Exam 1 Review

CS440/ECE448, Spring 2021

Exam date: Friday, March 5, 1:00pm

Question 1

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.

1 G_1	2 S	3 G_2
	4 G_3	

(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, $G \in \{\varnothing, G_1, G_2, G_3, G_1G_2, G_1G_3, G_2G_3, G_1G_2G_3\}$ 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 $1 + 3 \times 2 + 3 \times 3 = 16$ 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: A heuristic is consistent if, for any two neighboring nodes m and n,

$$h_1[m] - h_1[n] \leq d[m, n], \tag{1}$$

where $d[m, n]$ is the cost of getting from node m to node n.

In this graph, the distance between any pair of neighboring nodes is $d[m, n] = 1$, so we just have to determine whether or not $h_1[m] - h_1[n] \leq 1$.

From any node m in this graph, 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 $h_1[n] = 1$, so $h_1[m] - h_1[n] \leq 1$.

If $h_1[m] = 1$, however, then we can actually move to reach the nearest goal; call that node n. In that case, we reach the goal, so it's no longer "the nearest goal that has not been reached." If there are any more goals that have not been reached, then they are 2 steps away from us, so $h_1[n] = 2$. Thus $h_1[m] - h_1[n] = -1 \leq 1$.

(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 2

For each type of maze described below, specify the time complexity and space complexity of both breadth-first-search (BFS) and depth-first-search (DFS).

(a) The Albuquerque maze has b possible directions that you can take at each intersection. No path is longer than m steps, where m is finite. There is only one solution, which is known to require exactly d steps, where $d < m$.

Solution: BFS has time complexity $O\{b^d\}$, space complexity $O\{b^d\}$. DFS has time complexity $O\left\{b^m\right\}$, space complexity $O\left\{mb\right\}$.

(b) The Belmont maze has b possible directions that you can take at each intersection. No path is longer than m steps, where m is finite. All solutions require $d = m$ steps. About half of all available paths are considered solutions to the maze.

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.

Solution: BFS has time complexity $O\{b^m\}$, space complexity $O\{b^m\}$. DFS has space complexity $O\{mb\}$. Its worst-case time complexity is $O\{b^m\}$, but on average, it performs much better.

(c) The Crazytown maze has b possible directions that you can take at each intersection. The maze is infinite in size, so some paths have infinite length. There is only one solution, which is known to require $d = 25$ steps.

Solution: BFS has time complexity $O\{b^d\}$, space complexity $O\{b^d\}$. DFS has unbounded time complexity ($O\{b^m\}$, where m is infinite), and unbounded space complexity ($O\{mb\}$, where m is infinite).

Question 3

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.

(a) 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

(b) 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 4

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 $P \in \{1, \ldots, 11\}$ specifies the current position, $G \in \{\varnothing, G_1, G_2, G_1G_2\}$ 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 $h_1[n_1] - h_1[n_2] \leq d[n_1, n_2]$.

(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.

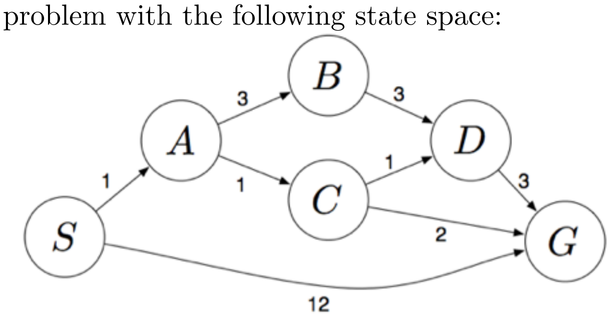
Solution: Proof by counter-example: for example, consider the state $n = (7, \varnothing)$. From this state, the shortest solution goes left, then up, then left: $d[n] = 3$ steps. The heuristic, however, is $h_2[n] = M[n, G_1] + M[G_1, G_2] = 3 + 2 = 5$, so $h[n] > d[n]$.

(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 $h_2[n] \geq 2 \geq h_1[n]$.

