CS6601 Midterm – Fall 2017

Please read the following instructions thoroughly.

Do not put your name or any personally identifiable information on this exam. The exam will be graded anonymously.

Fill out this PDF form and upload it to **Gradescope** when you are done. This is mandatory. You may also submit to T-Square so that we have a backup. You have unlimited resubmissions until the deadline. You can: **(a)** type directly into the form – we highly recommend using Adobe Reader DC (or Evince on Linux). 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. Short bullet points are appreciated over full sentences**. For questions that ask you to show your work, show it in the space provided and clearly indicate your final answer.**

Submit only a single PDF – no phone pictures, please! (You may use an app like Camscanner 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, **else there will be issues in grading.**

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. 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, "Midterm Exam #3").

Point breakdown:

Each question has sub-parts with varying points.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Total
Pts	**6**	**12**	**15**	**14**	**18**	**17**	**18**	**100**

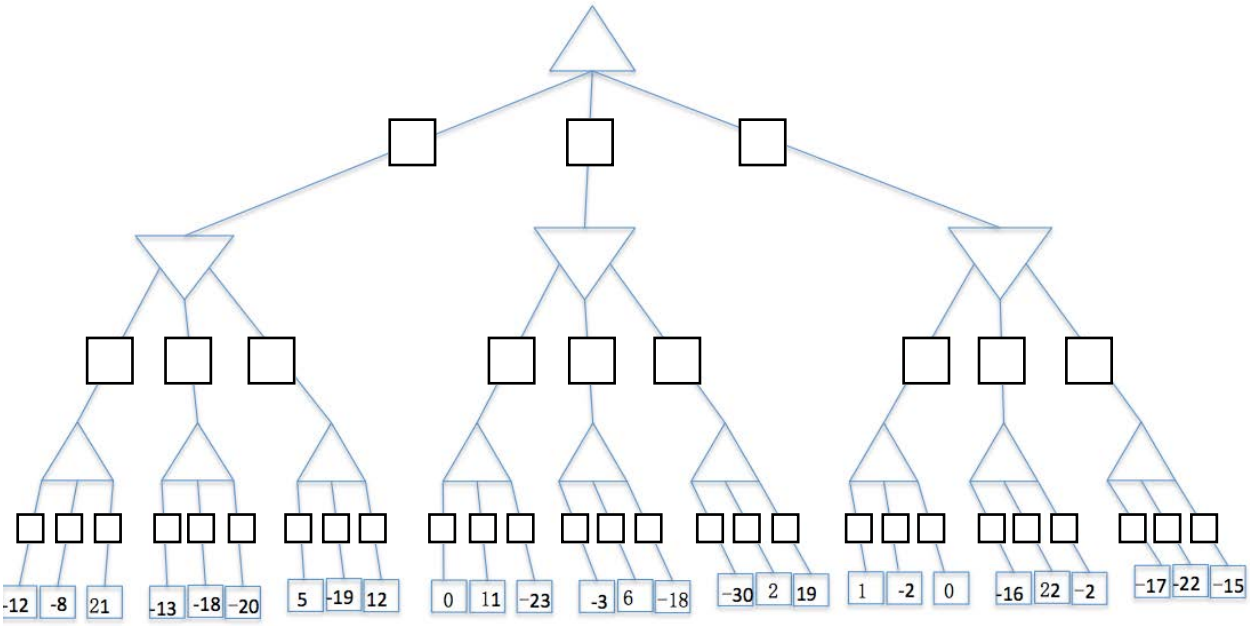
1. Game Playing

(6 points)

Question A.

(2 points)

Solve this game tree from left to right using MiniMax with alpha-beta pruning.
Show the value at each node. Make sure to use inequality signs wherever appropriate.
Check the boxes over branches that should be pruned.

