CS 6601 Midterm – Spring 2021

Please read the following instructions thoroughly.

Fill out this PDF form and submit it on **Gradescope**. Remember to also submit on Canvas. **You will be penalized with 5 points on this exam if you don't submit on both the platforms.**

You have unlimited resubmissions until the deadline. You can: **(a)** type directly into the form – we highly recommend using Adobe Reader DC (or Master PDF on Linux). If you are on MacOS, please do not use Preview, as we have seen **major** issues with it in the past. Other programs may not save your answers, so **please keep a backup**; or **(b)** print, hand-write & scan. You can combine the methods as well.

Submit only a single PDF – no phone pictures, please! (You may use an app like CamScanner or Office Lens if you do not have scanner access.) Do not add pages unless absolutely necessary; if you do, please add them at the end of the exam **only**, and clearly label **both** the extra page and the original question page. Submit **ALL** pages of the exam, not only the completed ones.

Do not forget to fill the checklist at the end before turning in the exam. The exam may not be graded if it is left blank.

The exam is open-book, open-note, open video lectures, with no time limit aside from the open period. No internet use is allowed, except for e-text versions of the textbook, this semester's CS6601 course materials, Piazza, and any links provided in the PDF itself. No resources outside this semester's 6601 class should be used. Do not discuss the exam on Piazza, Slack, or any other form of communication. More generally, do not post **publicly** about the exam. If there is a question for the teaching staff, **please make it private on Piazza and tag it as Midterm Exam with the question number in the subject line** (for example, a question on Search would be "Midterm Exam #2"). Please make **different posts for different questions.**

Please round all your final answers to 6 decimal places, don't round intermediate results. You can use `round(your_number, 6)` function in Python for help.
You will not receive full credit if your answers are not given to the specified precision.

Point breakdown (Each question has sub-parts with varying points):

	Q1	Q2	Q3	Q4	Q5	Q6	Total
Pts	**12**	**10**	**27**	**16**	**20**	**15**	**100**

1. Game Playing

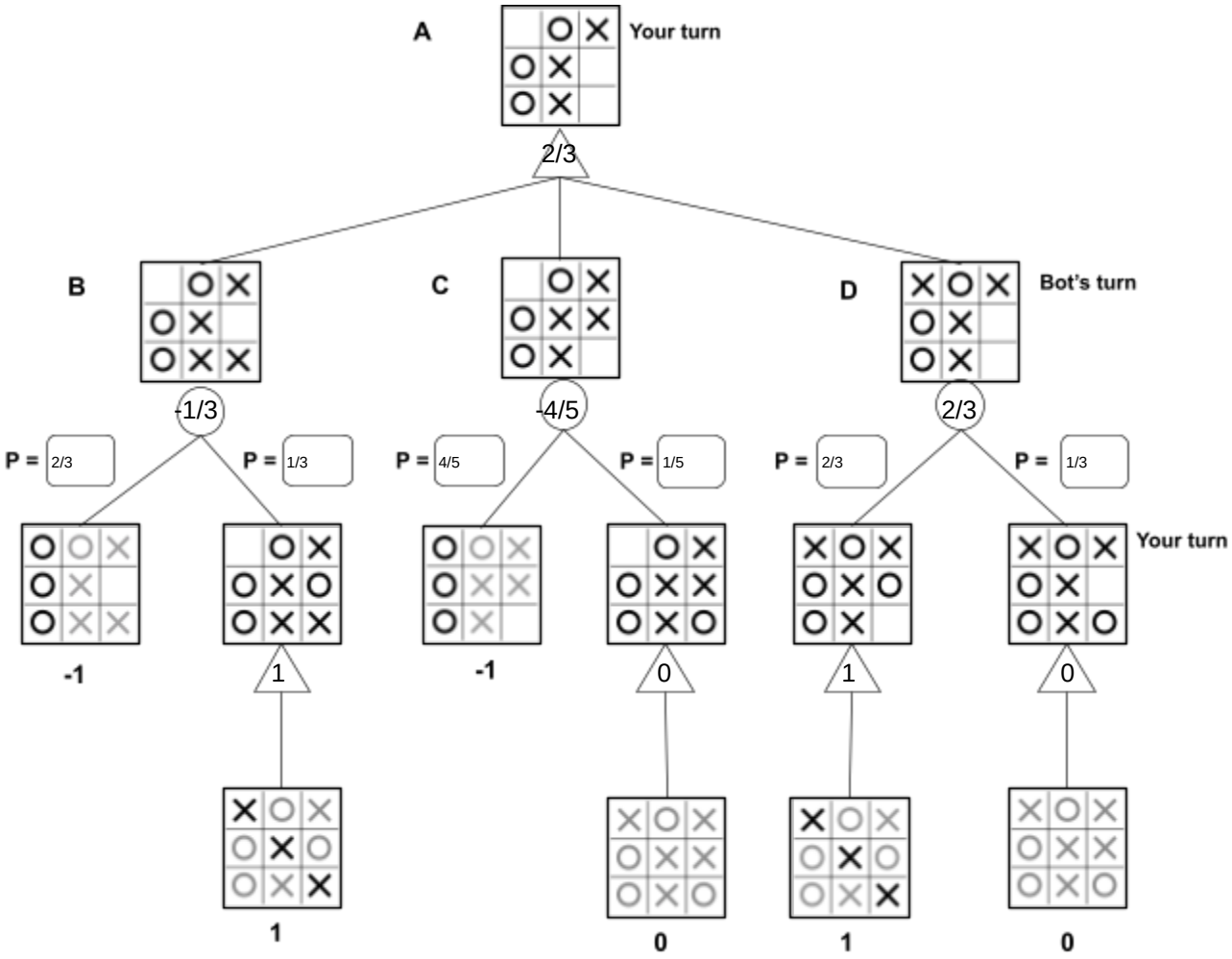
(12 points)

Question A - Expectimax

(6 points)

The Teaching Assistants of the CS 6601 course have built a random bot (“O” player) to play the game of Tic Tac Toe. However, you observe that this bot is not uniformly random. In fact, you observe that the bot is more likely to fill “O”s into the upper rows of the game. Specifically, the bot is **twice as likely** to pick a slot in the middle than the bottom row, and **again twice as likely** to pick a slot in the upper row than the middle row. With this observation in mind, you decide to play against this bot using a modified version of the expectimax algorithm (without the minimizing step) to maximize your expected outcomes.