Question 5

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 6

You're creating sentiment analysis. You have a training corpus with four movie reviews:

Review #	Sentiment	Review
1	+	what a great movie
2	+	I love this film
3	-	what a horrible movie
4	-	I hate this film

Let $Y = 1$ for positive sentiment, $Y = 0$ for negative sentiment.

(a) What's the maximum likelihood estimate of $P(Y = 1)$?

Solution: Maximum likelihood estimate is

$$P(Y = 1) = \frac{\#\text{ times } Y = 1}{\#\text{ training tokens}} = \frac{2}{4} = \frac{1}{2}$$

(b) Find maximum likelihood estimates $P(W|Y = 1)$ and $P(W|Y = 0)$ for the ten words $W \in \{\text{what,a,movie,I,this,film,great,love,horrible,hate}\}$.

Solution: There are three cases. For the words $W \in \{\text{what,a,movie,I,this,film}\}$, $P(W|Y = 0) = P(W|Y = 1) = 1/8$. For the words $W \in \{\text{great,love}\}$, $P(W|Y = 0) = 0$, and $P(W|Y = 1) = 1/8$. For the words $W \in \{\text{horrible,hate}\}$, $P(W|Y = 1) = 0$, and $P(W|Y = 0) = 1/8$.

(c) Use Laplace smoothing, with a smoothing parameter of $k = 1$, to estimate $P(W|Y = 1)$ and $P(W|Y = 0)$ for the ten words $W \in$ {what,a,movie,I,this,film,great,love,horrible,hate}.

> **Solution:** There are three cases. For the words $W \in$ {what,a,movie,I,this,film}, $P(W|Y = 0) = P(W|Y = 1) = 2/18$. For the words $W \in$ {great,love}, $P(W|Y = 0) = 1/18$, and $P(W|Y = 1) = 2/18$. For the words $W \in$ {horrible,hate}, $P(W|Y = 1) = 1/18$, and $P(W|Y = 0) = 2/18$.

(d) Using some other method (unknown to you), your professor has estimated the following conditional probability table:

Y	$P(\text{great}|Y)$	$P(\text{love}|Y)$	$P(\text{horrible}|Y)$	$P(\text{hate}|Y)$
1	0.01	0.01	0.005	0.005
0	0.005	0.005	0.01	0.01

and $P(Y = 1) = 0.5$. All other words (except great, love, horrible, and hate) can be considered out-of-vocabulary, and you can assume that $P(W|Y) = 1$ for all out-of-vocabulary words. Under these assumptions, what is the probability $P(Y = 1|R)$ that the following 14-word review is a positive review?

$R = \{$"I'm horrible fond of this movie, and I hate anyone who insults it."$\}$

> **Solution:**
>
> $$P(Y = 1|R) = \frac{P(Y = 1, R)}{P(Y = 1, R) + P(Y = 0, R)} = \frac{(0.5)(0.005)(0.005)}{(0.5)(0.005)(0.005) + (0.5)(0.01)(0.01)} = \frac{1}{5}$$

Question 7

Consider a Nave Bayes classifier with 100 feature dimensions. The label Y is binary with $P(Y = 0) = P(Y = 1) = 0.5$. All features are binary, and have the same conditional probabilities: $P(X_i = 1|Y = 0) = a$ and $P(X_i = 1|Y = 1) = b$ for $i = 1, \ldots, 100$. Given an item X with alternating feature values $(X_1 = 1, X_2 = 0, X_3 = 1, ..., X_{100} = 0)$, compute $P(Y = 1|X)$.

> **Solution:**
>
> $$\begin{aligned} P(Y = 1|X) &= \frac{P(Y = 1)\prod_{i=1}^{100} P(X_i|Y = 1)}{P(Y = 1)\prod_{i=1}^{100} P(X_i|Y = 1) + P(Y = 0)\prod_{i=1}^{100} P(X_i|Y = 0)} \\ &= \frac{0.5b^{50}(1-b)^{50}}{0.5b^{50}(1-b)^{50} + 0.5a^{50}(1-a)^{50}} \\ &= \frac{b^{50}(1-b)^{50}}{b^{50}(1-b)^{50} + a^{50}(1-a)^{50}} \end{aligned}$$

Question 8

Use the axioms of probability to prove that $P(\neg A) = 1 - P(A)$.

