UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

CS440/ECE448 Artificial Intelligence

Practice Exam 2

Spring 2022

Exam 2 will be April 4, 2022

Your Name: ____________________

Your NetID: ____________________

Instructions

- Please write your name and NetID on the top of every page.
- This will be a CLOSED BOOK exam. You will be permitted to bring one 8.5x11 page of handwritten notes (front & back).
- Calculators are not permitted. You need not simplify explicit numerical expressions.

Question 1 ***(0 points)***

Discuss the relative strengths and weaknesses of breadth-first search vs. depth-first search for AI problems.

Solution:

	BFS	DFS
Strength	Solution is guaranteed to be optimal. BFS is complete.	Can find goal faster than BFS if there are multiple solutions. Linear memory requirement $(O(bm))$.
Weakness	Exponential $(O(b^d))$ space complexity.	Solution not guaranteed to be optimal. Not complete. Computation is exponential in longest path $(O(b^m))$, rather than shortest path $(O(b^d))$.

Question 2 ***(0 points)***

In the tree search formulation, why do we restrict step costs to be non-negative?

Solution: If the cost around any loop is negative, then the lowest-cost path is to take that loop an infinite number of times.

Question 3 ***(0 points)***

What is the distinction between a world state and a search tree node?

Solution: A world state contains enough information to know (1) whether or not you've reached the goal, (2) what actions can be performed, (3) what will be the result of each action. A search tree node contains a pointer to the world state, plus a pointer to the parent node.

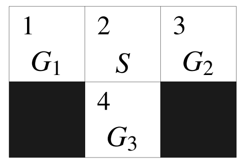
Question 4 ***(0 points)***

How do we avoid repeated states during tree search?

Solution: By keeping a set of "explored states." If expanding a search node results in a state that has already been explored, we don't add it to the frontier.

Question 5 ***(0 points)***

Imagine a maze with only four possible positions, numbered 1 through 4 in the following diagram. Position 2 is the start position (denoted S in the diagram below), while positions 1, 3, and 4 each contain a goal (denoted as G_1, G_2, and G_3 in the diagram below). Search terminates when the agent finds a path that reaches all three goals, using the smallest possible number of steps.



(a) Define a notation for the state of this agent. How many distinct non-terminal states are there?

> **Solution:** The state can be defined by a pair of variables: (P,G) where $P \in \{1,\ldots,4\}$ specifies the current position, $G \in \{\varnothing, G_1, G_2, G_3, G_1G_2, G_1G_3, G_2G_3\}$ specifies which of the goals have been reached. There is only one position (2) that can be reached without touching any goal. After touching G_1, there are two positions that can be reached without touching another goal (1 and 2). After touching G_1 and G_2, there are three positions that can be reached without touching G_3 (1, 2, and 3). Generalizing, there are a total of $1+3\times2+3\times3 = 16$ non-terminal states.

(b) Draw a search tree out to a depth of 3 moves, including repeated states. Circle repeated states.

> **Solution:** After the first move, possible states are $(1,G_1)$,$(3,G_2)$, and $(4,G_3)$. After the second move, possible states are $(2,G_1)$, $(2,G_2)$, and $(2,G_3)$. After the third move, possible states are $(1,G_1)$, $(1,G_1G_2)$, $(1,G_1G_3)$, $(3,G_2)$, $(3,G_1G_2)$, $(3,G_2G_3)$, $(4,G_3)$, $(4,G_1G_3)$, and $(4,G_2G_3)$. Of these, the states $(1,G_1)$, $(3,G_2)$, and $(4,G_3)$ in the last row are repeated states.

(c) For A* search, one possible heuristic, h_1, is the Manhattan distance from the agent to the nearest goal that has not yet been reached. Prove that h_1 is consistent.

Solution: From any node m, Manhattan distance to the nearest goal is always either $h_1[m]=1$ or $h_1[m]=2$. If $h_1[m]=2$, then any step we take will move us to a node n such that $h_1[n]=1$. If $h_1[m]=1$, then it is possible to move away from the goal (to position n such that $h_1[n]=2$), or it is possible to move toward the goal (in which case we reach the goal, and so the definition of "nearest goal that has not been reached" changes to one of the other goals, and again we have $h_1[n]=2$). So the change in heuristic is always $h_1[m]-h_1[n]\in\{-1,1\}$.

The distance traveled from any node m to its neighbor n is always one step. If n is closer to completing the maze than m, then the distance from m to the goal, $d[m]$, minus the distance from n to the goal, $d[n]$, is one: $d[m]-d[n]=1$, whereas $h_1[m]-h_1[n]\in\{-1,1\}$, so $d[m]-d[n]\geq h_1[m]-h_1[n]$.

(d) Another possible heuristic is based on the Manhattan distance $M[n,g]$ between two positions, and is given by

$$h_2[n]=M[G_1,G_2]+M[G_2,G_3]+M[G_3,G_1]$$

that is, h_2 is the sum of the Manhattan distances from goal 1 to goal 2, then to goal 3, then back to goal 1. Prove that h_2 is not admissible.

Solution: Notice that $h_2[n]=6$ for every node n, so we only have to find a counter-example for which the total cost of the best path is $d[n]<6$. But that's easy: the starting node S has a cost of $d[S]=5<h_2[S]$, so h_2 is not admissible.

(e) Prove that $h_2[n]$ is dominant to $h_1[n]$.

Solution: $h_1[n]\in\{1,2\}$, whereas $h_2[n]=6$ always, so $h_2[n]\geq h_1[n]$.

Question 6 ***(0 points)***

Consider the following maze. There are 11 possible positions, numbered 1 through 11. The agent starts in the position marked S (position number 3). From any position, there are from one to four possible moves, depending on position: Left, Right, Up, and/or Down. The agent's goal is to find the shortest path that will touch both of the goals (G_1 and G_2).

1	2 G_1	3 S	4
5		6 G_2	7
8	9	10	11

(a) Define a notation for the state of this agent. How many distinct non-terminal states are there?