Consider the game tree on the next page where the terminal values are the scores of the game. Answer the below questions assuming that you are the maximizing player (“X” player) while the bot is a chance player.



1.A.1. Solve the game tree above using the modified expectimax algorithm, then fill in the values you obtain for the following nodes for **1 point** each. You will not receive any points for filling out the game tree alone; you must fill in the blanks below. **(4 points)**

a. A = 0.666667

b. B = -0.333333

c. C = -0.8

d. D = 0.666667

For the below two questions (1.A.2 and 1.A.3), answer with respect to the score that you achieve at node A, using the modified expectimax algorithm.

1.A.2. What is your expected score if you take the expectimax optimal action? **(1 point)**

0.666667

1.A.3. Multiple outcomes are possible for your expectimax play. What is the worst possible score that you could get given you play optimally from the starting board state in A? **(1 point)**

0

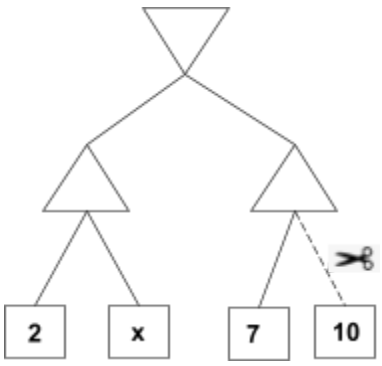
Question B - Alpha Beta Pruning

(6 points)

For each of the game-trees shown below, fill in the values of x for which the dashed branch with the scissors will be pruned. You can assume that the nodes are evaluated from left to right. Rules for filling in the blanks:

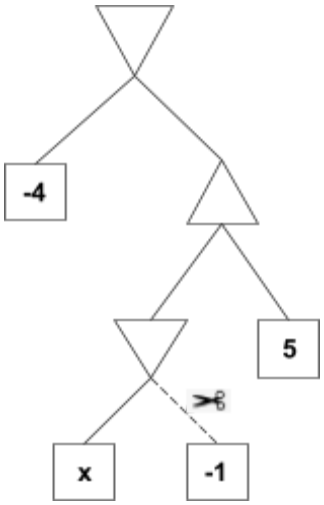
1. If pruning will not happen for any value of x fill "**None**".
2. If pruning will happen for all values of x fill "**All**".
3. If pruning occurs for specific values of x, write "**x >= [num]"** and/or "**x <= [num]"** to specify the range of values x can take on.
4. **To summarize, the string format you should use to specify the range of x can be from any of "x >= [num]" , "x <= [num]", "All", and/or "None"**.

1.B.1. (2 points)



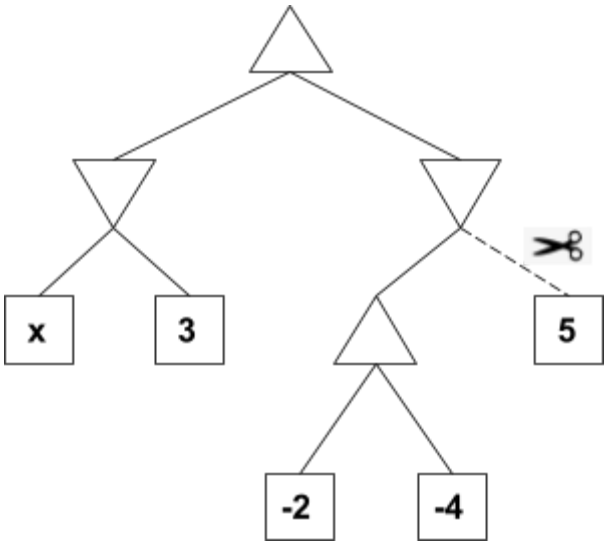
x <= 7

1.B.2. (2 points)