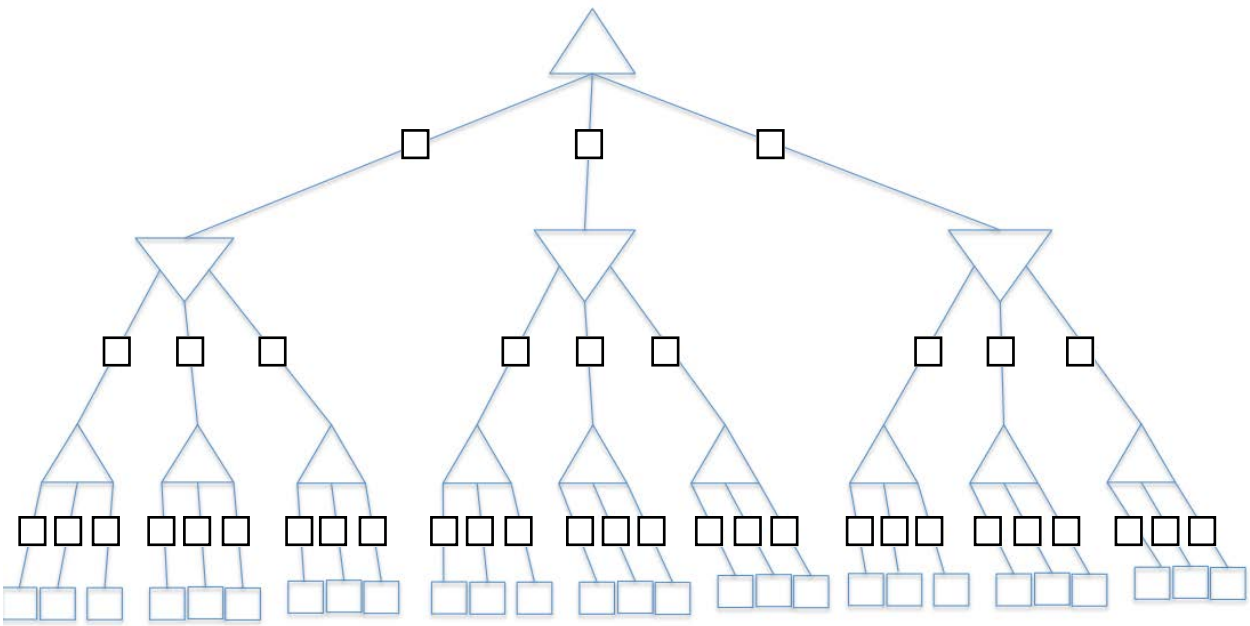


Question B.

(2 points)

Can the nodes be ordered in such a way that alpha-beta pruning can prune more nodes? If the answer is yes, describe your optimization method. Fill in the leaf nodes in the order that can cut off the maximum number of nodes and do alpha-beta pruning using this new node order. Otherwise, give your explanation.

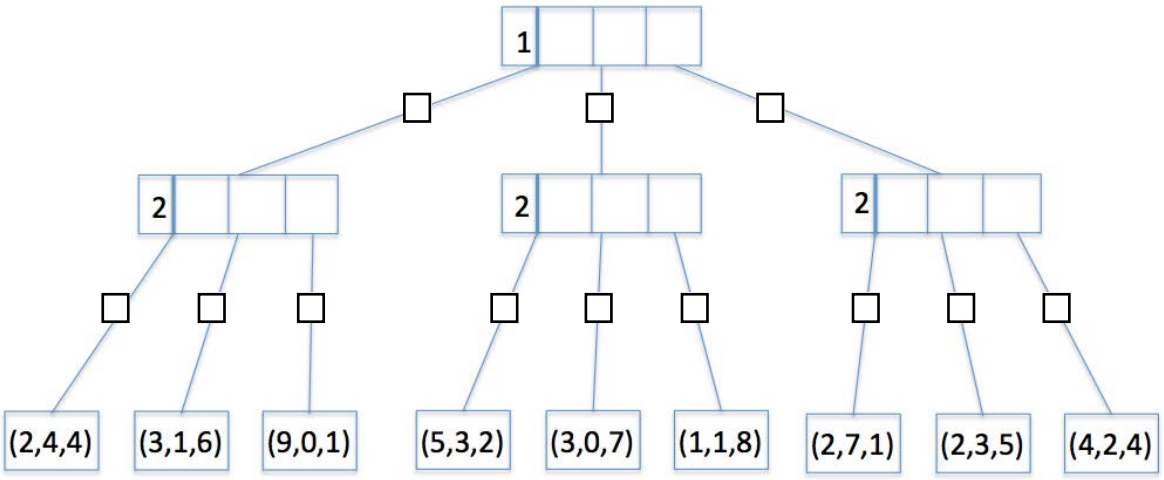
Optimization method / Explanation (less than 100 words):



Question C.

(2 points)

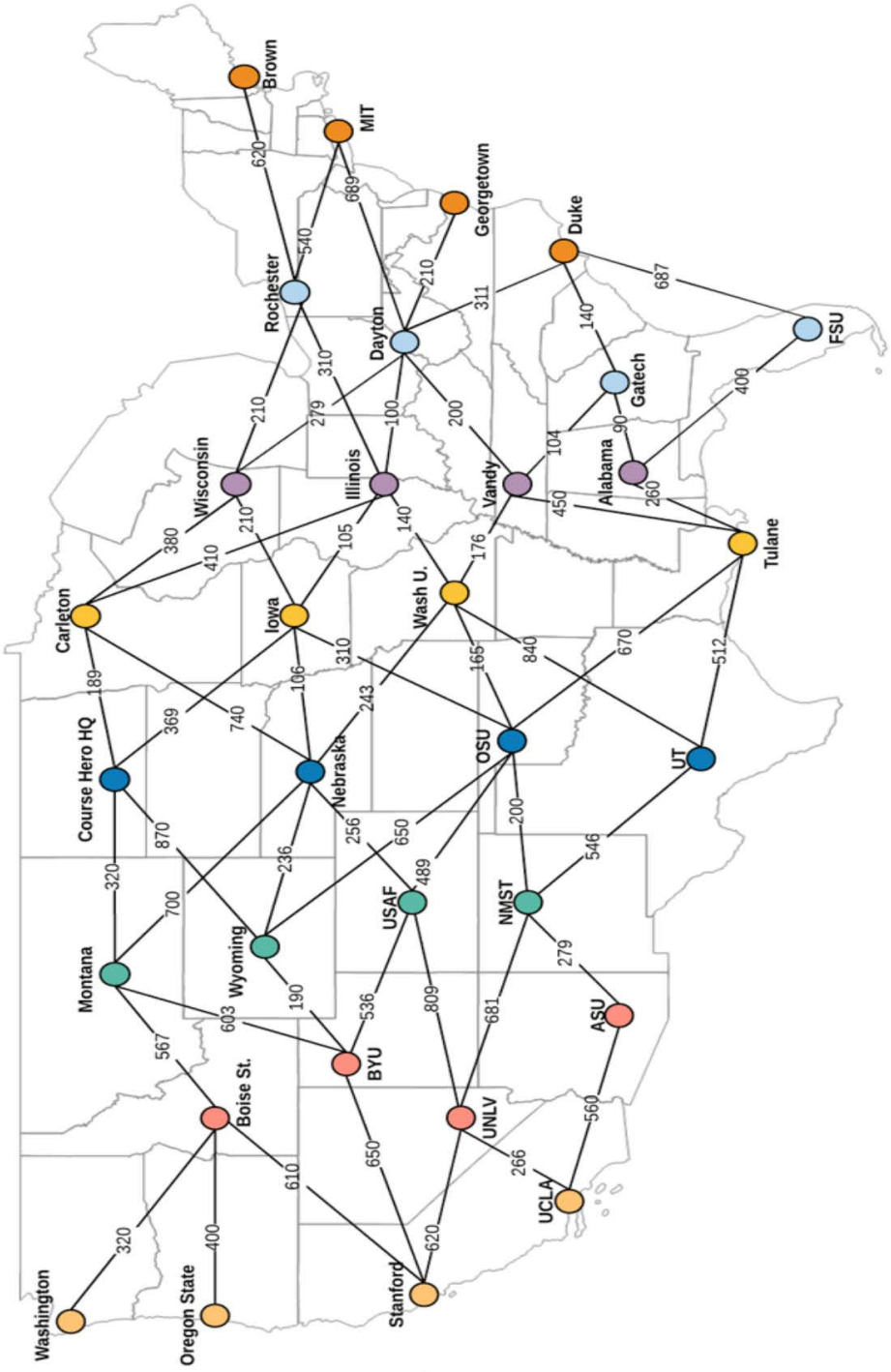
Solve this three-player game tree from left to right using alpha-beta pruning. Show the value of each node. Make sure to use inequality signs if appropriate. Check the boxes over branches that should be pruned. The 1, 2 at the front are the player numbers, indicating which player should be maximized. The maximum combined score is 10, which means the sum of all three scores is less than or equal to 10. All values are non-negative.



