

Introduction to AI – Probabilistic Reasoning

Fu Dan University FISS Summer 2022

Final Exam

87 Points Total

This exam is open book, open notes. You can read our slides (annotated or plain), read the class online textbook, and use a calculator on your mobile devices or computers. You are not allowed to browse other websites or seek help on social media. No collaboration is allowed.

By signing your name below, you are agreeing that you will never discuss or share any part of this exam with anyone who is not currently taking the exam or is not a member of the course staff. This includes posting any information about this exam on any website or any other social media. Discussing or sharing any aspect of this exam with anyone outside of this room constitutes a violation of the academic integrity policy of the summer program.

To express your understanding with the above policy, please write the phrase below in the box, and then sign your name below:

“I excel with integrity”

--

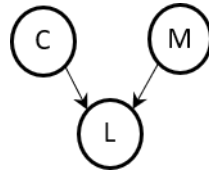
Signature:_____

Name (print):_____

Student ID:_____

1. Probabilistic Reasoning and Noisy-OR [12 points]

Suppose that each morning there are two events that might make me late for my class: car trouble and my children having a meltdown. Let $C \in \{0,1\}$ represent whether or not I have car trouble, $M \in \{0,1\}$ represent whether or not my kids have a meltdown and $L \in \{0,1\}$ represent whether or not I am late for class. The Bayes Net to model this situation is shown below:



- (a) (2 pts) The influence of C and M on L is modeled with a noisy-OR relationship. Fill in the missing values in the noisy-OR conditional probability table for this network. You do NOT need to show your work for part (a).

C	M	$P(L = 1 C, M)$
0	0	
0	1	0.2
1	0	
1	1	0.44

- (b) (4 pts) Order the following values from LEAST to GREATEST going down the page. If two or more values are the same, write them next to each other on the same line. You do not have to use all four lines.

$$P(M = 1)$$

$$P(M = 1|C = 0)$$

$$P(M = 1|C = 0, L = 1)$$

$$P(M = 1|C = 1, L = 1)$$

Least

Greatest

- (c) (6 pts) Let $P(M = 1) = 0.05$ and $P(C = 1) = 0.01$. Compute the value for $P(M = 1|C = 1, L = 1)$ and $P(M = 1|C = 0, L = 1)$ **You do not have to do the final arithmetic steps. You may leave your answers in terms of arithmetic operations on numerical values. SHOW YOUR WORK FOR PART (c).**

2. Conditional Independence [10 points]

(a) (2 pts) Consider the following equality:

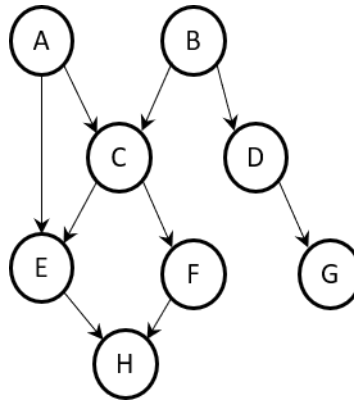
$$P(Q, R|S, T, H, I) = P(Q|S, I)P(R|Q, S, H, I)$$

Which of the following statements is/are implied by this equality? Write the letters of **all** (zero or more) that are implied in the blank. If none are implied write “None”.

- A. Q is conditionally independent from H given $\{S, I\}$
- B. R is conditionally independent from T given $\{Q, S, H, I\}$
- C. S is conditionally independent from T and H given $\{I\}$

Answer: _____

(b) (5 pts) With respect to the BN graph below, state whether each of the statements of conditional independence are true or false.

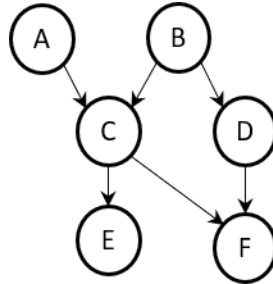


- (i) _____ $P(A, H|C, E) = P(A|C, E)P(H|C, E)$
- (ii) _____ $P(A|B) = P(A)$
- (iii) _____ $P(A, D|H) = P(A|H)P(D|H)$
- (iv) _____ $P(E|C, B, D, G) = P(E|C, B)$
- (v) _____ $P(G, H|C, E, A) = P(G|C, A)P(H|C, E, A)$

(c) (3 pts) Prove that $P(H|C, D) \neq P(H|C)$ using d-separation.

