Solutions: Homework 1
Statistical methods in AI/ML

Instructor: Vibhav Gogate
Vibhav.Gogate@utdallas.edu

Solutions Available: September 19, midnight

Please email me your marks out of 100 by September 26. If you miss this deadline, then I'll grade your homework.

- **Each problem is worth 10 points**

Problem 1: Propositional logic [Problems 2.6 to 2.9 from AD] This was perhaps the easiest one and there are many ways to prove the following statements. In the following, I'll describe two different approaches: one best on truth tables and the other based on definitions.

[(-2) if you only proved the "if" or the "only if" part.]

- Prove the Refutation theorem, namely $\alpha \models \beta$ iff $\alpha \wedge \neg\beta$ is inconsistent.

  Solution:

  If part: Assume that $\alpha \models \beta$. If a world ω satisfies α, then it must also satisfy β. But then it cannot satisfy $\neg\beta$. Therefore $\alpha \wedge \neg\beta$ is unsatisfiable.

  Only if part: Assume that $\alpha \wedge \neg\beta$ is unsatisfiable. Then every world ω that satisfies α must fail to satisfy $\neg\beta$. In other words, every world ω that satisfies α must satisfy β. Therefore, $\alpha \models \beta$.

- Prove the Deduction theorem, namely $\alpha \models \beta$ iff $\alpha \Rightarrow \beta$ is valid.

  Solution:

  Note that $\alpha \models \beta$ if models of α (namely whenever α is true) are models of β.

  We will prove equivalence using truth tables:

α	β	$\alpha \models \beta$	$\alpha \Rightarrow \beta$
T	T	T	T
T	F	F	F
F	T	T	T
F	F	T	T

  If part: clear from the truth table

  Only If part: Also clear from the truth table.

- Prove that if $\alpha \models \beta$ then $\alpha \wedge \beta$ is equivalent to α.

  Solution:

  We can again use the truth table method:

α	β	$\alpha \models \beta$	$\alpha \wedge \beta$	$\alpha \wedge \beta \Leftrightarrow \alpha$
T	T	T	T	T
T	F	F	F	F
F	T	T	F	T
F	F	T	F	T

  If part: From the truth table, whenever $\alpha \models \beta$ is true (rows 1, 3 and 4) $\alpha \wedge \beta \Leftrightarrow \alpha$ is also true.

- Prove that if $\alpha \models \beta$ then $\alpha \vee \beta$ is equivalent to β.

  Solution:

  We can again use the truth table method:

α	β	$\alpha \models \beta$	$\alpha \vee \beta$	$\alpha \vee \beta \Leftrightarrow \beta$
T	T	T	T	T
T	F	F	T	F
F	T	T	T	T
F	F	T	F	T

  If part: From the truth table, whenever $\alpha \models \beta$ is true (rows 1, 3 and 4) $\alpha \vee \beta \Leftrightarrow \beta$ is also true.

Problem 2: Probability theory

- Prove that the following two definitions of conditional independence are equivalent.

  1. $\Pr(\alpha|\beta \wedge \gamma) = \Pr(\alpha|\gamma)$
  2. $\Pr(\alpha \wedge \beta|\gamma) = \Pr(\alpha|\gamma)\Pr(\beta|\gamma)$

  [**Hint:** Derive one using the other using laws of probability.]

  You can start with (1) and derive (2) or you can start with (2) and derive (1). Below, I'll demonstrate how to derive (2) from (1).

  Solution:

$$
\begin{aligned}
\Pr(\alpha|\beta \wedge \gamma) &= \Pr(\alpha|\gamma) \\
\therefore \frac{\Pr(\alpha \wedge \beta \wedge \gamma)}{\Pr(\beta \wedge \gamma)} &= \Pr(\alpha|\gamma)\frac{\Pr(\gamma)}{\Pr(\gamma)} \\
\therefore \frac{\Pr(\alpha \wedge \beta \wedge \gamma)}{\Pr(\gamma)} &= \Pr(\alpha|\gamma)\frac{\Pr(\beta \wedge \gamma)}{\Pr(\gamma)} \\
\therefore \Pr(\alpha \wedge \beta|\gamma) &= \Pr(\alpha|\gamma)\Pr(\beta|\gamma)
\end{aligned}
$$

Problem 3: Probability theory

We solved the following problem in class.

- **Fact 1:** The probability that your partner fails a lie detector test given that he/she is cheating on you is 0.98 (or 98%). The probability that your partner fails the test given that he/she is not cheating on you is 0.02.
- **Fact 2:** You are a CS graduate student.
- **Fact 3:** You should break up with your partner if he/she is cheating on you with a probability greater than 0.05.