Solution: The state can be defined by a pair of variables: (P,G) where $P \in \{1,\dots,11\}$ specifies the current position, $G \in \{\varnothing, G_1, G_2\}$ specifies which of the goals have been reached. There are nine values of P that can be reached without touching either goal, ten that can be reached without touching G_1, and ten that can be reached without touching G_2, so the total number of non-terminal states is $9+10+10=29$.

(b) Draw a search tree out to a depth of 2 moves, including repeated states. Circle repeated states.

Solution: After the first move, possible states are $(2,G_1)$,$(6,G_2)$, and $(4,\varnothing)$. After the second move, possible states are $(1,G_1)$, $(3,G_1)$, $(3,G_2)$, $(10,G_2)$, $(7,G_2)$, $(3,\varnothing)$ (a repeated state), and $(7,\varnothing)$.

(c) For A* search, one possible heuristic, h_1, is the number of goals not yet reached. Prove that h_1 is consistent.

Solution: The heuristic difference between two neighboring states is either $h_1[n_1]-h_1[n_2]=1$ (if they differ in the number of goals remaining) or $h_1[n_2]-h_1[n_1]=0$ (if they have the same number of goals remaining. The distance is always $d[n_1,n_2]=1$. So $h_1[n_1]-h_1[n_2]\le d[n_1,n_2]$.

(d) Another possible heuristic is based on the Manhattan distance $M[n,g]$ between two positions, and is given by

$$h_2[n]=M[n,G_1]+M[G_1,G_2]$$

that is, h_2 is the sum of the Manhattan distance from the current position to G_1, plus the Manhattan distance from G_1 to G_2. Prove that h_2 is not admissible.

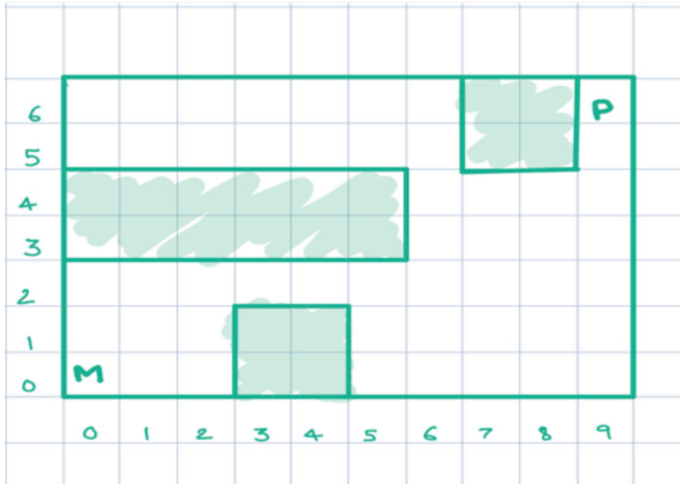
Solution: Proof by counter-example: for example, consider the state $n=(7,\varnothing)$. From this state, the shortest solution goes left, then up, then left: $d[n]=3$ steps. The heuristic, however, is $h_2[n]=M[n,G_1]+M[G_1,G_2]=3+2=5$, so $h[n]>d[n]$.

(e) Prove that $h_2[n]$ is dominant to $h_1[n]$.

Solution: $h_1[n]\in\{0,1,2\}$. The Manhattan distance $M[G_1,G_2]=2$, therefore $h_2[n]\ge 2\ge h_1[n]$.

Question 7 ***(0 points)***

Refer to the maze shown below. Here, 'M' represents Mario, 'P' represents Peach, and the goal of the game is to get Mario and Peach to find each other. In each move, both Mario and Peach take turns. For example, one move would consist of Peach moving a block to the bottom from her current position, and Mario moving one block to the left from his current position. Standing still is also an option.



(a) Describe state and action representations for this problem.

Solution:

- State = (x_M, y_M) position of Mario, (x_P, y_P) position of Peach.
- Action = Mario moves up,down,left,right, or stationary, Peach moves up,down,left,right, or stationary.

(b) What is the branching factor of the search tree?

Solution: 25

(c) What is the size of the state space?

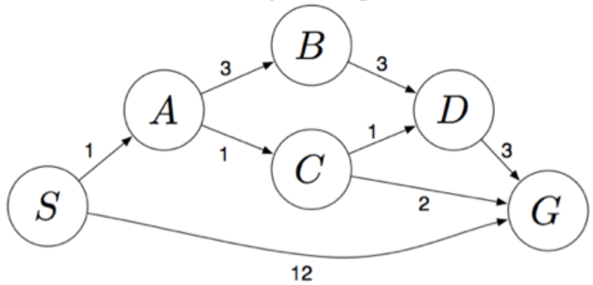
Solution: There are 50 possible locations for both Mario and Peach (if we assume that they can occupy the same square), for a total state space of 2500.

(d) Describe an admissible heuristic for this problem.

Solution: The heuristic $h(n) = 0$ is always admissible, but would receive at most a little bit of partial credit, because it's trivial. A more useful heuristic would measure the distance between Mario and Peach, e.g., $0.5 * (|x_M - x_P| + |y_M - y_P|)$.

Question 8 ***(0 points)***

Consider the search problem with the following state space:



S denotes the start state, G denotes the goal state, and step costs are written next to each arc. Assume that ties are broken alphabetically (i.e., if there are two states with equal priority on the frontier, the state that comes first alphabetically should be visited first).

(a) What path would BFS return for this problem?

> **Solution:** SG

(b) What path would DFS return for this problem?

> **Solution:** SABDG

(c) What path would UCS return for this problem?

Solution: SACG

(d) Consider the heuristics for this problem shown in the table below.

State	h_1	h_2
S	5	4
A	3	2
B	6	6
C	2	1
D	3	3
G	0	0

i. Is h1 admissible? Is it consistent?

Solution: Neither admissible nor consistent.

ii. Is h2 admissible? Is it consistent?

Solution: Admissible but not consistent.

Question 9 ***(0 points)***

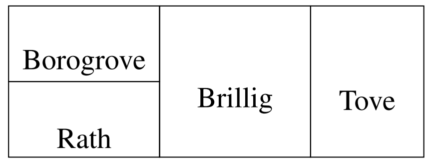
Explain why it is a good heuristic to choose the variable that is *most* constrained but the value that is *least* constraining in a CSP search.