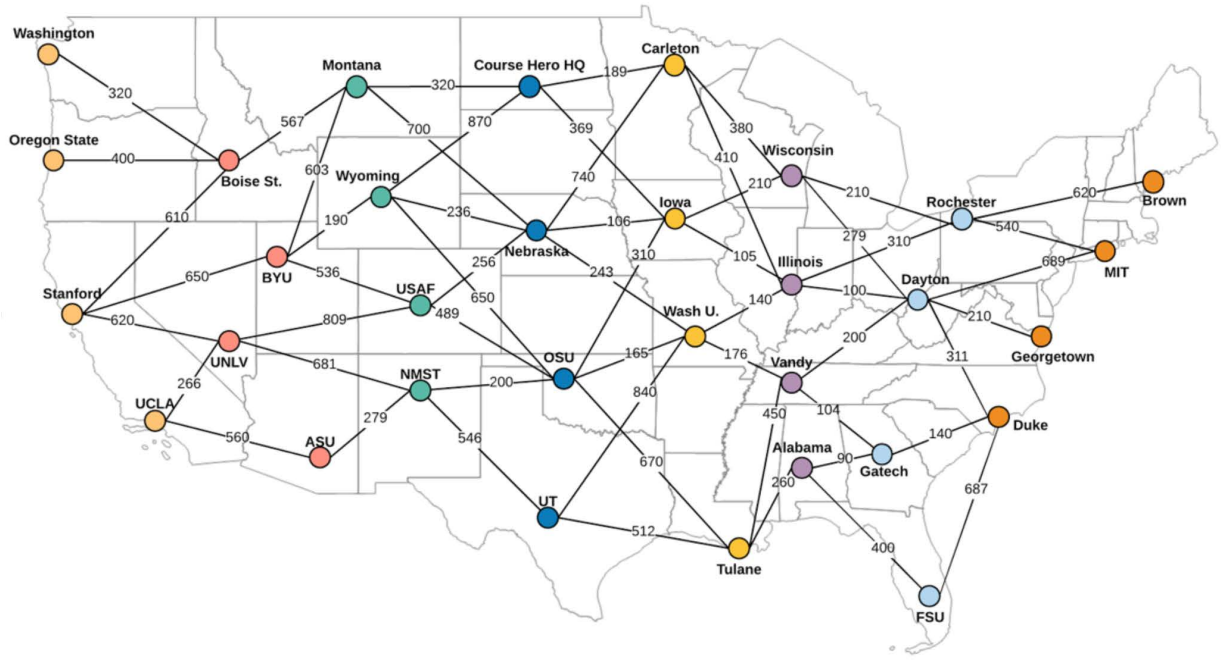
2. Search

(12 points)

For the following questions, please refer to the map below. A smaller version of this map is given on the next page.



Using the search techniques you have learned so far in this course, answer the questions that follow using the map of universities below. Note: the map is not to scale.



Question A.

(6 points)

Professor Starner is planning an epic 8-day cross country speaking tour which must start on the West Coast and end on the East Coast. He can visit one university per day. He must start from any university on the West Coast (Washington, Oregon State, Stanford, UCLA) and must end at any university on the East Coast (Brown, MIT, Georgetown, Duke). Costs to travel between two cities are shown on the edge between them. Perform Bidirectional UCS and answer the following questions. Keep 2 frontiers - one for the forward search (West coast to East coast) and one for the backward search (East coast to West coast). Start from the forward search and alternate between the two, exploring one node at a time.