Today, you found that you failed the lie detector test. You are in panic mode and are sure that your partner will break up with you. Suddenly you realize that you had previously found the following, pretty reliable statistic on the internet: only 1 out of 10000 CS graduate students (since they are boring people) cheat on their partners. Given this new information and using sound probabilistic arguments, how do you convince your partner that you are not cheating on him/her?

We derived in class that with this new information the probability of cheating given that you have failed the test is much smaller than 0.05.

Here is the twist; your partner is not totally convinced. He/She tells you to take the test three more times on three different days. Unfortunately, you fail two out of the three tests. Can you still convince your partner that you are not cheating on him/her (assume that the three tests are independent of each other). Justify your answer using purely probabilistic arguments.

Solution: Most of you got this wrong. Please review the following carefully.

Remember that you took 4 tests. Let us denote them by L_1, L_2, L_3 and L_4 respectively. Let C denote cheating.

L denotes that you failed the test and $\neg L$ denotes that you passed the test.

Given:

- $P(L_1|C) = P(L_2|C) = P(L_3|C) = P(L_4|C) = 0.98$
- $P(L_1|\neg C) = P(L_2|\neg C) = P(L_3|\neg C) = P(L_4|\neg C) = 0.02$
- $P(C) = \frac{1}{10000} = 0.0001$

To compute: $P(C|L_1, L_2, L_3, \neg L_4)$. Let us assume that you passed L_4.

From Bayes rule and the fact that all the tests are independent, we have:

$$\begin{aligned} P(C|L_1, L_2, L_3, \neg L_4) &= \frac{P(L_1, L_2, L_3, \neg L_4|C)P(C)}{P(L_1, L_2, L_3, \neg L_4)} \\ &= \frac{P(L_1|C)P(L_2|C)P(L_3|C)P(\neg L_4|C)P(C)}{P(L_1)P(L_2)P(L_3)P(\neg L_4)} \end{aligned} \tag{1}$$

[7 points for deriving this formula]

We know all the terms in the numerator. To compute the denominator, we use the conditioning rule.

$$\begin{aligned} P(L_1) = P(L_2) = P(L_3) = P(L_4) &= P(L_1, C) + P(L_2, \neg C) \\ &= P(L_1|C)P(C) + P(L_2|\neg C)P(\neg C) \\ &= (0.98 * 0.0001) + (0.02 * (1 - 0.0001)) \\ &= 0.020096 \end{aligned}$$

Therefore, $P(\neg L_4) = 1 - P(L_4) = 1 - 0.020096 = 0.979904$.

Substituting these values in Equation 1, we get:

$$P(C|L_1, L_2, L_3, \neg L_4) = \frac{0.98 \times 0.98 \times 0.98 \times 0.02 \times 0.0001}{0.020096 \times 0.020096 \times 0.020096 \times 0.979904} = 0.23669$$

Oops! We can no longer convince our partner.

[3 points for getting the correct answer]

Problem 4: Probability theory (Exercise 2.10 from Koller & Friedman)

The question investigates the way in which conditional independence relationships affect the amount of information needed for probabilistic calculations. Let α, β, and γ be three propositional variables.

- Suppose we wish to calculate $\Pr(\alpha|\beta, \gamma)$ and we have no conditional independence information. Which of the following sets of numbers is sufficient for the calculation?

  1. $\Pr(\alpha, \beta)$, $\Pr(\alpha)$, $\Pr(\beta|\alpha)$ and $\Pr(\gamma|\alpha)$.
  2. $\Pr(\beta, \gamma)$, $\Pr(\alpha)$ and $\Pr(\beta, \gamma|\alpha)$
  3. $\Pr(\beta|\alpha)$, $\Pr(\gamma|\alpha)$ and $\Pr(\alpha)$.

  For each case, justify your response either by showing how to calculate the desired answer or by explaining why this is not possible.

- Suppose we know that β and γ are conditionally independent given α. Now which of the preceding three sets is sufficient. Justify your response as before.

Solution: (1) and (3) are not possible while (2) is possible because of Bayes theorem.

With the given independence assumptions, (1) and (3) are still not possible. You will need $\Pr(\beta|\gamma)$ and it is given in neither (1) nor (3).

This question was little bit tricky. However, most of you got it.

5 points for getting part 1 correct and 5 points for getting part 2 correct.

Problem 5: Independence relations

- Prove that Weak Union and Contraction hold for any probability distribution Pr.
- Provide a counter-example to the intersection property. (You cannot use the counter example given in AD, you have to make your own.)

Solution:

- Weak union says that $I_{\text{Pr}}(X, Z, Y \cup W)$ only if $I_{\text{Pr}}(X, Z \cup W, Y)$.