Solution:

- When we choose a variable, we are not eliminating any possible solutions, we are only deciding on the order in which variables will be considered. It makes sense to choose the order that minimizes complexity.
- When we choose an assignment, we are eliminating possible solutions; if we're wrong, we'll have to back-track. Therefore it makes sense to choose the assignment that eliminates as few solutions as possible, to minimize the chance of back-tracking.

Question 10 ***(0 points)***

The figure below shows the map of a fictional country, with four provinces: Borogrove, Rath, Brillig, and Tove. The "map coloring problem" requires you to color each province red, blue, or green, without using the same color for any two neighboring provinces.



Remember that, in choosing an evaluation sequence for the depth-first search in a constraint satisfaction problem, three heuristics are often useful: LRV (least remaining values), MCV (most constraining variable), and LCV (least constraining value).

(a) According to the LRV, MCV, and LCV heuristics, which region should be colored first, and why?

> **Solution:** Brillig. MCV (3 constraints, versus 2 for Borogrove and Rath, 1 for Tove).

(b) Suppose Borogrove has already been colored red, all others are not colored yet. Would it make more sense to color Rath next, or Tove? Why?

> **Solution:** Rath. LRV (2 remaining values, versus 3 for Tove).

(c) Suppose Borogrove has already been colored red, all others are not colored yet. Now Tove is to be colored. What color should it be, and why?

> **Solution:** Red. LCV (coloring Tove red does not impose any further constraints on any other region. Coloring it blue would would impose the constraint that Brillig not be blue; likewise green).

Question 11 ***(0 points)***

For each of the following problems, determine whether an algorithm to optimally solve the problem requires worst-case computation time that is polynomial or exponential in the parameters d and m (assuming that P $\neq$ NP).

(a) A map has d regions. Colors have been applied to all d regions, drawing from a set of m possible colors. Your algorithm needs to decide whether or not any two adjacent regions have the same color.

> **Solution:** Polynomial in d.

(b) b. A map has d regions. Your algorithm needs to assign colors to all d regions, drawing colors from a set of m possible colors, in order to guarantee that no two adjacent regions have the same color.

> **Solution:** Exponential in d.

(c) Your algorithm needs to find its way out of a maze drawn on a d-by-d grid.

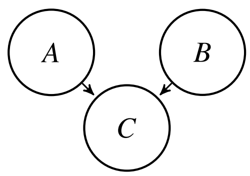
> **Solution:** Polynomial in d.

(d) Your algorithm needs to find the shortest path in a d-by-d maze while hitting m waypoints (equivalent to dots in MP1 part 1.2).

> **Solution:** Exponential in m.

Question 12 ***(0 points)***

Consider the following Bayes network (all variables are binary):



$P(A) = 0.4, P(B) = 0.1$

A, B	$P(C \mid A, B)$
False,False	0.7
False,True	0.7
True,False	0.1
True,True	0.9

(a) What is $P(C)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(C) &= P(\neg A, \neg B, C) + P(\neg A, B, C) + P(A, \neg B, C) + P(A, B, C) \\ &= (0.6)(0.9)(0.7) + (0.6)(0.1)(0.7) + (0.4)(0.9)(0.1) + (0.4)(0.1)(0.9) \end{aligned}$$

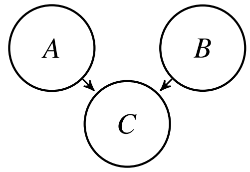
(b) What is $P(A \mid B = True, C = True)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(A \mid B, C) &= \frac{P(A, B, C)}{P(A, B, C) + P(\neg A, B, C)} \\ &= \frac{(0.4)(0.1)(0.9)}{(0.4)(0.1)(0.9) + (0.6)(0.1)(0.7)} \end{aligned}$$

Question 13 ***(0 points)***

Consider the following Bayes network (all variables are binary):



You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	True	False	False
2	False	False	True
3	True	True	False
4	False	False	False

(a) Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

Solution: $P(A) = 2/4$, $P(B) = 1/4$, and

A,B	$P(C\|A,B)$
F, F	$1/2$
F, T	UNKNOWN
T, F	$0/1$
T, T	$0/1$

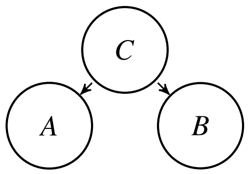
(b) Use the table of data, but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution: $P(A) = 3/6$, $P(B) = 1/3$, and

A,B	$P(C\|A,B)$
F, F	$2/4$
F, T	$1/2$
T, F	$1/3$
T, T	$1/3$

Question 14 ***(0 points)***

Consider the following Bayes network (all variables are binary):



$$P(C) = 0.1$$

C	$P(A\|C)$	$P(B\|C)$
False	0.8	0.7
True	0.4	0.7

(a) What is $P(A)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(A) &= P(\neg C, A) + P(C, A) \\ &= (0.9)(0.8) + (0.1)(0.4) \end{aligned}$$

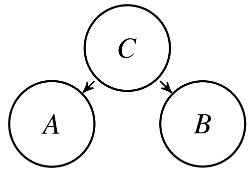
(b) What is $P(C|A = True, B = True)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(C|A,B) &= \frac{P(A,B,C)}{P(A,B,C) + P(A,B,\neg C)} \\ &= \frac{(0.1)(0.4)(0.7)}{(0.1)(0.4)(0.7) + (0.9)(0.8)(0.7)} \end{aligned}$$

Question 15 ***(0 points)***

Consider the following Bayes network (all variables are binary):



You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	False	True	False
2	True	True	False
3	False	False	True
4	False	False	True

(a) Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

Solution: $P(C) = 2/4$, and

C	$P(A\|C)$	$P(B\|C)$
F	$1/2$	$2/2$
T	$0/2$	$0/2$

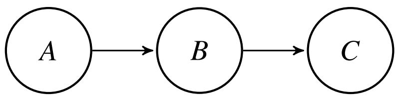
(b) Use the table of data, but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution: $P(C) = 3/6$, and