1. What is the shortest path? (Write down all nodes from start to end.) **(1.5 points)**

2. Where do the two frontiers first overlap? (There is an overlap if A is being explored on one side, and B is already in the explored set of opposite side, and there is a direct edge between A and B. You need to report A.) **(1.5 points)**

---

3. What is the total cost of the shortest path? **(1.5 points)**

---

4. Clearly describe your stopping condition. Limit your response to 2 sentences or less. **(1.5 points)**

Question B.

(3 points)

Using unidirectional UCS, how many nodes are explored to find the optimal path from Gatech to OSU? What is the optimal path? Note that we are no longer planning a conference schedule.

Question C.

(3 points)

Now let's take a moment and appreciate what A* search does and how it helps us limit the number of nodes we expand when finding a path. Using unidirectional A* search and the table provided below for the heuristic, how many nodes are explored to find the optimal path from Gatech to OSU? What is the optimal path? **Please answer on the next page.**

School	Euclidean Distance to the OSU
Gatech	500
Duke	560
Vandy	400
Alabama	450
Dayton	575
FSU	800
Tulane	600
Wash U	100
Nebraska	250
UT	100
NMST	120
Iowa	400
Illinois	350

3. Optimization Algorithms

(15 points)

Question A.

(9 points)

You are building an AI agent that can pass as an art critic. Your agent analyzes images of what it sees and critiques them. As a starting point, you decide to find the pixel with the highest value in an image. A pixel's value is defined as the sum of squares of its RGB components. Each component is a value in the range [0, 255].

1. Assuming standard additive color used in today's computer graphics, what is the name of the color with the highest possible value? **(1 point)**

   _____

2. You first try hill climbing. Your agent starts at the top left of the image and goes through it pixel by pixel, left to right, top to down, flattening the image into a 1D array. It sees the following RGB values (starting at [22, 98, 75]). Using the column on the left, check which pixel is returned by hill climbing if it encounters these values? **(1 point)**

Checkboxes	R	G	B
◯	22	98	75
◯	22	100	90
◯	65	100	84
◯	100	67	90
◯	120	60	187
◯	150	80	200
◯	140	70	210
◯	230	230	230
◯	95	200	145
◯	153	197	214

3. You decide to try Simulated Annealing next. However, due to the nature of the algorithm, you now encounter pixels in a different order (top to bottom in the table below). What is the probability these values will be accepted by Simulated Annealing? Give your answer correct to 4 decimal places. You must compare a pixel with the value just before it, even if the previous value has partial probability. Do not compute a joint probability. **(3 points)**

R	G	B	Temperature	Probability?
22	98	75	100000	
22	100	90	90000	
150	80	200	80000	
65	100	84	70000	
230	230	230	60000	
120	60	187	50000	
100	67	90	40000	
153	197	214	30000	
95	200	145	20000	
140	70	210	10000	

4. You feel that using sum of squares isn't working out well for your agent. You decide to try a new objective function. You calculate the average Euclidean color difference for a pixel and all its neighbors.
(Wikipedia page on color difference)

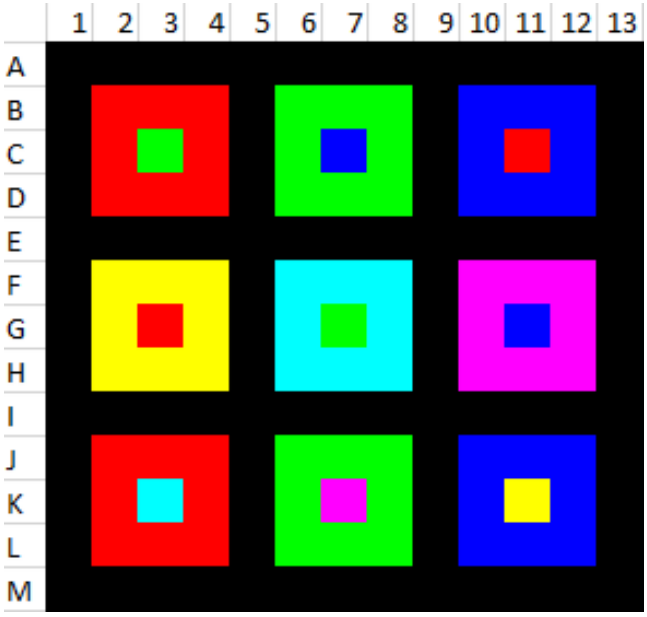
   The color difference between two pixels is defined as:

$$d_{xy} = \sqrt{\left(\mathrm{R_x} - \mathrm{R_y}\right)^2 + \left(\mathrm{G_x} - \mathrm{G_y}\right)^2 + \left(\mathrm{B_x} - \mathrm{B_y}\right)^2}$$

$$Objective\ function, f(x) = Average\ of\ d_{xy}\ \forall\ y \in neighbors(x)$$

   A pixel in 2D has 8 neighbors. (4 cardinal and 4 inter-cardinal). If any neighbors are not present (like at an edge), simply reduce the number of neighbors for that pixel. E.g. in the below image, A1 only has 3 neighbors: B1, B2 and A2. B2 has 8 neighbors: A1, A2, A3, B1, B3, C1, C2, C3

   For the below image, which coordinates would an optimal Simulated Annealing algorithm stop on? (If you think there is more than one possible answer, enter any one.) **(2 points)**



Color	RGB Value
	0, 0, 0
	255, 0, 0
	0, 255, 0
	0, 0, 255
	255, 255, 0
	0, 255, 255
	255, 0, 255

5. If you wanted to instead minimize the average color difference, how would you change the objective function? **(5 words or less) (1 point)**

_____

6. Which coordinates would be returned by minimizing the color difference? (If you think there is more than one possible answer, enter any one.) **(1 point)**

_____

Question B.

