Project`. You do not need to use the Bayesian network for this question. (Hint: Consider using the `cov` function in Matlab.)

★ **SOLUTION:** The code for the solution is given at http://www.cs.cmu.edu/~ggordon/10601/hws/hw2/solution.m. The correlation is computed by taking:

$$\rho_{\text{hw,project}} = \frac{\sigma^2_{\text{hw,project}}}{\sqrt{\sigma^2_{\text{hw,hw}}\sigma^2_{\text{project,project}}}} = 0.35574$$

2. From the model structure, identify a potential common cause variable which may explain the correlation between `HW` and `Project`.

★ **SOLUTION:** Being Smart can explain success on homework and the project. From the graph structure we see that intelligence behaves like a common cause variable. Of course in practice it is possible that success on the homework could lead to an increased intelligence and then ultimately to success on the project.

4.4 Counting [15 Points]

Use Matlab and `students.csv` to calculate the parameters for each conditional probability table by counting with Laplace smoothing. Please consider formatting your conditional probability tables as shown in Table 3.

A	B	C	$\mathbf{P}(X = 1 \mid A, B, C)$
T	T	T	0.4
T	T	F	0.8
T	F	T	0.15
T	F	F	0.16
F	T	T	0.23
F	T	F	0.42
F	F	T	0.4
F	F	F	0.8

Table 3: An example conditional probability table for $\mathbf{P}(X \mid A, B, C)$.

★ **SOLUTION:** For the solution code see http://www.cs.cmu.edu/~ggordon/10601/hws/hw2/solution.m. The CPTs are given in Table 4.

4.5 Inference [15 Points]

Using any of the following software,

- **Recommended:** AISpace Graphical Tool http://www.aispace.org/bayes/version5.1.6/bayes.jnlp other formats (jar, exe, applet) are available http://www.aispace.org/downloads.shtml
- The Matlab Bayes Net Toolbox: http://people.cs.ubc.ca/~murphyk/Software/BNT/bnt.html
- WinBUGS (Bayesian Inference Using Gibbs Sampling) http://www.mrc-bsu.cam.ac.uk/bugs/

along with your conditional probability table estimates, calculate the following probabilities:

★ **SOLUTION:** You can download a complete Bayesian Network for the AISpace Graphical Tool from http://www.cs.cmu.edu/~ggordon/10601/hws/hw2/bayesnet.xml.

- What is the probability of being happy?

$$\mathbf{P}(\text{creative} = T) = 0.69932$$
$$\mathbf{P}(\text{smart} = T) = 0.70472$$
$$\mathbf{P}(\text{party} = T) = 0.60216$$

creative	smart	$\mathbf{P}(\text{project} = T \mid \text{creative}, \text{smart})$
T	T	0.90484
T	F	0.40307
F	T	0.79326
F	F	0.10731

creative	smart	$\mathbf{P}(\text{mac} = T \mid \text{creative}, \text{smart})$
T	T	0.68564
T	F	0.89635
F	T	0.41347
F	F	0.12329

smart	party	$\mathbf{P}(\text{hw} = T \mid \text{smart}, \text{party})$
T	T	0.80252
T	F	0.89790
F	T	0.09447
F	F	0.30556

project	hw	$\mathbf{P}(\text{success} = T \mid \text{project}, \text{hw})$
T	T	0.89633
T	F	0.20737
F	T	0.30714
F	F	0.05066

success	mac	party	$\mathbf{P}(\text{happy} = T \mid \text{success}, \text{mac}, \text{party})$
T	T	T	0.95842
T	T	F	0.35837
T	F	T	0.72082
T	F	F	0.30769
F	T	T	0.49234
F	T	F	0.20619
F	F	T	0.42043
F	F	F	0.09646

Table 4: Learned conditional probability tables.

★ **SOLUTION:**

$$\mathbf{P}(\mathsf{happy} = T) = 0.51575$$

- What is the probability of being happy given that you party often, are wicked smart, but not very creative?

★ **SOLUTION:**

$$\mathbf{P}(\mathsf{happy} = T \mid \mathsf{party} = T, \mathsf{smart} = T, \mathsf{creative} = F) = 0.6922$$

- What is the probability of being happy given that you are wicked smart and very creative?

★ **SOLUTION:**

$$\mathbf{P}(\mathsf{happy} = T \mid \mathsf{smart} = T, \mathsf{creative} = T) = 0.58132$$

- What is the probability of being happy given you do not party, and do well on all your homework and class project?

★ **SOLUTION:**

$$\mathbf{P}\left(\mathsf{happy} = T \,|\, \mathsf{party} = F, \mathsf{hw} = T, \mathsf{project} = T\right) = 0.32108$$

- What is the probability of being happy given you own a mac?

★ **SOLUTION:**

$$\mathbf{P}\left(\mathsf{happy} = T \,|\, \mathsf{mac} = T\right) = 0.56269$$

- What is the probability that you party often given you are wicked smart?

★ **SOLUTION:**

$$\mathbf{P}\left(\mathsf{party} = T \,|\, \mathsf{smart} = T\right) = 0.60216$$

- What is the probability that you party often given you are wicked smart and happy?

★ **SOLUTION:**

$$\mathbf{P}\left(\mathsf{party} = T \,|\, \mathsf{smart} = T, \mathsf{happy} = T\right) = 0.79204$$