C	$P(A\|C)$	$P(B\|C)$
F	$2/4$	$3/4$
T	$1/4$	$1/4$

Question 16 ***(0 points)***

Consider the following Bayes network (all variables are binary):



$P(A) = 0.8$

A	$P(B\|A)$
False	0.7
True	0.3

B	$P(C\|B)$
False	0.5
True	0.7

(a) What is $P(C)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(C) &= P(\neg A, \neg B, C) + P(\neg A, B, C) + P(A, \neg B, C) + P(A, B, C) \\ &= (0.2)(0.3)(0.5) + (0.2)(0.7)(0.7) + (0.8)(0.7)(0.5) + (0.8)(0.3)(0.7) \end{aligned}$$

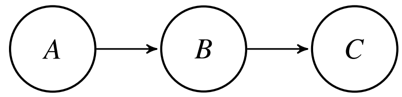
(b) What is $P(A|B = True, C = True)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(A|B,C) &= \frac{P(A,B,C)}{P(A,B,C) + P(\neg A,B,C)} \\ &= \frac{(0.8)(0.3)(0.7)}{(0.8)(0.3)(0.7) + (0.2)(0.7)(0.7)} \end{aligned}$$

Question 17 ***(0 points)***

Consider the following Bayes network (all variables are binary):



You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	True	False	False
2	False	False	True
3	True	True	False
4	False	False	False

(a) Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

Solution:

$$P(A) = 2/4$$

A	$P(B\|A)$
False	0/2
True	1/2

B	$P(C\|B)$
False	1/3
True	0/1

(b) Use the table of data, but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution: $P(A) = 3/6$, $P(B) = 1/3$, and

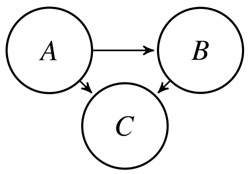
$$P(A) = 3/6$$

A	$P(B\|A)$
False	1/4
True	2/4

B	$P(C\|B)$
False	2/5
True	1/3

Question 18 ***(0 points)***

Consider the following Bayes network (all variables are binary):



$P(A) = 0.4$

A	$P(B\|A)$
False	0.1
True	0.2

A,B	$P(C\|A,B)$
False,False	0.9
False,True	0.3
True,False	0.7
True,True	0.5

(a) What is $P(C)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(C) &= P(\neg A, \neg B, C) + P(\neg A, B, C) + P(A, \neg B, C) + P(A, B, C) \\ &= (0.6)(0.9)(0.9) + (0.6)(0.1)(0.3) + (0.4)(0.8)(0.7) + (0.4)(0.2)(0.5) \end{aligned}$$

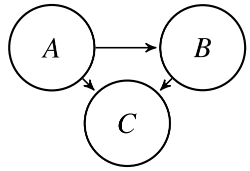
(b) What is $P(A|B = True, C = True)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(A|B,C) &= \frac{P(A,B,C)}{P(A,B,C) + P(\neg A,B,C)} \\ &= \frac{(0.4)(0.2)(0.5)}{(0.4)(0.2)(0.5) + (0.6)(0.1)(0.3)} \end{aligned}$$

Question 19 ***(0 points)***

Consider the following Bayes network (all variables are binary):



You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	True	True	False
2	False	True	True
3	False	True	False
4	False	False	True

(a) Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

Solution:

$$\begin{aligned} P(A) &= 1/4 \\ P(B|\neg A) &= 2/3 \\ P(B|A) &= 1/1 \\ P(C|\neg A, \neg B) &= 1/1 \\ P(C|\neg A, B) &= 1/2 \\ P(C|A, \neg B) &= UNKNOWN \\ P(C|A, B) &= 0/1 \end{aligned}$$

(b) Use the table of data, but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution:

$$\begin{aligned} P(A) &= 2/6 \\ P(B|\neg A) &= 3/5 \\ P(B|A) &= 2/3 \\ P(C|\neg A, \neg B) &= 2/3 \\ P(C|\neg A, B) &= 2/4 \\ P(C|A, \neg B) &= 1/2 \\ P(C|A, B) &= 1/3 \end{aligned}$$

Question 20 ***(0 points)***

We have a bag of three biased coins, a, b, and c, with probabilities of coming up heads of 20%, 60%, and 80%, respectively. One coin is drawn randomly from the bag (with equal likelihood of drawing each of the three coins), and then the coin is flipped three times to generate the outcomes X1, X2, and X3.

(a) Draw the Bayesian network corresponding to this setup and define the necessary conditional probability tables (CPTs).

Solution: You need an intermediate variable, $C \in \{a, b, c\}$, to specify which coin is drawn, then the graph is

$$\begin{tikzcd} & C \arrow[dl] \arrow[d] \arrow[dr] & \\ X_1 & X_2 & X_3 \end{tikzcd}$$

and the CPTs are

C	$P(C)$	$P(X_1 = H\|C)$	$P(X_2 = H\|C)$	$P(X_3 = H\|C)$
a	1/3	0.2	0.2	0.2
b	1/3	0.6	0.6	0.6
c	1/3	0.8	0.8	0.8

(b) Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

Solution:

$$P(C = a, HHT) = (0.2)(0.2)(0.8)/3 = 32/3000$$
$$P(C = b, HHT) = (0.6)(0.6)(0.4)/3 = 144/3000$$
$$P(C = c, HHT) = (0.8)(0.8)(0.2)/3 = 128/3000$$

The maximum-posterior-probability event is also the maximum-joint-probability event, which is the event $C = b$.

Question 21 ***(0 points)***

Two astronomers in different parts of the world make measurements M_1 and M_2 of the number of stars N in some small region of the sky, using their telescopes. Under normal circumstances, this experiment has three possible outcomes: either the measurement is correct (with probability $1-2e-f$), or the measurement overcounts the stars by one (one star too high) with probability e, or the measurement undercounts the stars by one (one star too low) with probability e. There is also the possibility, however, of a large measurement error in either telescope (events F_1 and F_2, respectively, each with probability f), in which case the measured number will be **at least** three stars too low (regardless of whether the scientist makes a small error or not), or, if N is less than 3, fail to detect any stars at all.