3. General Inference in Bayesian Networks [10 points]

Consider the Bayesian Network given below where all variables are *binary*.



- (a) (2 pts) Using the product rule and marginalization, the posterior probability $P(C = 1|F = 0, E = 1)$ can be expressed as:

$$P(C = 1|F = 0, E = 1) = \frac{P(C = 1, F = 0, E = 1)}{\sum_c P(C = c, F = 0, E = 1)}$$

Express the joint probability $P(C = c, F = 0, E = 1)$ in terms of only the probabilities given in the conditional probability tables in the Bayesian Network.

- (b) (6 pts) Fill in the table with the number of additions and multiplications it takes to compute $P(C = 1, F = 0, E = 1)$, $P(C = 0, F = 0, E = 1)$, and then the total number to compute $P(C = 1|F = 0, E = 1)$ using the brute force (enumeration) expression you wrote in part (a).

Phase of algorithm	# multiplications	# additions	# divisions
Compute $P(C = 1, F = 0, E = 1)$			0
Compute $P(C = 0, F = 0, E = 1)$			0
Combine above values to calculate $P(C = 1 F = 0, E = 1)$	0	1	1
Total			1

- (c) (2 pts) Does the Variable Elimination algorithm using the elimination order E, F, A, D, B use fewer arithmetic operations than the brute-force (enumeration) approach to compute $P(C = 1 | F = 0, E = 1)$ in this network? Briefly explain why or why not.

4. Maximum Likelihood Learning [20 points]

(a) (6 pts) Recall the derivation of the log-likelihood for a Bayesian Network that we did in class, shown below. Briefly justify each step.

- (i) $\mathcal{L} = \log P(\text{data})$
- (ii) $= \log \prod_{t=1}^T P(X_1 = x_1^{(t)}, X_2 = x_2^{(t)}, \dots, X_n = x_n^{(t)})$
- (iii) $= \log \prod_{t=1}^T \prod_{i=1}^n P(X_i = x_i^{(t)} | \text{Parents}(X_i) = \pi_i^{(t)})$
- (iv) $= \sum_{i=1}^n \sum_{t=1}^T \log P(X_i = x_i^{(t)} | \text{Parents}(X_i) = \pi_i^{(t)})$
- (v) $= \sum_{i=1}^n \sum_x \sum_{\pi} \text{count}(X_i = x, \text{Parents}(X_i) = \pi) \log P(X_i = x | \text{Parents}(X_i) = \pi)$

Step	Justification
(i)	
(ii)	
(iii)	Move log in and swap sums
(iv)	

- (b) Suppose that A , B , and C are binary random variables and that we observe $T = 1000$ samples from their joint distribution $P(A, B, C)$ with the following frequencies:

A	B	C	count(A, B, C)
0	0	0	100
0	0	1	50
0	1	0	200
0	1	1	150
1	0	0	50
1	0	1	100
1	1	0	150
1	1	1	200

- (i) (3 pts) From the data in the table above, calculate the Maximum Likelihood (ML) estimates for the following probabilities.

$$P_{ML}(A = 1), P_{ML}(B = 1|A = 0), P_{ML}(B = 1|A = 1)$$

$$P_{ML}(A = 1) = \underline{\hspace{2cm}}$$

$$P_{ML}(B = 1|A = 0) = \underline{\hspace{2cm}}$$

$$P_{ML}(B = 1|A = 1) = \underline{\hspace{2cm}}$$

- (ii) (2 pt) From your answers to part (i), does the data support the claim that A and B are marginally independent? Briefly justify your answer.