(6 points)

A struggling record label has seen one too many of its artists becoming one-hit wonders. The people in charge decide to take matters into their own hands. They are considering the prospect of automatically generating music, and you are assigned the task of using the genetic algorithm to accomplish this. All music can be generated using musical sequences named A-G, such that A has the lowest pitch and G has the highest pitch. You are free to decide the length of the song you make. Please note that these letters have nothing to do with the actual musical notes A-G.

1. Which of these are properties of the above problem that make genetic algorithm a good solution for this task? One or more options may be correct. **(2 points)**
   - ❑ The search space has a high degree of stochasticity.
   - ❑ The task at hand can be looked at as a combinatorial path planning technique, where each individual in the search space can be derived as a permutation of 2 or more parents.
   - ❑ There are many possible optimal solutions for this problem.
   - ❑ There is only one global optimum for this search space.
   - ❑ There are very few local minima.

2. You need to generate music such that all the notes need to be in an increasing order of pitch, i.e., ABCDEF is good and FEDCBA is bad. If your population is

   { GHFABB, DACFEC, FFFFAA, DDAADD }

   Is it possible to find an optimal solution without doing mutation? **(1 point)**
   - ○ Yes
   - ○ No

Let's say that you need to generate music which is a combination of two different genres, where both genres must be intertwined throughout the song. There are two methods proposed to you.

In the first method, you have to use two fitness functions (f1 and f2), one for each genre, and during the crossover phase, you perform crossover between the top n/2 candidates sorted based on f1, and the top n/2 candidates sorted based on f2.

In the second method you have to use the genetic algorithm with just one fitness function, which is a combination of f1 and f2.

3. Which of the following statements are true? One or more options may be correct. **(2 points)**
   - ❑ Both methods will produce the same result.
   - ❑ Method(1) may get you an optimal solution, but is not guaranteed to find one.
   - ❑ Method(2) may get you an optimal solution, but is not guaranteed to find one.
   - ❑ Method(1) will definitely give you an optimal solution.
   - ❑ Method(2) will definitely give you an optimal solution.

4. Let f be a fitness function that is being optimized by Simulated Annealing or the Genetic Algorithm. Which algorithm would perform better if you drastically change f to $1/f$ or $-f$ halfway through the algorithm? **(1 point)**
   - ◯ Simulated Annealing
   - ◯ Genetic Algorithm
   - ◯ Both algorithms would perform equally well

4. Constraint Satisfaction Problems

(14 points)

Question A.

(14 points)

At Georgia Tech, the newly formed Department of Constraint Science needs to schedule classes for the upcoming term.

They have hired 3 professors: Professor X, Professor Y, and Professor Z, who between them must teach five classes. Professor X teaches at another school *in the morning* so he'll only be teaching one class *in the afternoon*. Their class assignments have been formulated as a Constraint Satisfaction Problem with the following variables that represent the course number:

$$X = \{X_1, Y_1, Y_2, Z_1, Z_2\}$$

In the above assignments,$Y_1$and $Z_1$are in the morning while $Y_2$and $Z_2$are taught in the afternoon. The following classes need to be taught:

- CON101: Introduction to Constraints
- CON123: Introduction to Forward Checking
- CON321: Advanced Backtracking
- CON333: Consistency and You
- CON601: Graduate Consistency Graphing

The domain of possible values for each assignment is the set of course number suffixes for these five classes:

$$D_i = \{ 101, 123, 321, 333, 601 \}$$

The assignments of classes to professors must conform to the following constraints:

- All five classes must be assigned to one of the open slots.
- The department has found freshmen are up earlier than other students, so a professor's afternoon class must be a higher course number than their morning class
- Since CON 601 and CON333 are often taken concurrently, they cannot both be morning classes or both be afternoon classes.
- Professor Y refuses to teach CON123 or CON321.
- Professor X does not teach CON601.
- Professor Z does not teach CON101.

1. After unary constraints are applied, what are the domains of each variable? Write down the domains corresponding to each variable.
   (example: 101, 123, 321, 333, 601)
   (2 points)

X1	
Y1	
Y2	
Z1	
Z2	

2. Prof Y chooses class CON333. Ensure forward checking with the above unary constraints and write down the domains in front of each variable (Choose alphabetically if equivalent). **(5 points)**

X1	
Y1	
Y2	
Z1	
Z2	

3. Do you need backtracking to solve the problem after forward checking is ensured? **(2 points)**
   - ❑ Yes
   - ❑ No
   - ❑ Can't say

4. Draw the constraint graph for this problem. **(5 points)**

5. Probability

(18 points)

Question A.

(9 points)

The AI class in the Georgia Institute of Wearable Technology has given out three Homeworks this semester. Out of 448 students, 349 got an A for HW1 while 308 students got an A for HW2. The rest got a B. Furthermore, we know that 267 students got an A for both HW1 and HW2.