(a) Draw a Bayesian network for this problem.

Solution: A solution must include the variables N, M_1, M_2 with the dependencies shown below. The variables F_1, F_2 are optional:

$$
\begin{tikzcd}
F_1 \arrow[dr] & & N \arrow[dl] \arrow[dr] & & F_2 \arrow[dl] \\
& M_1 & & M_2 &
\end{tikzcd}
$$

(b) Write out a conditional distribution for $P(M_1|N)$ for the case where $N \in \{1,2,3\}$ and $M_1 \in \{0,1,2,3,4\}$. Each entry in the conditional distribution table should be expressed as a function of the parameters e and/or f.

Solution:

	M_1				
N	0	1	2	3	4
1	$e+f$	$1-2e-f$	e	0	0
2	f	e	$1-2e-f$	e	0
3	f	0	e	$1-2e-f$	e

(c) Suppose $M_1 = 1$ and $M_2 = 3$. What are the possible numbers of stars if you assume no prior constraint on the values of N?

Solution: $N = 2$ is possible, if both made small mistakes. $N = 4$ is possible, if M_2 made a small and M_1 a big mistake. $N \geq 6$ is possible, if both M_1 and M_2 made big mistakes.

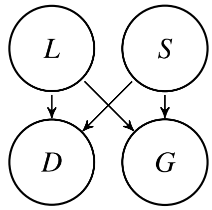
(d) What is the most likely number of stars, given the observations $M_1 = 1, M_2 = 3$? Explain how to compute this, or if it is not possible to compute, explain what additional information is needed and how it would affect the result.

Solution: We need to find the value of N that maximizes $P(N, M_1 = 1, M_2 = 3)$. We have that $P(N = 2, M_1 = 1, M_2 = 3) = P(N = 2)e^2$. We know that $P(N = 4, M_1 = 1, M_2 = 3) \leq P(N = 4)fe$; we don't know exactly how much it is, because we don't know $P(M_1 = 1|N = 4)$, but we know that $P(M_1 = 1|N = 4) \leq f$. So if $P(N = 2)e > P(N = 4)f$, $N = 2$ is the most probable value. If $P(N = 2)e \leq P(N = 4)f$, then it depends on the way in which big errors are distributed among the various values that are "at least three stars" too small.

Question 22 ***(0 points)***

Maria likes ducks and geese. She notices that when she leaves the heat lamp on (in her back yard), she is likely to see ducks and geese. When the heat lamp is off, she sees ducks and geese in the summer, but not in the winter.

(a) The following Bayes net summarizes Maria's model, where the binary variables D,G,L, and S denote the presence of ducks, geese, heat lamp, and summer, respectively:



On eight randomly selected days throughout the year, Maria makes the observations shown in the table below:

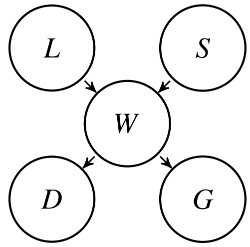
day	D	G	L	S	day	D	G	L	S
1	0	1	1	0	5	1	0	0	1
2	1	0	1	0	6	1	0	1	1
3	0	0	0	0	7	0	1	1	1
4	0	0	0	0	8	0	1	0	1

Write the maximum-likelihood conditional probability tables for D, G, L and S.

Solution: We have that $P(S) = 0.5$, $P(L) = 0.5$, and

S	L	$P(D\|S,L)$	$P(G\|S,L)$
0	0	0	0
0	1	0.5	0.5
1	0	0.5	0.5
1	1	0.5	0.5

(b) Maria speculates that ducks and geese don't really care whether the lamp is lit or not, they only care whether or not the temperature in her yard is warm. She defines a binary random variable, W, which is 1 when her back yard is warm, and she proposes the following revised Bayes net:



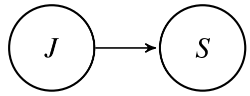
She forgot to measure the temperature in her back yard, so W is a hidden variable. Her initial guess is that $P(D|W) = \frac{2}{3}$, $P(D|\neg W) = \frac{1}{3}$, $P(G|W) = \frac{2}{3}$, $P(G|\neg W) = \frac{1}{3}$, $P(W|L \wedge S) = \frac{2}{3}$, $P(W|\neg(L \wedge S)) = \frac{1}{3}$. Find the posterior probability $P(W|\text{day})$ for each of the 8 days, day $\in \{1, \ldots, 8\}$, whose observations are shown in the Table in part (a).

Solution:

day	1	2	3	4	5	6	7	8
$P(W\|\text{day})$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$

Question 23 ***(0 points)***

Suppose you have a Bayes net with two binary variables, Jahangir (J) and Shahjahan (S):



This network has three trainable parameters: $P(J) = a$, $P(S|J) = b$, and $P(S|\neg J) = c$. Suppose you have a training dataset in which S is observed, but J is hidden. Specifically, there are N training tokens for which $S = \text{True}$, and M training tokens for which $S = \text{False}$. Given current estimates of a, b, and c, you want to use the EM algorithm to find improved estimates $\hat{a}$, $\hat{b}$, and $\hat{c}$.

(a) Find the following expected counts, in terms of M, N, a, b, and c:

$$E[\#\text{ times } J \text{ True}] =$$
$$E[\#\text{ times } J \text{ and } S \text{ True}] =$$
$$E[\#\text{ times } J \text{ True and } S \text{ False}] =$$

Solution:

$$E[\#\text{ times } J \text{ True}] = \frac{abN}{ab+(1-a)c} + \frac{a(1-b)M}{a(1-b)+(1-a)(1-c)}$$
$$E[\#\text{ times } J \text{ and } S \text{ True}] = \frac{abN}{ab+(1-a)c}$$
$$E[\#\text{ times } J \text{ True and } S \text{ False}] = \frac{a(1-b)M}{a(1-b)+(1-a)(1-c)}$$

(b) Find re-estimated values $\hat{a}$, $\hat{b}$, and $\hat{c}$ in terms of M, N, $E[\#\text{ times } J \text{ True}]$, $E[\#\text{ times } J \text{ and } S \text{ True}]$, and $E[\#\text{ times } J \text{ True and } S \text{ False}]$.