- (iii) (4 pts) From the data in the table above, calculate the Maximum Likelihood (ML) estimates for the following probabilities.

$$P_{ML}(C = 1|A = 0, B = 0), P_{ML}(C = 1|A = 0, B = 1), P_{ML}(C = 1|A = 1, B = 0), \\ P_{ML}(C = 1|A = 1, B = 1)$$

$$P_{ML}(C = 1|A = 0, B = 0) = \underline{\hspace{2cm}}$$

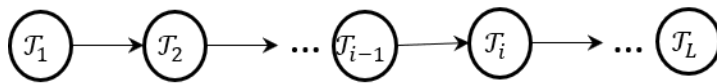
$$P_{ML}(C = 1|A = 0, B = 1) = \underline{\hspace{2cm}}$$

$$P_{ML}(C = 1|A = 1, B = 0) = \underline{\hspace{2cm}}$$

$$P_{ML}(C = 1|A = 1, B = 1) = \underline{\hspace{2cm}}$$

- (iv) (1 pt) From your answer to part (iv), does the data support the claim that C is conditionally independent from B given A? Briefly justify your answer.

- (v) (4 pts) Finally, consider a first-order (bigram) Markov Model defined over the 3-letter alphabet $\mathcal{A} = \{x, y, z\}$. Let $\mathcal{T}_i \in \mathcal{A}$ denote the i^{th} token in a sequence. The model is shown below:



List all of the parameters of this model (i.e. the parameters whose value you need to specify in order to use the model). You can use a shorthand notation as long as it's clear what all the parameters are.

5. Implementing Hangman [6 points]

Recall from your implementation of the Hangman game in HW2 that the probability of evidence E given word w is:

$$P(E|W = w) = \begin{cases} 1 & \text{if } E \text{ is consistent with } w \\ 0 & \text{otherwise} \end{cases}$$

where E contains the letters guessed so far along with the places (if any) they appear, and W represents a word in the dictionary. E is consistent with w if all the letters in E that have places listed appear in w in those places, and all the letters in E that do not have places listed do not appear in w .

Assume that E is represented as a map named `evidence` where the keys are letters that have been guessed and the values are lists of integers (which may be empty) containing all the places that the letter appears. For example, for the following state of the game:

Correctly guessed: D - - I D

Incorrectly guessed: {O, M}

`evidence['D']` is [1, 5],

`evidence['I']` is [4],

`evidence['O']` and `evidence['M']` are [],

and no other keys appear in the map.

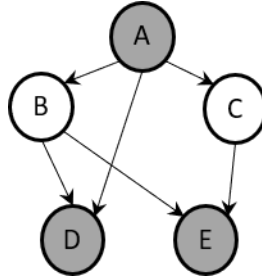
Complete the pseudocode to compute and return $P(E|W = w)$. You do NOT have to use correct syntax. Precise English statements are acceptable.

```
evidenceGivenWord(evidence, w):  
    for each letter which is a key in evidence:  
        let places = evidence[letter] // places is a list
```

6. EM Algorithm (18 pts)

For all parts of this problem except part (d) show your work, justify each step, and simplify your answers as much as possible. In part (d) simply write your final simplified answer.

Consider the belief network shown below, with CPTs for $P(A)$, $P(B|A)$, $P(C|A)$, $P(D|A, B)$, $P(E|B, C)$. Suppose the values of the nodes B and C are unobserved (hidden), while the values of the shaded nodes A, D, and E are observed.