None

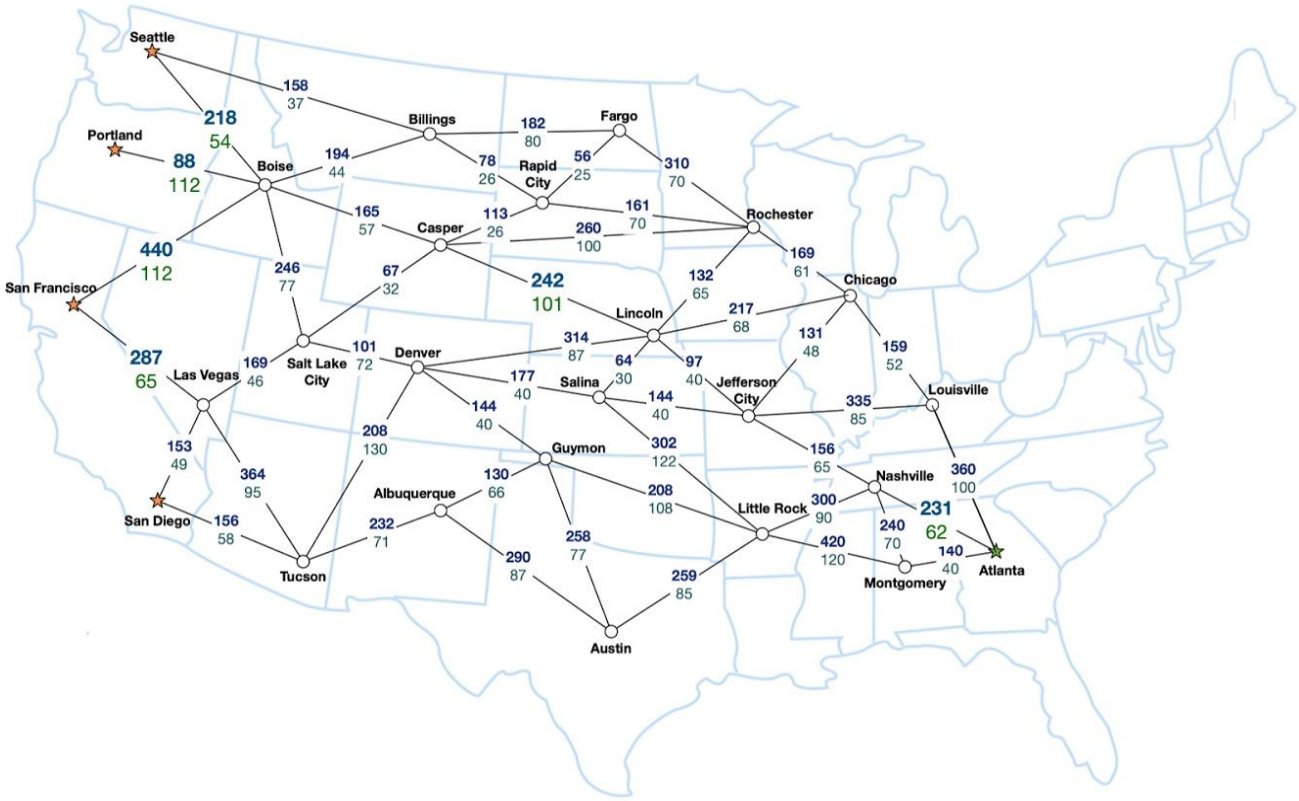
1.B.3. (2 points)



x >= -2

2. Search

(10 points)



It's March 2021 and COVID-19 vaccination is in full swing. Vaccine shots which are to be administered to people are being transported around the country. As vaccination is a time sensitive task, we want to take the quickest path to our destinations.

Given that a shot of vaccine is manufactured in Atlanta, your task is to find a path that will take the shortest amount of time (absolute quickest) to go to one of the West Coast cities (Seattle, Portland, San Francisco, San Diego; each marked with stars on the map provided).

The vaccines are being transported in electric powered trucks. There are two distinct costs for every edge/path. The top value (blue & **bold**) represents the time it takes to go from one city to the other in **minutes**, and the bottom value (green) represents the energy power consumption in **kWh**.

Using your knowledge about search from CS 6601, you find the **quickest** path(with respect to time taken) from Atlanta to the West Coast, by running a search algorithm starting from Atlanta. Specifically, you find that the following path: **Atlanta - Nashville - Jefferson City - Lincoln - Casper - Boise - Portland** is the **quickest.**

Now assume that the battery pack of the truck has a capacity of **365 kWh**, which may not be enough to make the trip you previously found on a single charge. Suppose also that you do not want to spend time at a charging station. You want to find a new quickest path, such that you'll be able to **cover the journey on a single charge**. You'd like to avoid stopping to recharge, even if it ends up extending trip time. Note that the energy cost corresponding to kWh used is the green & non-bolded value specified below every edge in the map.

If you look closely at the energy costs on the map, you will notice that the energy consumption is not necessarily a function of only time/distance. This happens because of the regenerative braking system of the truck, combined with the terrain properties. In other words the truck recharges itself while going downhill and braking, which lets you travel across some cities consuming less energy.

2.1. Now it's time to find the path from Atlanta to the West Coast (either of Seattle, Portland, San Diego, San Francisco) which can be taken given the battery pack restrictions. Report the **quickest** such path and the total **power and time** used on the path. **(6 points)**

(Use the 'index' to 'city' names mapping given in Table below to report your answer)

Index	City	Index	City
ALB	Albuquerque	**LTR**	Little Rock
ATL	Atlanta	**LOU**	Louisville
AUS	Austin	**MNT**	Montgomery
BLG	Billings	**NVL**	Nashville
BSE	Boise	**PTD**	Portland
CSP	Casper	**RPD**	Rapid City
CGO	Chicago	**RCT**	Rochester
DNV	Denver	**SLN**	Salina
FRG	Fargo	**SLT**	Salt Lake City
GYN	Guymon	**SND**	San Diego
JFC	Jefferson City	**SNF**	San Francisco
LSV	Las Vegas	**SEA**	Seattle
LCN	Lincoln	**TCS**	Tucson

Eg. For the path **Atlanta - Nashville - Jefferson City - Lincoln - Casper - Boise - Portland**, report the answer like this: **ATL, NVL, JFC, LCN, CSP, BSE, PTD**

a. Path: **(4 points)**

ATL, NVL, JFC, LCN, RCT, RPD, BLG, SEA

b. Time duration: **(1 point)**

1013 minutes

c. Power consumption: **(1 point)**

365kWh

2.2. Suppose we decide to make a **30 minute stop** to recharge **50 kWh of energy**. Assuming we **must** make the stop at some point in the trip, which of the following is the quickest path from Atlanta to the West Coast satisfying the given constraints? **(2 points)**

- (●) Quickest Path found in question 2.1.a
- ( ) Atlanta - Nashville - Jefferson City - Lincoln - Casper - Boise - Portland
- ( ) Neither the first nor second option is the quickest path satisfying the given constraints

2.3. If you were to design an admissible heuristic for the problem from question 1 (energy consumption), which of the following would be the best for your design? Select **one** answer below. **(2 points)**

Note: Mileage is estimated based on the euclidean distance between the cities.