Solution:

$$\hat{a} = \frac{E[\#\text{ times } J \text{ True}]}{M+N}$$
$$\hat{b} = \frac{E[\#\text{ times } J \text{ and } S \text{ True}]}{E[\#\text{ times } J \text{ True}]}$$
$$\hat{c} = \frac{N - E[\#\text{ times } J \text{ and } S \text{ True}]}{M+N-E[\#\text{ times } J \text{ True}]}$$

Question 24 ***(0 points)***

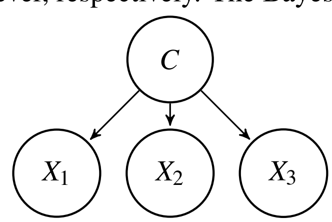
Consider the data points in Table 1, representing a set of seven patients with up to three different symptoms. We want to use the Naïve Bayes assumption to diagnose whether a person has the flu based on the symptoms.

Sore Throat	Stomachache	Fever	Flu
No	No	No	No
No	No	Yes	Yes
No	Yes	No	No
Yes	No	No	No
Yes	No	Yes	Yes
Yes	Yes	No	Yes
Yes	Yes	Yes	No

Table 1: Symptoms of seven patients, three of whom had the flu.

(a) Define random variables, and show the structure of the Bayes network representing a Naïve Bayes classifier for the flu, using the variables shown in Table 1.

Solution: The binary variables could be called F, T, S, and E, representing the presence of flu, sore throat, stomach ache, and fever, respectively. The Bayes net is then



(b) Calculate the maximum likelihood conditional probability tables.

Solution:

F	$P(F)$	$P(T\|F)$	$P(S\|F)$	$P(E\|F)$
0	$4/7$	$1/2$	$1/2$	$1/4$
1	$3/7$	$2/3$	$1/3$	$2/3$

(c) If a person has stomachache and fever, but no sore throat, what is the probability of him or her having the flu (according to the conditional probability tables you calculated in part (b))?

Solution:

$$
\begin{aligned}
P(F|\neg T,S,E) &= \frac{P(\neg T,S,E,F)}{P(\neg T,S,E)} \\
&= \frac{P(F,\neg T,S,E)}{P(F,\neg T,S,E)+P(\neg F,\neg T,S,E)} \\
&= \frac{(3/7)(1/3)(1/3)(2/3)}{(3/7)(1/3)(1/3)(2/3)+(4/7)(1/2)(1/2)(1/4)} \\
&= \frac{8}{17}
\end{aligned}
$$

Question 25 ***(7 points)***

There is a lion in a cage in the dungeons under Castle Rock.

- The zookeeper goes on vacation with a probability of P.
- If the zookeeper is on vacation, the lion doesn't get fed. If not, the lion gets fed with probability Q, and goes hungry with probability $1-Q$.
- If the lion has not been fed, and you try to pet it, then it will bite your hand with probability R. If it has been fed, it will only bite you with probability S.

(a) (2 points) Draw a Bayes network with three random variables: $Z=1$ if the zookeeper is on vacation, $F=1$ if the lion gets fed today, $B=1$ if it will bite the hand of the next person who tries to pet it. Draw edges to show the dependencies specified by the problem statement above.

Solution:

$$Z \longrightarrow F \longrightarrow B$$

(b) (3 points) Circe pets the lion, and it bites her hand. In terms of the unknown parameters P, Q, R, and S, what is the probability that the zookeper is on vacation?

Solution:

$$P(Z|B) = \frac{P(Z,F,B)+P(Z,\neg F,B)}{P(Z,F,B)+P(Z,\neg F,B)+P(\neg Z,F,B)+P(\neg Z,\neg F,B)}$$

$$= \frac{PR}{PR+(1-P)QS+(1-P)(1-Q)R}$$

(c) (2 points) Lord Lucky, the Lord of Castle Rock, hires a troupe of circus performers to pet the lion, once per day, in an attempt to learn the parameters P, Q, R, and S. Over the course of seven days, he collects the following observations. Based on these observations, find maximum-likelihood estimates of P, Q, R, and S.

Day	Z	F	B
1	1	0	1
2	0	1	1
3	0	1	0
4	1	0	0
5	0	0	1
6	0	1	0
7	0	1	0

Solution: The probability estimates are

$$P = P(Z) = \frac{2}{7}$$

$$Q = P(F|\neg Z) = \frac{4}{5}$$

$$R = P(B|\neg F) = \frac{2}{3}$$

$$S = P(B|F) = \frac{1}{4}$$

Question 26 ***(0 points)***

A particular hidden Markov model (HMM) has state variable X_t, and observation variables E_t, where t denotes time. Suppose that this HMM has two states, $X_t \in \{0,1\}$, and three possible observations, $E_t \in \{0,1,2\}$. The initial state probability is $P(X_1 = 1) = 0.3$. The transition and observation probability matrices are

X_{t-1}	$P(X_t = 1 \vert X_{t-1})$	X_t	$P(E_t = 0 \vert X_t)$	$P(E_t = 1 \vert X_t)$
0	0.6	0	0.4	0.1
1	0.4	1	0.1	0.6

Suppose that, in a particular test of the HMM, the observation sequence is

$$\{E_1, E_2\} = \{2, 1\}$$

(a) What is the joint probability $P(X_1 = 1, E_1 = 2, X_2 = 0)$?

Solution:

$$\begin{aligned} P(X_1 = 1, E_1 = 2, X_2 = 0) &= P(X_1 = 1)P(E_1 = 2|X_1 = 1)P(X_2 = 0|X_1 = 1) \\ &= (0.3)(0.3)(0.6) \end{aligned}$$

(b) What is the probability of the most likely state sequence ending in $X_2 = 0$? In other words, what is $\max_{X_1} P(X_1, E_1 = 2, X_2 = 0, E_2 = 1)$?