(a) Posterior probability (4 pts)

Compute the posterior probability $P(B, C|A, D, E)$ over the hidden nodes in terms of the CPTs of this belief network. **Show your work, justify each step, and simplify your answer as much as possible.**

(b) Posterior probability (4 pts)

Compute the posterior probabilities $P(B|A, D, E)$ and $P(C|A, D, E)$ in terms of the posterior probability $P(B, C|A, D, E)$ from part (a). **Show your work, justify each step and simplify your answer as much as possible.**

(c) Log-likelihood (4pts)

Consider a data set of T partially labeled examples $\{a_t, d_t, e_t\}_{t=1}^T$ over the nodes A, D and E of the network. The log-likelihood of the data set is given by:

$$\mathcal{L} = \sum_t \log P(A = a_t, D = d_t, E = e_t)$$

Compute this log-likelihood in terms of the CPTs of the belief network above. **Show your work, justify each step and simplify your answer as much as possible.**

(d) EM Algorithm (6 pts)

Consider the EM algorithm that updates the CPTs to maximize the log-likelihood of the data set in part (c). Complete the numerator and denominator in the expressions for the EM updates shown below.

Suggested notation. Use $P(b, c|a_t, d_t, e_t)$, $P(b|a_t, d_t, e_t)$, $P(c|a_t, d_t, e_t)$ for the posterior probabilities computed in parts (a) and (b). Also use indicator functions such as $I(a, a_t)$ where

$$I(a, a_t) = \begin{cases} 1 & \text{if } a = a_t \\ 0 & \text{if } a \neq a_t \end{cases}$$

Place your answers in the corresponding blanks. **You do NOT need to show your work.**

$$P(B = b|A = a) \leftarrow \underline{\hspace{10cm}}$$

$$P(D = d|A = a, B = b) \leftarrow \underline{\hspace{10cm}}$$

$$P(E = e|B = b, C = c) \leftarrow \underline{\hspace{10cm}}$$

7. Hidden Markov Models (8 pts)

For all parts of this problem, show your work and justify each step in your derivation.

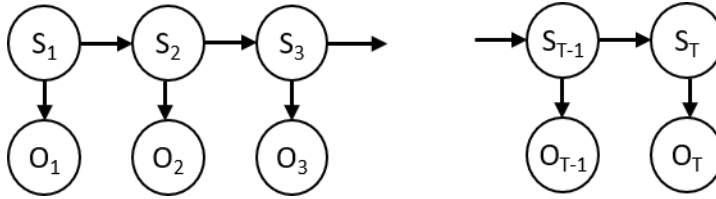
Consider a discrete hidden Markov Model (HMM) with the belief network shown below. Let $S_t \in \{1, 2, \dots, n\}$ denote the hidden state of the HMM at time t , and let $O_t \in \{1, 2, \dots, m\}$ denote the observation at time t . As in class, let

$$a_{ij} = P(S_t = j | S_{t-1} = i)$$

$$b_{ik} = P(O_t = k | S_t = i)$$

$$\pi_i = P(S_1 = i)$$

denote the transition matrix, the emission matrix and the initial state distribution, respectively.



(a) Initial condition (4 pts)

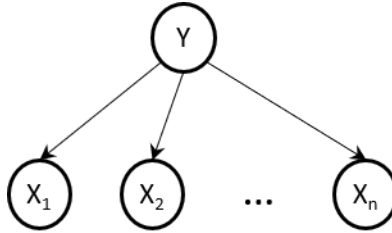
Compute the probability $P(O_1 = o_1, S_1 = i)$ in terms of the parameters $\{\pi_i, a_{ij}, b_{ik}\}$ of the HMM. **Show your work and justify each step.**

(b) Forward algorithm (4 pts)

Consider the probabilities $\alpha_{it} = P(o_1, o_2, \dots, o_t, S_t = i)$ for an observed sequence $\{o_t\}_{t=1}^T$. Derive an efficient recursion to compute these probabilities at time $t + 1$ from those at time t and the parameters $\{\pi_i, a_{ij}, b_{ik}\}$ of the HMM. **Show your work and justify each step.**

8. Naïve Bayes (3 pts)

Consider the belief network shown below for binary random variables $X_i \in \{0,1\}$ and $Y \in \{0,1\}$. Compute the posterior probability $P(Y = 1|X_1, X_2, \dots, X_n)$ in terms of the conditional probability tables $P(Y)$ and $P(X_i|Y)$.



$P(Y = 1|X_1, X_2, \dots, X_n)$. Show your steps