1. How many students got a B on both assignments? **(1.5 points)**

   ---

2. How many students got different grades on the first two assignments? **(1.5 points)**

   ---

3. Suppose that, for any given student, he or she spends the same amount of time on each assignment (i.e., Student #1 spends 10 hours on HW1, 10 hours on HW2 and so on; Student #2 spends 16 hours on HW1, 16 hours on HW2, and so on; etc.).
   A survey was given after HW1 and HW2 were completed, asking each student the amount of time they spent on the assignments.
   We found students spending 15 hours or more on each assignment have an 80% chance of getting an A and 20% of getting B for that assignment, while a student spending less than 15 hours has a 50% chance of getting an A and 50% chance of getting a B.
   Based on the above statistics, what is the expected number of students scoring an A in HW3? (Round to the nearest whole number.) **(3 points)**

   ---

4. We sent out another survey for HW3 and found the following: 12.5% of students who had previously spent more than 15 hours on their assignments and got an A for their HW2 have slacked off, so they ended up spending less than 15 hours on HW3, while 20% of the students who previously spent less than 15 hours on their assignments and got B for their HW2, have started working harder, so they spent more than 15 hours on HW3. What is the expected number of students scoring an A in HW3? (Round to the nearest whole number.) **(3 points)**

Question B.

(9 points)

For this section (Question B), please give answers with at least 3 significant figures.
A normal deck of cards has 52 cards, comprised of 13 cards of each suit (spades, hearts, diamonds, clubs). You are given a deck of cards, but some of the cards are missing. The deck only has 45 cards in it.

1. If you draw a card from the deck, what is the probability that it is a spade? **(0.5 points)**

   _____

2. If you draw 3 cards from the deck, what is the probability that they are all spades? **(1 point)**

   _____

3. You've discovered that all 7 of the missing cards are hearts. Now what is the probability that drawing 3 cards gives you all spades? **(1 point)**

   _____

You are given four eight-sided dice. Each die is labeled with numbers from 1 to 8 that are all equally likely to be face up when rolled.

4. What is the probability that, when you roll the dice, they all show 1's? **(0.5 points)**

   _____

5. What is the probability that they are all the same number? **(1 point)**

   _____

6. If you use a single 8-sided die, fill out the probabilities of the following events in the table below: **(4 points)**

Roll a number greater than 1	
Roll a 1, then roll a number greater than 1	
Roll 1 twice in a row, then roll a number greater than 1	
Roll 1 three times in a row	

7. With a single 8-sided die, what is the expected number of rolls to make before you roll a 1 three times in a row? **(1 point)**

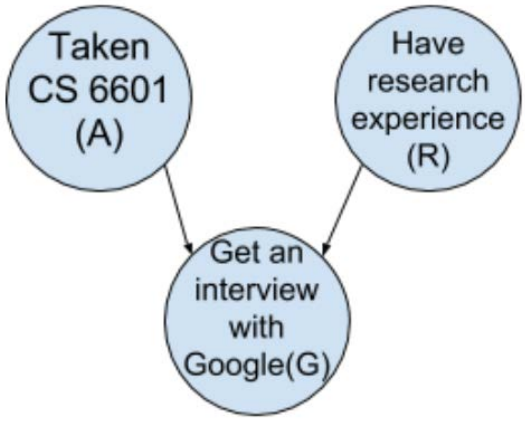
_____

6. Bayes Nets and D-Separation

(17 points)

Question A.

(2 points)



Given that
P(A) = 0.5
P(R) = 0.4
and the following table:

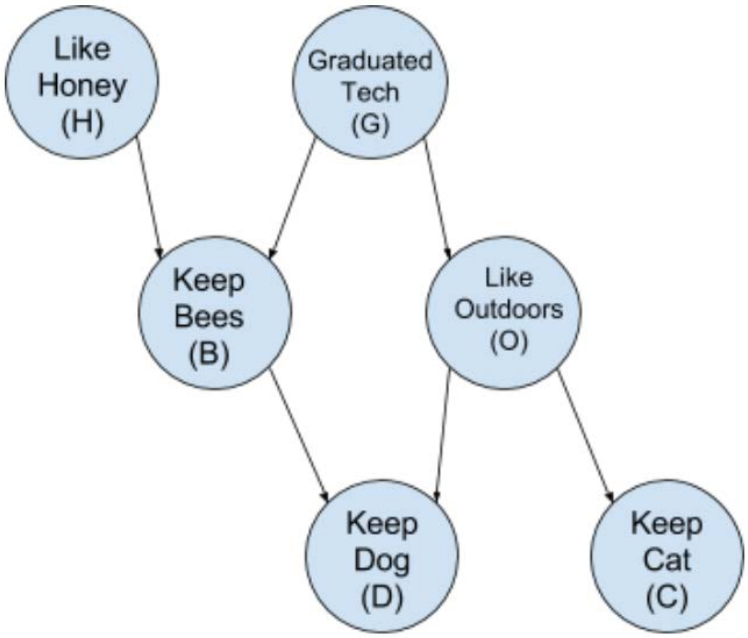
A	**R**	**P(G)**
T	T	0.8
T	F	0.6
F	T	0.7
F	F	0.5

Fill out the full joint distribution table using the given information.

	A		**~A**	
	R	**~R**	**R**	**~R**
G				
~G				

Question B.