Solution:

$$\begin{aligned} \max_{X_1} P(X_1, E_1 = 2, X_2 = 0, E_2 = 1) &= \max_{X_1} P(X_1)P(E_1 = 2|X_1)P(X_2 = 0|X_1)P(E_2 = 1|X_2 = 0) \\ &= \max\left((0.7)(0.5)(0.4)(0.1), (0.3)(0.3)(0.6)(0.1)\right) \\ &= (0.7)(0.5)(0.4)(0.1) \end{aligned}$$

Question 27 ***(0 points)***

A particular hidden Markov model (HMM) has state variable X_t, and observation variables E_t, where t denotes time. Suppose that this HMM has two states, $X_t \in \{0,1\}$, and three possible observations, $E_t \in \{0,1,2\}$. The initial state probability is $P(X_1 = 1) = 0.3$. The transition and observation probability matrices are

X_{t-1}	$P(X_t = 1 \| X_{t-1})$	X_t	$P(E_t = 0 \| X_t)$	$P(E_t = 1 \| X_t)$
0	0.6	0	0.4	0.1
1	0.4	1	0.1	0.6

Suppose that, in a particular test of the HMM, the observation sequence is

$$\{E_1, E_2\} = \{2, 1\}$$

(a) What is the joint probability $P(X_1 = 0, E_1 = 2, E_2 = 1)$?

Solution:

$$\begin{aligned} P(X_1 = 0, E_1 = 2, E_2 = 1) &= \sum_{X_2} P(X_1 = 0) P(E_1 = 2 | X_1 = 0) P(X_2 | X_1 = 0) P(E_2 = 1 | X_2) \\ &= (0.7)(0.5)(0.4)(0.1) + (0.7)(0.5)(0.6)(0.6) \end{aligned}$$

(b) What is the probability of the most likely state sequence ending in $X_2 = 1$? In other words, what is $\max_{X_1} P(X_1, E_1 = 2, X_2 = 1, E_2 = 1)$?

Solution:

$$\begin{aligned} \max_{X_1} P(X_1, E_1 = 2, X_2 = 1, E_2 = 1) &= \max_{X_1} P(X_1) P(E_1 = 2 | X_1) P(X_2 = 1 | X_1) P(E_2 = 1 | X_2 = 1) \\ &= \max\left((0.7)(0.5)(0.6)(0.6), (0.3)(0.3)(0.4)(0.6)\right) \\ &= (0.7)(0.5)(0.6)(0.6) \end{aligned}$$

Question 28 ***(0 points)***

A particular hidden Markov model (HMM) has state variable X_t, and observation variables E_t, where t denotes time. Suppose that this HMM has two states, $X_t \in \{0,1\}$, and three possible observations, $E_t \in \{0,1,2\}$. The initial state probability is $P(X_1 = 1) = 0.3$. The transition and observation probability matrices are

X_{t-1}	$P(X_t = 1 \vert X_{t-1})$	X_t	$P(E_t = 0 \vert X_t)$	$P(E_t = 1 \vert X_t)$
0	0.6	0	0.4	0.1
1	0.4	1	0.1	0.6

Suppose that, in a particular test of the HMM, the observation sequence is

$$\{E_1, E_2\} = \{2, 1\}$$

(a) What is the joint probability $P(E_1 = 2, X_2 = 1, E_2 = 1)$?

Solution:

$$\begin{aligned} P(E_1 = 2, X_2 = 1, E_2 = 1) &= \sum_{X_1} P(X_1)P(E_1 = 2|X_1)P(X_2 = 1|X_1 = 0)P(E_2 = 1|X_2 = 1) \\ &= (0.7)(0.5)(0.6)(0.6) + (0.3)(0.3)(0.4)(0.6) \end{aligned}$$

(b) If it is observed that $X_2 = 0$, what is the most likely value of X_1? In other words, what is $\arg\max_{X_1} P(X_1, E_1 = 2, X_2 = 0, E_2 = 1)$?

Solution:

$$\begin{aligned} \arg\max_{X_1} P(X_1, E_1 = 2, X_2 = 0, E_2 = 1) &= \arg\max_{X_1} P(X_1)P(E_1 = 2|X_1)P(X_2 = 0|X_1) \\ &= \arg\max\left((0.7)(0.5)(0.4), (0.3)(0.3)(0.6)\right) \\ &= 0 \end{aligned}$$

Question 29 ***(0 points)***

A particular hidden Markov model (HMM) has state variable X_t, and observation variables E_t, where t denotes time. Suppose that this HMM has two states, $X_t \in \{0,1\}$, and three possible observations, $E_t \in \{0,1,2\}$. The initial state probability is $P(X_1 = 1) = 0.3$. The transition and observation probability matrices are

X_{t-1}	$P(X_t = 1 \vert X_{t-1})$	X_t	$P(E_t = 0 \vert X_t)$	$P(E_t = 1 \vert X_t)$
0	0.6	0	0.4	0.1
1	0.4	1	0.1	0.6

Suppose that, in a particular test of the HMM, the observation sequence is

$$\{E_1, E_2\} = \{2, 1\}$$

(a) What is the total probability $P(E_1 = 2, E_2 = 1)$?

Solution:

$$\begin{aligned} P(E_1 = 2, E_2 = 1) &= \sum_{X_1, X_2} P(X_1)P(E_1 = 2|X_1)P(X_2|X_1)P(E_2 = 1|X_2) \\ &= (0.7)(0.5)(0.4)(0.1) + (0.7)(0.5)(0.6)(0.6) + (0.3)(0.3)(0.6)(0.1) + (0.3)(0.3)(0.4)(0.6) \end{aligned}$$

(b) If it is observed that $X_2 = 1$, what is the most likely value of X_1? In other words, what is $\arg\max_{X_1} P(X_1, E_1 = 2, X_2 = 1, E_2 = 1)$?