- (●) Best case (minimum) energy consumption as a function of mileage
- ( ) Worst case (maximum) energy consumption as a function of mileage
- ( ) Average case energy consumption as a function of mileage
- ( ) Best case (minimum) time as a function of mileage (assuming constant travel at the fastest speed limits in the US, which is 80 mph)

3. Genetic Algorithms

(27 points)

Question A - Creating a Superkitty Society (Genetic Algorithms)

(15 points)

Suppose we have a population of kittens, some of whom have superpowers. You are a genetic engineer at StarnerCorp, whose job it is to maximize the quality of superpowers future generations of kittens will have through the use of genetic algorithms.
In the table below, each number corresponds to a superpower and the larger the number, the greater the superpower. The list of superpowers is given below (note the individual superpowers are not important for the question, but they provide fun context):

Superpower Gene	Superpower
0	Teleportation
1	Ability to communicate with humans
2	Invisibility
3	Flight
4	Mind Control
5	Generate Air
6	Generate Fire
7	Generate Water
8	Generate Earth
9	Ability to steal superpowers

A gene is ***only*** expressed if the DNA of a kitten contains ***at least 3 instances*** of the gene anywhere in the DNA sequence (does not have to be consecutive). The DNA of a kitten is composed of ***exactly 12*** genes.

As an example, here are some samples of DNA you might find:

2398 - 1078 - 2534	This kitten has no superpowers.
1088 - 2**9**34 - **99**56	This kitten has the ability to steal superpowers
0101 - 0101 - 0101	This kitten has teleportation and the ability to communicate with humans

Next, you evaluate the ***fitness*** of a kitten's DNA using the following formula:

Fitness(DNA sequence) = (**A** + **B**) / 12
where **A** = number of *expressed* superpowers in the DNA sequence
where **B** = sum of the total power of the unique *expressed* superpowers of the DNA sequence

As an example, we fill in the fitness values corresponding to each DNA sequence in the table that was given earlier:

2398 - 1078 - 2534	0/12 = 0
1088 - 2**9**34 - **99**56	(1 + 9) / 12 = 10/12 = 0.833333
0101 - 0101 - 0101	(2 + (0 + 1))/12 = 3/12 = 0.25

3.A.1. Suppose we have the following 8 DNA sequences of the kittens. Calculate the corresponding fitness values. Please remember to round your answers as per the rounding rules of the exam. **(3 points)**

1111 - 2222 - 3333	0.75
3790 - 1029 - 0758	0.083333
7721 - 4211 - 4006	0.166667
4669 - 1897 - 8334	0
2013 - 3020 - 5433	0.416667
2210 - 5914 - 1814	0.166667
0452 - 5419 - 5890	0.5
5733 - 5888 - 1132	1.083333

3.A.2. List the top 3 kittens with the highest fitness values from Question 1 in decreasing fitness order and list the DNA sequence below. **(3 points)**

Note we have already done part of the problem for you; your answer should be consistent with ours. Please remember to round your answers as per the rounding rules of the exam.

Ranking	DNA Sequence	Fitness Score
#1	5733 - 5888 - 1132	1.083333
#2	1111 - 2222 - 3333	0.75
#3	0452 - 5419 - 5890	0.5

3.A.3. In this part, we use the top 3 individuals in Question 3.A.2 for crossover with one another.

Recall that in crossover, the 2 DNA sequences are combined together depending on the crossover point. For this problem, the crossover point we use is defined as the point ***after the 6th*** gene. We compute all 6 permutations for crossover among the top 3 individuals from the previous question. Note that the "Order" column contains ordered pairs, which tell us which DNA sequences from the final table in Question 2 to use for the first and second half of the crossover. For example, [1, 2] indicates that the first half of the DNA sequence ranking #1 and the second half of the DNA sequence ranking #2 should be used.

For clarity, here is an example of the Crossover operation we use in this problem. In the first row, we take the first half of the sequence ranked **A** and combine it with the second half of the sequence ranked **B**. In the second row we do it vice versa in accordance with the "Order" column. Please note that these are just examples and may not be the correct DNA sequences.

Ranking	DNA Sequence	Order	Crossover DNA Sequence
A	**1234-78**23-4576	[A,B]	1234-7863-8751
B	9876-23**63-8751**	[B,A]	9876-2323-4576

We have already done part of the problem for you; your answer should be consistent with ours. Please remember to round your answers as per the rounding rules of the exam. **(3 points)**

DNA Sequences to Crossover	DNA Sequence	New Fitness Score
[1,2]	5733-5822-3333	0.333333
[1,3]	5733-5819-5890	0.5
[2,1]	1111-2288-1132	0.416667
[2,3]	1111-2219-5890	0.166667
[3,1]	0452-5488-1132	0
[3,2]	0452-5422-3333	0.583333

Mutation:
Now assume we roll a 12-sided die to denote which index in the DNA sequence that gets mutated, and then we roll a 10-sided die to denote which superpower that index gets changed to. The DNA sequences we mutate are those we generate using Crossover in Question 3.A.3. They are presented in the same order as in the Crossover table from Question 3.A.3. Assume the rolls corresponding to each sequence are as follows:

DNA Sequence being mutated	DNA Index being mutated (begins from 0)	Superpower to change to
[1,2]	7	5
[1,3]	11	9
[2,1]	1	8
[2,3]	0	2
[3,1]	5	2
[3,2]	9	0

Mutation Example:

DNA Sequence	DNA Index being mutated (begins from 0)	Superpower to change to	Mutated DNA Sequence
1234-7**8**23-4576	5	4	1234-7**4**23-4576

3.A.4. What are the DNA sequences and fitness scores of the children after applying mutation? Again, the DNA sequences we mutate are those we generate using Crossover in Question 3. They are presented in the same order as in the Crossover table from Question 3 and the Mutation table above. Use the corresponding rows in the Mutation table to apply mutation. **(3 points)**

Note: this calculation is for the children that are produced ***after*** crossover. We have already done part of the problem for you; your answer should be consistent with ours. Please remember to round your answers as per the rounding rules of the exam.