(4 points)



Given the following probabilities:

P(H) = 0.4
P(G) = 0.25

H	**G**	**P(B)**
T	T	0.65
T	F	0.45
F	T	0.25
F	F	0.08

G	**P(O)**
T	0.55
F	0.45

B	**O**	**P(D)**
T	T	0.35
T	F	0.15
F	T	0.62
F	F	0.28

O	**P(C)**
T	0.25
F	0.45

1. What is the P(Like Outdoors = T | Keep Bees =T) ?
   Do this by hand and show your equations; give the answer correct to 6 decimal places.
   (1 point)

   Final Answer: _____

   Equations:

2. Given that Keep Bees = True, are Like honey and Graduated Tech independent of each other? **(1 point)**
   - ❏ Yes
   - ❏ No

3. Suppose we do not know if any nodes are True or False (in other words, you do not have any evidence nodes), which has the higher probability: Keep Dog or Keep Cat? **(1 point)**
   - ◯ Dog
   - ◯ Cat
   - ◯ Equal

4. What is the absolute difference of the two probabilities? **(1 point)**
   Give a non-negative number to represent the absolute value |P(D) - P(C)|.
   Do this by hand and show your equations; give the answer to 6 decimal places.

   Final Answer: _____

   Equations:

Question C.

(11 points)

(If your answer has more than 6 decimal places, please give it correct to 6 decimal places.)

The war between the Democratic Republic of Starneria and the Autocratic Dictatorship of Isbelland continues with no end in sight. Starneria's intelligence agency, AI-6, continues to try to collect intelligence on the schemes planned by Isbelland's ML-4.

The President of Starneria (hereby called "The President") continues to govern in peace, aided by his trusty wearable computer that gives him a steady flow of intelligence from AI-6 and keeps him aware of the Dictator of Isbelland ("The Dictator")'s attempts to overthrow him.

In a desperate attempt to end the peaceful presidency, The Dictator has sanctioned the development of an Electromagnetic Pulse (EMP) Weapon, which he hopes will disable all electronics in the capital of Starneria, including The President's wearable. AI-6 is aware of this, of course (AI-6 sees everything), and are trying to infiltrate ML-4's weapons division so that they can sabotage the weapon. What AI-6 doesn't know, however, is that Isbelland suffers from a severe dearth of scientists proficient in weapons technology, and there is no guarantee that they can make this weapon work.

The Dictator is obviously worried about the potential failure of his weapon, so he has his scientists give him the run-down on their chances of making it work. Here's what they tell him:

- There's only an 80% chance ML-4 can find scientists good enough to make a working weapon.
- There's also a 60% chance ML-4 has been infiltrated by spies from AI-6 already.
- With competent scientists and no AI-6 spies, there's a 90% chance the weapon will work. However, if ML-4 is infiltrated, this drops to 70%.
- With incompetent scientists and no AI-6 spies, there's a 40% chance they can hack up a working weapon; however, this drops to 20% if they've been infiltrated.

1. The Dictator is confused. Probability was never his strong suit; he was always more of a genetics kinda guy. He says, "These numbers confuse me. Just give it to me straight. What is the probability ML-4 can deploy a working weapon?"
   Give him the correct answer below. **(1 point)**

   _____

2. The Dictator then says, "Okay, that doesn't sound too bad. What is the probability that we find competent scientists and manage to avoid being infiltrated, but still can't get the weapon working?"
   Give him the correct answer below. **(1 point)**

   _____

Given these numbers, The Dictator seems to like his chances, and asks them to go ahead with building the weapon.

A few months later, the head of ML-4 ("The Head") tells The Dictator the weapon is ready. The Dictator promptly asks him to launch it at Starneria City, and he does. As they wait to find out whether it worked or not, The Head reveals to The Dictator that they may have underestimated the power of the weapon, and that the effects may be seen as far away as… Isbelland City.

THE DICTATOR: "But… That's where *we* are! What effects would it have?"
THE HEAD: "Respected Leader… It might disable most advanced technology here as well. We would need to resort to using primitive devices."
THE DICTATOR: "What sort of advanced technology are we talking about?"
THE HEAD: "Benevolent General… There is a chance your electric car will stop functioning. You will need to resort to a much older vehicle for transportation."

In fact, The Head goes on to detail the following scenario:

- If the weapon is successful, there is an 85% chance it will disable The President's wearable, but also a 45% chance it will disable The Dictator's car.
- Due to the extensive amount of experimental technology present in The President's wearable, there is a 12% chance it will fail even if the weapon fails.
- Since The Dictator's car is notoriously unreliable, there is a 25% chance it will stop working even if the weapon fails.

The Dictator is, understandably, not pleased. ML-4 is a mess! They might be compromised, they find it hard to hire competent scientists, they still don't know if the weapon worked, they might have messed up his car, and they keep talking to him in numbers! He demands to know whether ML-4 has been infiltrated or not.

The Head says, "Exalted Lordship, perhaps I can set a test that must be taken by all our scientists. It will tell us whether we have hired competent scientists or not. That will give us a better idea of whether ML-4 has been compromised."

3. Is he right? **(0.5 points)**
    - ❏ Yes
    - ❏ No

A few seconds later, The Dictator's chauffeur bursts into the room and says, "Oh Iron-Willed Commander of the Glorious Nation, your car has suddenly stopped functioning, and I do not know why!" The Dictator turns to The Head and says, "Have your men administer the test. Get me the results. It will tell me more about whether there are spies in ML-4."

4. Is he right? **(1 point)**
    - ❑ Yes
    - ❑ No

5. The Dictator then wonders, “Given that I know all this, what are the chances the weapon actually worked?” Give him the answer below. **(1 point)**

    _____

6. A member of ML-4 then comes in and whispers something in The Head’s ear. The Head turns to The Dictator and says, “Peerless Politician, it appears our weapon has not worked. However, we know that your vehicle no longer works, so perhaps that give us some insight into whether we have disabled the enemy’s wearable!” Is he right? **(0.5 points)**
    - ❑ Yes
    - ❑ No

Thoroughly disappointed, The Dictator asks everyone to leave the room, and decides to analyze what went wrong, and what the chances were for each event occurring. He prepares a questionnaire, and hires you to do the math for him:

7. Assume that we administered the competence test before we even started on the weapon. We do not know the outcome of any other events. **(1.5 points)**
    a. What would have been the probability ML-4 was infiltrated, given that we do not know the results of the test?