Solution:

$$\begin{aligned} \arg\max_{X_1} P(X_1, E_1 = 2, X_2 = 1, E_2 = 1) &= \arg\max_{X_1} P(X_1)P(E_1 = 2|X_1)P(X_2 = 1|X_1) \\ &= \arg\max_{0,1} ((0.7)(0.5)(0.6), (0.3)(0.3)(0.4)) \\ &= 0 \end{aligned}$$

Question 30 ***(0 points)***

The University of Illinois Vaccavolatology Department has four professors, named Aya, Bob, Cho, and Dale. The building has only one key, so we take special care to protect it. Every day Aya goes to the gym, and on the days she has the key, 60% of the time she forgets it next to the bench press. When that happens one of the other three TAs, equally likely, always finds it since they work out right after. Bob likes to hang out at Einstein Bagels and 50% of the time he is there with the key, he forgets the key at the shop. Luckily Cho always shows up there and finds the key whenever Bob forgets it. Cho has a hole in her pocket and ends up losing the key 80% of the time somewhere on Goodwin street. However, Dale takes the same path to campus and always finds the key. Dale has a 10% chance to lose the key somewhere in the Vaccavolatology classroom, but then Cho picks it up. The professors lose the key at most once per day, around noon (after losing it they become extra careful for the rest of the day), and they always find it the same day in the early afternoon.

(a) Let $X_t =$ the first letter of the name of the person who has the key ($X_t \in \{A,B,C,D\}$). Find the maximum likelihood estimates of the Markov transition probabilities $P(X_t|X_{t-1})$.

Solution:

	X_t			
X_{t-1}	A	B	C	D
A	0.4	0.2	0.2	0.2
B	0	0.5	0.5	0
C	0	0	0.2	0.8
D	0	0	0.1	0.9

(b) Sunday night Bob had the key (the initial state distribution assigns probability 1 to $X_0 = B$ and probability 0 to all other states). The first lecture of the week is Tuesday at 4:30pm, so one of the professors needs to open the building at that time. What is the probability for each professor to have the key at that time? Let X_0, X_{Mon} and X_{Tue} be random variables corresponding to who has the key Sunday, Monday, and Tuesday evenings, respectively. Fill in the probabilities in the table below.

Professor	$P(X_0)$	$P(X_{Mon})$	$P(X_{Tue})$
A	0		
B	1		
C	0		
D	0		

Solution:

Professor	$P(X_0)$	$P(X_{Mon})$	$P(X_{Tue})$
A	0	0	0
B	1	0.5	0.25
C	0	0.5	0.35
D	0	0	0.4

Question 31 ***(0 points)***

Consider a hidden Markov model (HMM) whose hidden variable denotes part of speech (POS), $X_t \in \{N,V\}$ where $N=$noun, $V=$verb, the initial state probability is $P(X_1=N)=0.8$, and the transition probabilities are $P(X_t=N|X_{t-1}=N)=0.1$ and $P(X_t=V|X_{t-1}=V)=0.1$. Suppose we have the observation probability matrix given in Table 2. You are given the sentence "bill rose." You want to

E_t	rose	bill	likes
$P(E_t\|X_t=N)$	0.4	0.4	0.2
$P(E_t\|X_t=V)$	0.2	0.2	0.6

Table 2: Observation probabilities for a simple POS HMM.

figure out whether each of these two words, "bill" and "rose", is being used as a noun or a verb.

(a) List the four possible combinations of (X_1,X_2). For each possible combination, give $P(X_1,E_1,X_2,E_2)$.

Solution:

$P(X_1,E_1,X_2,E_2)$	$X_2=N$	$X_2=V$
$X_1=N$	(0.8)(0.4)(0.1)(0.4)	(0.8)(0.4)(0.9)(0.2)
$X_1=V$	(0.2)(0.2)(0.9)(0.4)	(0.2)(0.2)(0.1)(0.2)

(b) Find $P(X_2=V|E_1=\text{bill},E_2=\text{rose})$.

Solution: Using the forward algorithm, we can compute the joint probabilities as

$$\begin{aligned}P(E,X_2=V) &= P(X_1=N,E_1,X_2=V,E_2)+P(X_1=V,E_1,X_2=V,E_2)\\ &= (0.8)(0.4)(0.9)(0.2)+(0.2)(0.2)(0.1)(0.2)\\ P(E,X_2=N) &= P(X_1=N,E_1,X_2=N,E_2)+P(X_1=V,E_1,X_2=N,E_2)\\ &= (0.8)(0.4)(0.1)(0.4)+(0.2)(0.2)(0.9)(0.4)\end{aligned}$$

Dividing the first row by the sum of the two rows, we get

$$P(X_2=V|E)=\frac{(0.8)(0.4)(0.9)(0.2)+(0.2)(0.2)(0.1)(0.2)}{(0.8)(0.4)(0.9)(0.2)+(0.2)(0.2)(0.1)(0.2)+(0.8)(0.4)(0.1)(0.4)+(0.2)(0.2)(0.9)(0.4)}$$

(c) Use the Viterbi algorithm to find the most likely state sequence for this sentence.

Solution:

- To find the backpointer from $X_2=N$, we find the maximum among the two possibilities $P(X_1=N,E_1,X_2=N,E_2)$ and $P(X_1=V,E_1,X_2=N,E_2)$. The larger of the two is $P(X_1=V,E_1,X_2=N,E_2)=(0.2)(0.2)(0.9)(0.4)$, so the backpointer from $X_2=N$ points to $X_1=V$.
- To find the backpointer from $X_2=V$, we find the maximum among the two possibilities $P(X_1=N,E_1,X_2=V,E_2)$ and $P(X_1=V,E_1,X_2=V,E_2)$. The larger of the two is $P(X_1=$

$N, E_1, X_2 = V, E_2) = (0.8)(0.4)(0.9)(0.2)$, so the backpointer from $X_2 = V$ points to $X_1 = N$.

- To find the best terminal state, then, we find the maximum among the two possibilities $P(X_1 = V, E_1, X_2 = N, E_2)$ and $P(X_1 = N, E_1, X_2 = V, E_2)$. The larger of the two is $P(X_1 = N, E_1, X_2 = V, E_2) = (0.8)(0.4)(0.9)(0.2)$, so the maximum likelihood state sequence is $(X_1, X_2) = (N, V)$.