Mutation	DNA Sequence	New Fitness Score
[1,2]	5733-5825-3333	0.833333
[1,3]	5733-5819-5899	1.333333
[2,1]	1811-2288-1132	1.166667
[2,3]	2111-2219-5890	0.416667
[3,1]	0452-5288-1132	0.25
[3,2]	0452-5422-3033	0.583333

3.A.5. Finally, select the top 4 individuals for the next generation of the genetic algorithm (after crossover and mutation) and sort them in decreasing fitness order. We have already done part of the problem for you; your answer should be consistent with ours. Please remember to round your answers as per the rounding rules of the exam. **(3 points)**

Ranking	DNA Sequence	Fitness Score
#1	5733-5819-5899	1.333333
#2	1811-2288-1132	1.166667
#3	5733-5825-3333	0.833333
#4	0452-5422-3033	0.583333

Question B - Knapsack (Hill Climbing / Simulated Annealing)

(12 points)

Knapsack is a popular problem in combinatorial optimization which is generally solved using dynamic programming algorithms. In this section, however, we will solve this optimization problem using hill climbing and simulated annealing. The problem is as follows.

Consider a set of N items numbered $1, 2, \ldots, N$. Each item i is associated with a value $V[i]$ and a weight $W[i]$. The task is to choose a subset of these items such that the sum of values of the items in the subset is maximal while the total weight (sum of the weights of the items) is not more than some given weight limit B.

We are given a set of 7 items, numbered $1, 2, \ldots, 7$. The corresponding values and weights in order are as follows:

item #	**V**alue	**W**eight
1	50	5
2	40	4
3	30	6
4	50	3
5	30	2
6	24	6
7	36	7

The weight limit is $B = 20$

In order to make the computation easier we employ a few restrictions. We will limit the length of the subsets we choose to either 4 or 5. The subset chosen will also always be ordered in ascending order. For example, [2,3,5,6] is a valid subset, but [3,5,6,2] is not. This ordering ensures that different permutations of the same subset are not considered.

Given an ordered subset, the neighboring subsets are obtained by applying one of the following operations.

1. Replace **exactly one** item. For example, [2,3,**4**,6] is a neighbor of [2,3,**5**,6]. A valid neighbor should also be in ascending order. Hence, [2,**7**,5,6] is NOT a neighbor of [2,**3**,5,6].
2. Add exactly one item to the subset. This is only possible when the current size is 4, since we are only considering subsets of size 4 or 5. The new element can be added in any position but the resulting subset should also be in ascending order. For example, [**1**,2,3,5,6] and [2,3,**4**,5,6] are neighbors of [2,3,5,6].
3. Remove exactly one item from the subset. This is only possible when the current size is 5 since we are only considering subsets of size 4 or 5. Any one of the 5 elements can be removed to get a neighbor subset. For example, [2,3,4,6] is a neighbor of [2,3,4,**5**,6]

IMPORTANT NOTE: If the total weight of the subset obtained by performing any of the above three operations exceeds the weight limit B, then it is NOT a valid neighbor. For example, [1,2,**6**,7] is NOT a neighbor of [1,2,**4**,7] since ($W[1] + W[2] + W[6] + W[7]) = 22 > B$.

We define the fitness function as the sum of the values of all the elements in the subset. The goal of the optimization is to maximize this fitness function.

We start with a subset $S_0 = [1, 2, 5, 7]$, whose sum of values is 156 and weight is 18.

3.B.1. Which of the following are neighbors of [1,2,5,7]? (Mark all that apply) **(2 point)**

- ☐ [1,2,3,7]
- ☐ [1,6,5,7]
- ■ [1,2,4,7]
- ☐ [1,2,4,5,7]
- ■ [1,2,5,6]

3.B.2. To solve the optimization problem, we first employ hill climbing starting with **[1,2,5,7]**.

a. What is the next subset (a neighbor of [1,2,5,7]) chosen by the hill climbing algorithm? (In case of the same value, choose the one with the lower weight.) **(2 points)**

[1,2,4,7]

b. What is the fitness of the subset found in **3.B.2.a**? **(1 point)**

176

c. What is the maximum fitness amongst all the valid neighbours of the state in **3.B.2.a**? **(2 points)**

170

d. Is the point you found in **3.B.2.a** a local maximum? **(0.5 points)**

- ◉ Yes
- ○ No

3.B.3. We will now employ simulated annealing to find a solution. Fill in the table for the probability of acceptance for a subset, given the previous subset in the table. You should use the annealing temperature (given) from the same row for calculating the probability of acceptance for each subset. When performing the calculation for a given row, assume that the previous subset was accepted irrespective of its probability. Do not compute a joint probability. Please note that we are starting with the same subset as hill climbing - i.e. [1,2,5,7] **(2 points)**

$$Prob\,(Acceptance) = e^{(\frac{\Delta E}{T})}\ if\ \Delta E < 0,\ else\ 1$$

$$\Delta E = fitness\,(Current\ subset) - fitness(Previous\ \ subset)$$

Please remember to round your answers as per the rounding rules of the exam.

Subset	Temperature	Probability of acceptance	Total weight
[1,2,5,6]	1000	0.988072	17
[1,2,4,5,6]	300	1	20
[1,2,4,5]	100	0.786628	14
[1,2,3,4,5]	20	1	20

Note: The subset [1,2,3,4,5] provides a better local maximum point than the hill climbing. In this case, this subset is the global maximum.