  Let us assume that $I_{\text{Pr}}(X, Z, Y \cup W)$ holds.

  $$\begin{aligned}
  \Pr(X, Y|Z, W) &= \frac{\Pr(X, Y, W|Z)}{\Pr(W|Z)} \\
  &= \frac{\Pr(X|Z)\Pr(Y, W|Z)}{\Pr(W|Z)} \text{(This follows from } I_{\text{Pr}}(X, Z, Y \cup W)) \\
  &= \Pr(X|Z)\Pr(Y|Z, W)
  \end{aligned}$$

  From Decomposition property, we have $I_{\text{Pr}}(X, Z, Y \cup W) \Rightarrow I_{\text{Pr}}(X, Z, Y)$. Therefore, $\Pr(X|Z) = \Pr(X|Z, W)$ and the proof follows.

- To prove contraction, we have to prove that $I_{\text{Pr}}(X, Z, Y)$ and $I_{\text{Pr}}(X, Z \cup Y, W)$ only if $I_{\text{Pr}}(X, Z, Y \cup W)$.

  Let us assume that both $I_{\text{Pr}}(X, Z, Y)$ and $I_{\text{Pr}}(X, Z \cup Y, W)$ hold.

  Consider:

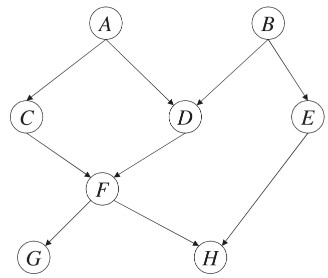
  $$\begin{aligned}
  \Pr(X, Y, W|Z) &= \Pr(X, W|Z, Y)\Pr(Y|Z) \\
  &= \Pr(X|Z, Y)\Pr(W|Z, Y)\Pr(Y|Z) \text{ (This follows from } I_{\text{Pr}}(X, Z \cup Y, W)) \\
  &= \Pr(X|Z, Y)\Pr(W, Y|Z) \\
  &= \Pr(X|Z)P(W, Y|Z) \text{ (This follows from } I_{\text{Pr}}(X, Z, Y))
  \end{aligned}$$

  This proves contraction.

- The simplest counter example is $X = Y = W$. Here both $I(X, Z \cup Y, W)$ and $I(X, Z \cup W, Y)$ hold. However, $I(X, Z, Y \cup W)$ does not hold.

  (2 points for proving weak union and 4 points each for the remaining two).

Problem 6: Bayesian networks (AD Exercise 4.1)



A	Θ_A
1	.2
0	.8

B	Θ_B
1	.7
0	.3

B	E	$\Theta_{E\|B}$
1	1	.1
1	0	.9
0	1	.9
0	0	.1

A	B	D	$\Theta_{D\|AB}$
1	1	1	.5
1	1	0	.5
1	0	1	.6
1	0	0	.4
0	1	1	.1
0	1	0	.9
0	0	1	.8
0	0	0	.2

Consider the Bayesian network given above.

1. List the Markovian assumptions asserted by the DAG.
2. Express $\Pr(a, b, c, d, e, f, g, h)$ in terms of network parameters.
3. Compute $\Pr(A = 0, B = 0)$ and $\Pr(E = 1|A = 1)$. Justify your answers.
4. True or false? Why?
   (a) $dsep(A, BH, E)$
   (b) $dsep(G, D, E)$
   (c) $dsep(AB, F, GH)$

Solution:

3 points The Markovian assumptions are:

$I(\{A\}, \emptyset, \{B, E\})$
$I(\{B\}, \emptyset, \{A, C\})$
$I(\{C\}, \{A\}, \{B, D, E\})$
$I(\{D\}, \{A, B\}, \{C, E\})$
$I(\{E\}, \{B\}, \{A, C, D, F, G\})$
$I(\{F\}, \{C, D\}, \{A, B, E\})$
$I(\{G\}, \{F\}, \{A, B, C, D, E, H\})$
$I(\{H\}, \{E, F\}, \{A, B, C, D, G\})$

2 points Using the chain rule for Bayesian networks, we have:

$$\Pr(a,b,c,d,e,f,g,h) = \Pr(a)\Pr(b)\Pr(c|a)\Pr(d|a,b)\Pr(e|b)\Pr(f|c,d)\Pr(g|f)\Pr(h|e,f)$$

2 points $\Pr(A=0,B=0) = \Pr(A=0)\Pr(B=0) = 0.8 \times 0.3 = 0.24$ since A and B are d-separated.

$\Pr(E=1|A=1) = \Pr(E=1) = 0.34$ since E and A are d-separated.

3 points None of the *dsep*'s are true.

Problem 7: Bayesian networks (AD Exercise 4.12)
Construct two distinct DAGs over variables A, B, C, and D. Each DAG must have exactly four edges and the DAGs must agree on d-separation.

Solution: [5 points each] First DAG: $A \to B$, $A \to C$, $B \to D$, $C \to D$

Second DAG: $A \to B$, $C \to A$, $B \to D$, $C \to D$

Many other solutions are possible.