        _____

    b. What would have been the probability ML-4 was infiltrated, given that the scientists were competent?

        _____

    c. What would have been the probability ML-4 was infiltrated, given that the scientists were incompetent?

        _____

8. Assume that we found the weapon to be successful. We do not know the outcome of any other events. **(1.5 points)**
    a. What is the probability ML-4 was infiltrated?

    _____

    b. What is the probability ML-4 was infiltrated, given the scientists were competent?

    _____

    c. What is the probability ML-4 was infiltrated, given the scientists were incompetent?

    _____

9. Assume that we do not know the outcomes of any events. **(1.5 points)**
    a. What is the probability The Dictator's car would have stopped working?

    _____

    b. What is the probability the car would have stopped working, given The President's wearable stopped working?

    _____

    c. What is the probability the car would have stopped working, given The President's wearable continued to work?

    _____

10. Assume that we know the weapon worked. **(1.5 points)**
    a. What is the probability the car would have stopped working?

    _____

    b. What is the probability the car would have stopped working, given The President's wearable stopped working?

    _____

    c. What is the probability the car would have stopped working, given The President's wearable continued to work?

    _____

7. Machine Learning

(18 points)

Question A.

(9 points)

The nations of Topisbell and Topistarn have had disagreements for a long time over the similarity of their names. Recently, someone proposed that a change be made to each in order to better differentiate the two: one will be given the prefix "U-", and the other "Dys-". Of course, the two nations couldn't agree which prefix each should have - and so they decided to have a vote between ambassadors from the largest nations: Topisbell, Topistarn, and Goeland.

There's a lot of speculation leading up to the vote as to what each nation will be called. Reading the news one day, you notice an advertisement for a job predicting the results of the vote. Since you just learned a number of machine learning methods in AI, you think this is a great opportunity to apply what you know!

Talking with the hiring company, you learn that they have historical information for prior ambassadors. The table below summarizes this information, where each row in a nation's column is an ambassador (6 ambassadors per nation, 18 in total). Ambassadors can be assumed not to hold any allegiance to their own nation when voting (i.e. nationality is independent of the vote).

F_I= Has family in Topisbell (N=0, Y=1)
F_S= Has family in Topistarn (N=0, Y=1)
V= Vote cast in favor of Topistarn getting a "U-" prefix (N=0, Y=1)

Topisbell			**Topistarn**			**Goeland**		
F_I	F_S	V	F_I	F_S	V	F_I	F_S	V
1	0	0	1	0	1	0	0	0
1	1	0	1	1	0	0	1	0
0	1	1	0	1	0	1	0	1
1	0	1	0	0	1	1	1	1
1	1	0	1	1	1	1	0	1
0	1	0	0	1	1	0	1	0

1. The company gives you the information on the current three ambassadors, and asks you to determine the probability for each ambassador of voting in favor of Topisbell ($P(V = 0 \mid F_I, F_S)$) using Bayes Rule (maximum a posteriori (MAP) without argmax). Be sure to include denominators, and show your work. **(4 points)**

Topisbell			**Topistarn**			**Goeland**		
F_I	F_S	V	F_I	F_S	V	F_I	F_S	V
1	1		1	0		0	0	

2. After giving them your results, they clarify that an ambassador having family in a nation is independent of having family in the other nation (i.e. $P(F_I, F_S) = P(F_I)P(F_S)$). Calculate the likelihoods in this situation using Naïve Bayes, this time giving the more likely vote (1 or 0), and showing your work. **(4 points)**

Topisbell			**Topistarn**			**Goeland**		
F_I	F_S	V	F_I	F_S	V	F_I	F_S	V
1	1		1	0		0	0	

3. What is most likely to be the new name of Topistarn? **(1 point)**

_____

Question B.

(9 points)

1. (Decision Trees) For the following question, consider the data given below.

Examples	Input Attributes			Goal
	A	B	C	
x_1	Yes	Yes	Yes	y_1= α
x_2	No	Yes	Maybe	y_2= γ
x_3	No	Yes	No	y_3= γ
x_4	Yes	No	Yes	y_4= α
x_5	Yes	No	No	y_5= α
x_6	Yes	Yes	Maybe	y_6= β

a. Calculate the information gain for each of the input attributes, accurate to 3 decimal places. **(3 points)**

Gain (A)	
Gain (B)	
Gain (C)	

b. Draw the optimal decision tree. **(4 points)**

2. Overfitting does happen in decision trees and a process called decision tree pruning is often used. **Use your own words** to describe the process of decision tree pruning. (no more than 50 words) **(2 points)**

Checklist

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

- ❏ All pages are being uploaded in the correct order that they were presented to you.
- ❏ Any extra pages are only attached at the END of this exam, after page 35 with clear pointers to wherever the actual answer is in the PDF (reference properly).
- ❏ Are you submitting only one PDF and nothing else (no docx, doc, etc.)?
- ❏ Did you avoid putting your name on the exam?
- ❏ Is the PDF you are submitting not blank (unless you 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.**