3.B.4. Simulated annealing with the temperature parameter equal to or tending to which of the following values at all times, will resemble hill climbing? **(1 point)**

- (●) 0
- ( ) ∞
- ( ) 1
- ( ) None of these

3.B.5. Which of the following statements are true? Mark all that apply. **(1.5 points)**

- [■] Local beam search with k=1 is same as hill climbing
- [■] Simulated Annealing is guaranteed to converge to the global optimum solution under a temperature schedule that starts with a very high value and reduces infinitesimally in each step.

4. Constraint Satisfaction Problems

(16 points)

Exercise Scheduling

Due to the global health pandemic, John has been self-quarantining for several months. Although he has been able to keep himself and his family safe, the coziness of his safe home is overwhelming and he has forgotten his regular exercise routine. Therefore, John decides to create a new exercise routine to feel healthier.

John's schedule is a **six** day routine and each day is identified numerically. The first Day is Day1, the second is Day2, and so on.

There are three types of exercises: Pull, Push, and Leg. Here are the exercises:

Pull

- Pulldown
- Pullup
- Cable Row
- Dumbbell Curl

Push

- Bench Press
- Overhead Press
- Pushdown

Leg

- Squat
- Leg Press
- Deadlift

The exercises Bench Press, Squat, and Deadlift are also referred to as the **Big Three** exercises.

John will abide by the constraints given below when he makes his exercise schedule. Note that for order based constraints, the order of exercises John does on any given day does not matter (i.e. having Day1: exercise Z, exercise X is equivalent to Day1: exercise X, exercise Z). We only care about order for exercises on different days (i.e. exercise X is done **before** exercise Z if we have Day1: exercise X and Day2: exercise Z)

a. All exercises must be done only once except the **Big Three** exercises.
b. All **Big Three** exercises must be done twice within the six days schedule (i.e. **exactly two days** should be allocated for each one of the **Big Three**).
c. Pulldown must be done within the first 3 days. (i.e. Day1, Day2, or Day3)
d. John must finish all 10 exercises within the 6 days.
e. Pulldown must be done before Pullup.

f. Cable Row must be done after Pullup.
g. Leg Press must be done on the third day.
h. Pullup must be done on the second day.
i. Dumbbell Curl must be done after Pullup and before Cable Row.
j. On each day, John must do at least 2 exercises.
k. John should do at least 1 Pull exercise per day on the first four days.
l. The exercises in the first 3 days should be all different (i.e. if John does Pushdown once on Day1, he cannot do Pushdown at any other time during Day1 or at all on Day2 and Day3).

John took the AI course at Georgia Tech and he wants to apply what he learned to his life. He decides to model the exercise scheduling problem as a Constraint Satisfaction Problem. He sets the exercises as variables and days as domain values.

4.1. Mark all of the exercises that have unary constraints. **(2 points)**

- [x] Pulldown
- [x] Pullup
- [ ] Squat
- [x] Leg Press
- [ ] Cable Row
- [ ] Dumbbell Curl
- [ ] None of the above

4.2. A single variable is called **node-consistent** if all the values in the variable's domains satisfy the variable's unary constraints. What are the domains for Pulldown, Pullup, and Cable Row after applying **node-consistency**? Identify each day in the domain with its corresponding number and write your answer in increasing order of domain values (i.e. 1,2,6 instead of 6,1,2). **(3 points)**

a. Pulldown:

1, 2, 3

b. Pullup:

2

c. Cable Row:

1, 2, 3, 4, 5, 6

4.3. Mark all of the exercises that have binary constraints. **(2 points)**

- [x] Pulldown
- [x] Pullup
- [ ] Squat
- [ ] Leg Press
- [x] Cable Row
- [x] Dumbbell Curl
- [ ] None of the above

4.4. What could be the domains of Pulldown, Pullup, and Cable Row after applying **both arc-consistency** and **node-consistency**? Identify each day in the domain with its corresponding number and write your answer in increasing order of domain values (i.e. 1,2,6 instead of 6,1,2). **(3 points)**

a. Pulldown:

1

b. Pullup:

2

c. Cable Row:

4

4.5. Min-conflict is a technique used to solve CSPs. It starts with an initial assignment of values and randomly selects a variable that is violating at least one constraint and reassigns a value to the variable such that the number of constraints violated by that variable is minimized. We apply it to this question by initializing a tentative schedule and rescheduling an exercise to reduce the number of constraint violations it has. For this problem, if one constraint is violated multiple times by exercises on various days, we count it as one violation. For example, if constraint **a.** is violated by two exercises, we will count this as one violation.

John is testing out the tentative schedule below, and is using min-conflicts to reduce conflicts with constraints **a.** through **m.** (listed above). Note that this is **NOT** the final version of his schedule.

Day 1: Pulldown, Bench Press, Overhead Press
Day 2: Pushdown, Squat, Deadlift
Day 3: Dumbbell Curl,Cable Row
Day 4: Squat, Leg Press, Bench Press
Day 5: Pullup, Deadlift
Day 6: Overhead Press

a. How many violations are there among all the constraints? If one constraint is violated multiple times by the schedule, consider that as one violation. **(2 point)**

   Number of violations: 7

b. Suppose we want to reschedule Pulldown on **Day 1** (to a **different** day). Which day will minimize the number of violations? Break ties using the earliest day possible. Each constraint should be counted as at most one violation similar to part **a**. **(4 points)**

   2

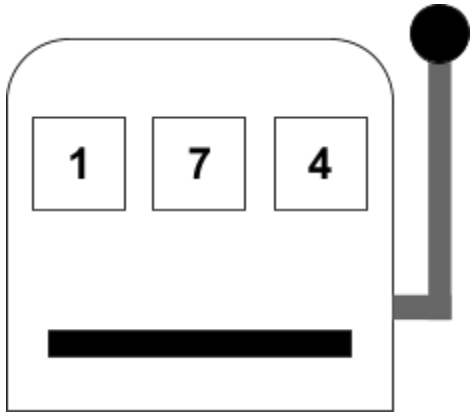
5. Probability

(20 points)

Question A - Probability and Expected Value of Discrete Random Variable

(8 points)

You decide to head to the casino and try out your luck at a new special slot machine recently added to the casino. This slot machine replaces the standard spinning reels with three digital screens, each one capable of displaying any real number. At the time of your visit, however, the machine is programmed to randomly display an integer between 1 and 7 (inclusive), on each of the three screens when the lever of the machine is pulled. Your score is calculated by summing the three slot values that are displayed (e.g. the score of the machine seen below would be 1+7+4 = 13). You may assume that the individual slots of the machine are independent and uniformly distributed.