Solution:

- From the third axiom, $P(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$.
- The event $(A \vee \neg A)$ is always true, so from the second axiom, $P(A \vee \neg A) = 1$. The event $(A \wedge \neg A)$ is always false, so from the second axiom, $P(A \wedge \neg A) = 0$.
- Combining the two statements above, $1 = P(A) + P(\neg A)$. Q.E.D.

Question 9

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 10

A friend who works in a big city owns two cars, one small and one large. Three-quarters of the time he drives the small car to work, and one-quarter of the time he drives the large car. If he takes the small car, he usually has little trouble parking, and so is at work on time with probability 0.9. If he takes the large car, he is at work on time with probability 0.6. Given that he was on time on a particular morning, what is the probability that he drove the small car?

Solution: Let S be the event "takes the small car," and T is the event "arrives on time." Then

$$P(S|T) = \frac{P(T|S)P(S)}{P(T)} = \frac{P(T|S)P(S)}{P(T|S)P(S) + P(T|\neg S)P(\neg S)} = \frac{0.9(3/4)}{0.9(3/4) + 0.6(1/4)} = \frac{27}{33}$$

Question 11

You're trying to determine whether a particular newspaper article is of class $Y = 0$ or $Y = 1$. The prior probability of class $Y = 1$ is $P(Y = 1) = 0.4$. The newspaper is written in a language that only has four words, so that the i^{th} word in the article must be $W_i \in \{0, 1, 2, 3\}$, with probabilities given by:

$$\begin{array}{ll} P(W_i = 0|Y = 0) = 0.3 & P(W_i = 0|Y = 1) = 0.1 \\ P(W_i = 1|Y = 0) = 0.1 & P(W_i = 1|Y = 1) = 0.1 \\ P(W_i = 2|Y = 0) = 0.1 & P(W_i = 2|Y = 1) = 0.3 \end{array}$$

The article is only three words long; it contains the words

$$A = (W_1 = 3, W_2 = 2, W_3 = 0)$$

What is $P(Y = 1, A)$?

Solution:

$$P(Y = 1, A) = P(Y = 1)P(W_1 = 3|Y = 1)P(W_2 = 2|Y = 1)P(W_3 = 0|Y = 1) = (0.4)(0.5)(0.3)(0.1)$$

Question 12

20% of students at U of I are part of the Greek system. Amongst these students, 10% study engineering. Furthermore, 15% of the entire student body studies engineering. Given that we know that a student studies engineering, what is the probability that the student is not part of the Greek system?

Solution: Define G=student part of the Greek system, E=student studies engineering. We are given that $P(G) = 0.2$ and $P(E|G) = 0.1$, from which we may infer that $P(E, G) = 0.02$. We are also given that $P(E) = 0.15$, from which we may infer that

$$\begin{aligned} P(\neg G, E) &= P(E) - P(E, G) = 0.13 \\ P(\neg G|E) &= \frac{P(\neg G, E)}{P(E)} \\ &= \frac{0.13}{0.15} = \frac{13}{15} \end{aligned}$$

Question 13

Consider the following joint probability distribution:

$$\begin{aligned} P(A, B) &= 0.12 \\ P(A, \neg B) &= 0.18 \\ P(\neg A, B) &= 0.28 \\ P(\neg A, \neg B) &= 0.42 \end{aligned}$$

What are the marginal distributions of A and B? Are A and B independent and why?

Solution: $P(A) = 0.3, P(\neg A) = 0.7, P(B) = 0.4, P(\neg B) = 0.6$. They are independent, because $P(A)P(B) = P(A, B) = 0.12, P(A)P(\neg B) = P(A, \neg B) = 0.18$, and so on.