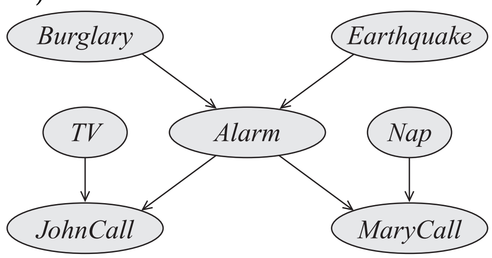
Problem 8: Bayesian networks (AD Exercise 4.15)
Identify a DAG that is a D-MAP for all distributions Pr over variables $\mathbf{X}$. Similarly, identify another DAG that is an I-MAP for all distributions Pr over variables $\mathbf{X}$.

Solution: [5 points each] An empty DAG is a D-map of all distributions. A complete DAG constructed along any variable order is an I-map for all distributions.

Problem 9: Bayesian networks (AD Exercise 4.17)
Prove that for strictly positive distributions, if $\mathbf{B}_1$ and $\mathbf{B}_2$ are Markov blankets for some variable X, then $\mathbf{B}_1 \cap \mathbf{B}_2$ is also a Markov blanket for X. [**Hint:** Appeal to the intersection axiom.]

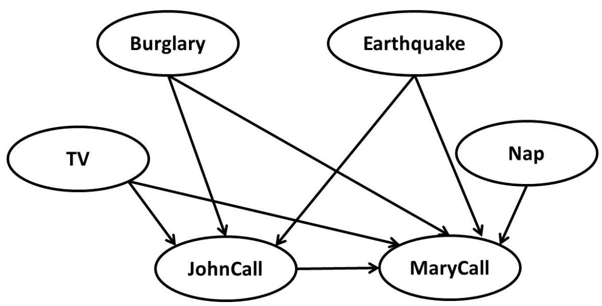
Solution: Let $\mathbf{X}$ be all variables and $\mathbf{Q} = \mathbf{X} - (\{X\} \cup \mathbf{B}_1 \cup \mathbf{B}_2)$. If we assume that $\mathbf{Y}$ contains all common variables between $\mathbf{B}_1$ and $\mathbf{B}_2$ ($\mathbf{Y}$ is possibly empty), we can then write the sets as $\mathbf{B}_1 = \mathbf{Y} \cup \mathbf{Z}$ and $\mathbf{B}_2 = \mathbf{Y} \cup \mathbf{W}$ for some $\mathbf{Z}$ and $\mathbf{W}$. Since $\mathbf{B}_1$ and $\mathbf{B}_2$ are blankets, we have $I_{\Pr}(\mathbf{X}, \mathbf{Y} \cup \mathbf{Z}, \mathbf{W} \cup \mathbf{Q})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Y} \cup \mathbf{W}, \mathbf{Z} \cup \mathbf{Q})$. By decomposition, we have $I_{\Pr}(\mathbf{X}, \mathbf{Y} \cup \mathbf{Z}, \mathbf{W})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Y} \cup \mathbf{W}, \mathbf{Z})$. By Intersection, we have $I_{\Pr}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} \cup \mathbf{W})$. By Decomposition, we have $I_{\Pr}(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. By Contraction, we have $I_{\Pr}(\mathbf{X}, \mathbf{Y}, \mathbf{W} \cup \mathbf{Z} \cup \mathbf{Q})$. Hence, $\mathbf{Y} = \mathbf{B}_1 \cap \mathbf{B}_2$ is a Markov blanket for $\mathbf{X}$.

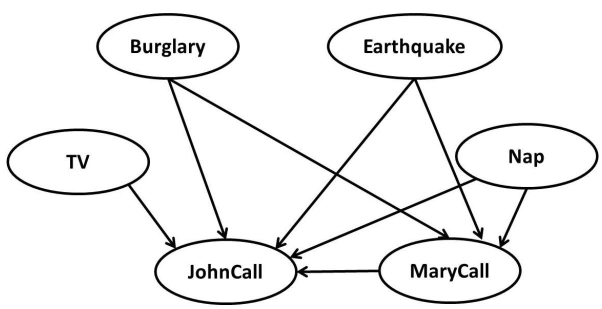
Problem 10: Bayesian networks (Exercise 3.11 from Koller & Friedman)



Consider the Burglary Alarm network given above. Construct a Bayesian network over all the node **except** the Alarm that is a minimal I-map for the marginal distribution over the remaining variables (namely, over B, E, N, T, J, M). Be sure to get all the dependencies from the original network.

Solution: The two possible solutions are:





[1 point each] for the edges from Burglary and Earthquake to JohnCall and MaryCall.

[3 points each] for the edge between MaryCall to JohnCall (or from JohnCall to MaryCall) and the edge between Nap to JohnCall (or from TV to MaryCall).