5.A.1 What is the expected value of your score after pulling the lever once? Please remember to round your answers as per the rounding rules of the exam. **(3 points)**

12.0

5.A.2 Suppose now that you can earn prize tickets by earning a certain score. Specifically, for any odd score earned, the number of prize tickets awarded is equivalent to the score earned. For any even score, no prize tickets are given out. What is the expected number of prize tickets earned after pulling the lever once? Assume you are able to receive fractional prize tickets for this calculation. Please remember to round your answers as per the rounding rules of the exam. **(5 points)**

6.017493

Question B - Variable Independence

(5 points)

There has been a recent upsurge in the price of Bitcoin. Having studied the basics of Probability you are excited to apply this knowledge and have fun!

Consider a boolean variable S (social media) which represents whether bitcoin was trending on social media platforms lately (S = 1 if trending). Also, consider another boolean variable V (value) which tells us whether the price of bitcoin changed substantially (1 if the price changed more than 5%). Given the partial joint probability distribution table below between the variables S and V, find the values of P(S = 0, V = 0) and P(S = 0, V = 1) such that the binary variables S and V become independent.

S	V	P(S, V)
1	1	0.6
1	0	0.2

5.B.1. What is the required value of P(S = 0, V = 0)? Please remember to round your answers as per the rounding rules of the exam. **(2.5 points)**

0.05

5.B.2. What is the required value of P(S = 0, V = 1)? Please remember to round your answers as per the rounding rules of the exam. **(2.5 points)**

0.15

Question C - Conditional Probability and Bayes Rule

(7 points)

5.C.1. With some cash in hand, you are now considering whether to invest some money in the Bitcoin market. You first talk to one of your friends and she claims that the price of Bitcoin will go up. However, based on the past experience you have deduced that your friend's advice turns out to be true only 60% of the time. You also know that thus far the price of bitcoin has gone up only 20% of the time. Assume that the probability that your friend's advice is correct and the probability that the price of bitcoin will actually go up are independent. Given that your friend claims the price of bitcoin will rise, what is your updated belief (probability) that the price of Bitcoin will go up? Note that this problem is entirely independent of part Question 5.B.
Please remember to round your answers as per the rounding rules of the exam. **(3.5 points)**

0.272727

5.C.2. After talking to your friend, you consult a professional investment advisor to gather more information about the price of bitcoin. She also claims that the price of Bitcoin will rise. Suppose that the probability that the advisor is correct is 85%. Assume independence among the probabilities associated with the accuracy of your friend, the accuracy of the advisor and the price of bitcoin increasing. What is your updated belief (probability) that the price of bitcoin will rise, given that **both** your friend and the investor claim that the price of bitcoin will rise? If required, you should reuse rounded probability values from 5.C.1. Please remember to round your answers as per the rounding rules of the exam. **(3.5 points)**

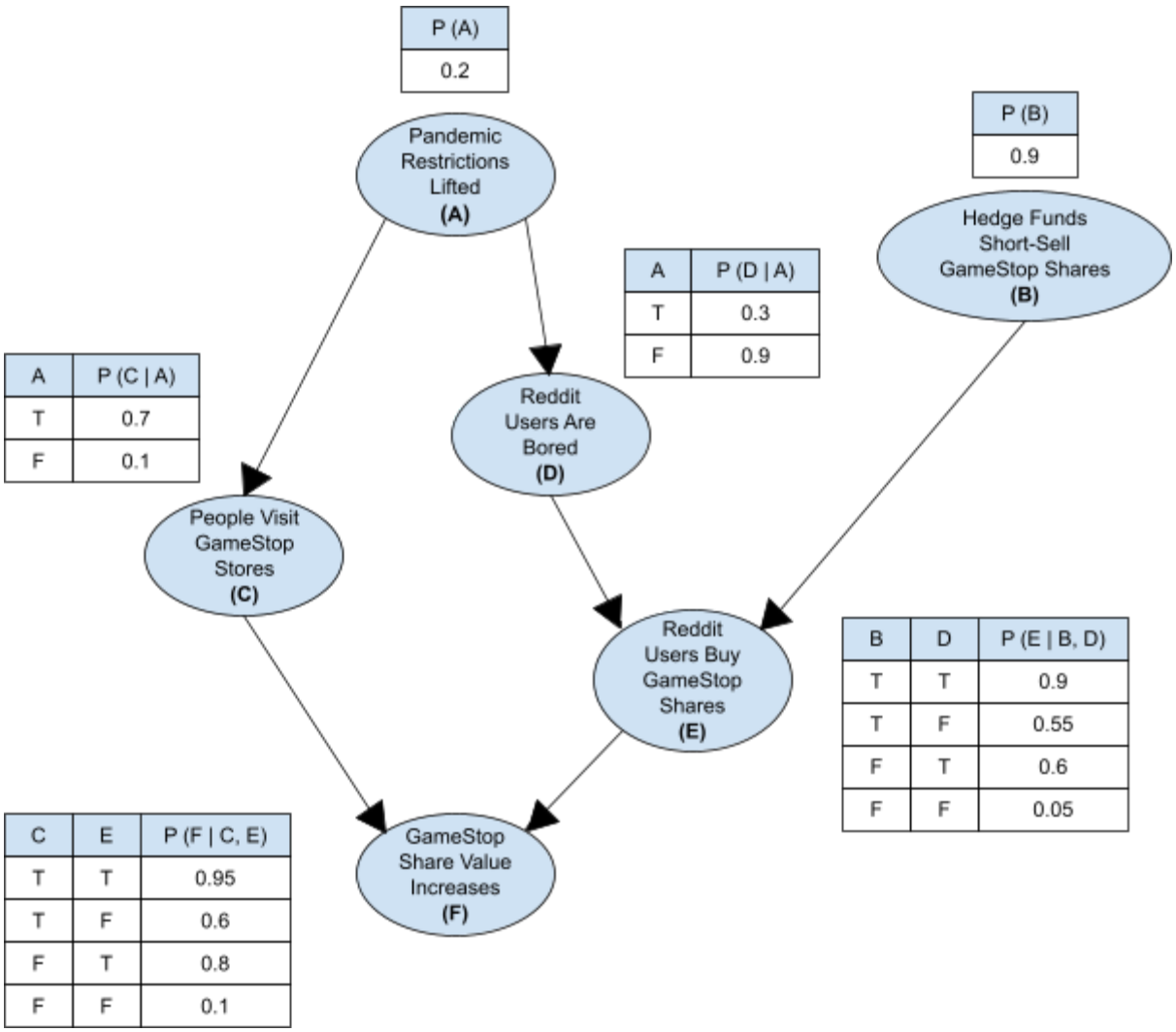
0.68

6. Bayes Nets

(15 points)

GameStop is a video game company which has been struggling for business since the pandemic hit. Recently, GameStop's stock prices surged because a bunch of people from Reddit decided to go against the big investors who made a profit from GameStop's misfortune.

This scenario can be represented by the following Bayes' Net.



6.1 Which of the following are a part of the Markov Blanket of **D**? Mark all that apply. **(3 points)**

- [x] A
- [x] B
- [ ] C
- [ ] D
- [x] E
- [ ] F

6.2 Which of the following are a part of the Markov Blanket of **E**? Mark all that apply. **(3 points)**

- [ ] A
- [x] B
- [x] C
- [x] D
- [ ] E
- [x] F

6.3 Which of these conditional independence relationships are true? Mark all that apply. Note that $P \perp Q \mid R, S$ means that P is conditionally independent of Q given R and S. **(4 points)**

- [x] $A \perp F \mid C, D, E$
- [x] $C \perp D \mid A$
- [x] $C \perp E \mid D$
- [x] $D \perp B \mid A$
- [x] $E \perp A \mid B, D$
- [ ] $F \perp A \mid C$
- [x] $F \perp A \mid C, E$
- [ ] $F \perp D \mid E$

6.4 What is the probability that the pandemic restrictions are lifted, given that Reddit users are bored, i.e., what is **P(A | ¬D)**? You must show your work below. You cannot use a Variable Elimination API to solve this (whether it was written by you or someone else). Please remember to round your answers as per the rounding rules of the exam. **(2 points)**

0.636364

Provide some ***brief*** reasoning to your answers in 6.4.

P(A|¬D) = mu*P(¬D|A)*P(A)
P(¬A|¬D) = mu*P(¬D|¬A)*P(¬A)

From Bayes' Net:
P(¬D|A) = 0.7
P(¬D|¬A) = 0.1
P(A) = 0.2
P(¬A) = 0.8

Thus,
mu = 4.545455
P(A|¬D) = 0.636364

6.5 What is the probability that Reddit users are bored given that people do not visit stores, i.e., what is **P(D | ¬C)**? You must show your work below. You cannot use a Variable Elimination API to solve this (whether it was written by you or someone else). Please remember to round your answers as per the rounding rules of the exam. **(3 points)**

0.853846

Provide some ***brief*** reasoning to your answers in 6.5.

Based on the Bayes' rule:
P(D|¬C) = P(D|¬C, A)*P(A|¬C) + P(D|¬C, ¬A)*P(¬A|¬C)

Because C,D is independant given A:
P(D|¬C) = P(D|A)*P(A|¬C) + P(D|¬A)*P(¬A|¬C)

P(A|¬C) = mu*P(¬C|A)*P(A)
P(¬A|¬C) = mu*P(¬C|¬A)*P(¬A)

From Bayes' Net:
P(¬C|A) = 0.3
P(¬C|¬A) = 0.9
P(D|A) = 0.3
P(D|¬A) = 0.9
P(A) = 0.2
P(¬A) = 0.8

Thus:
P(A|¬C) = 0.076923
P(¬A|¬C) = 0.923077
P(D|¬C) = 0.853846

Checklist

And now mark the checklist below making sure you have taken care of each of the points mentioned:

- ■ I have read the pinned Piazza post with the title ‘Midterm Exam Clarifications Thread’, and I am familiar with all of the clarifications made by the Teaching staff.
- ■ All answers with more than 6 digits after the decimal point have been rounded to 6 decimal places.
- ■ All pages are being uploaded in the correct order that they were presented to me, and none of the pages are missing/removed.
- ■ Any extra pages **(including blanks)** are only attached at the END of this exam, after page 28 with clear pointers to wherever the actual answer is in the PDF (reference properly).
- ■ I am submitting only one PDF and nothing else (no docx, doc, etc.).
- ■ The PDF I am submitting is not blank (unless I want it to be).
- ■ **I will go over the uploaded pictures on Gradescope and make sure that all the answers are clearly visible. I acknowledge that I am aware that dull / illegible / uneven scans will not be graded.**
- ■ I have submitted a copy of the PDF to Canvas.