6.034 Quiz 2

18 October 2017

Name	King Kandy
Email	kkandy@mit.edu

For 1 extra credit point: Circle the TA whose recitations you attend, so that we can more easily enter your score in our records and return your quiz to you promptly:

Suri Bandler	**Erin Hong**	**Samarth Mohan**
Jake Barnwell	**Nathan Landman**	**Michael Shum**
Abigail Choe	**Amanda Liu**	**Jackie Xu**
Francesca Cicileo	**Nick Matthews**	

Problem number	Maximum	Score	Grader
1 - ID Trees	30		
2 - kNN	30		
3 - Constraints	40		
Total	**100**		

SRN	6		

Survey

We are investigating the false-positive rate for SRN questions. We will **not** consider your answer to this question when determining grades. Out of 5 SRN lectures (Sussman, Boyden, Kanwisher, line drawings, deep nets), how many did you attend?

6

There are 18 pages in this quiz, including this one, but not including tear-off sheets. A tear-off sheet with duplicate drawings and data is located after the final page of the quiz.

This quiz is open book, open notes, open just about everything, including a calculator, but no computers.

Problem 1: Identification Trees (30 Points)

Part A: Run, Barry, Run! (25 points)

Barry Allen, from the TV show *The Flash*, has reset the timeline again, and is having trouble identifying which of two parallel earths people are from: *Earth 1* or *Earth 2*. He has gathered information on seven (7) individuals about their **STATUS**, **POWERS**, **EVIL** nature, and **BACKSTORY**. Having taken 6.034, you offer your assistance by creating an identification tree which Barry can use to identify which earth people are from.

For your convenience, a copy of this data is provided on a tear-off sheet at the end of the quiz.

Name	Earth	Status	Powers	Evil	Backstory
Barry Allen	1	Alive	Yes	No	Tragic
Caitlin Snow	1	Alive	Yes	No	Tragic
Iris West	1	Alive	No	No	Neutral
Lord Savitar	1	Dead	Yes	Yes	Tragic
Detective Joe	2	Dead	No	No	Neutral
Killer Frost	2	Dead	Yes	Yes	Neutral
Oliver Queen	2	Dead	No	No	Tragic

A1 (8 points) Compute the average test disorder of each of the following feature tests. Use the table of logarithms on the next page to express your answer as **sums and products of decimals and fractions only.** Your final answer should contain **no logarithms**. *Space is provided on the next page for you to show your work.*

Feature Test	Disorder (as sums and products of decimals and fractions)
Evil	$\frac{2}{7} + \frac{5}{7}(0.97) \quad = .977$
Powers	$\frac{4}{7}(0.8) + \frac{3}{7}(0.9) \quad = .842$

You can earn partial credit for part A1 by showing your work here:

$$\text{disorder}_{evil} = \frac{2}{7}\cdot 1 + \frac{5}{7}\left(-\frac{2}{5}\lg\frac{2}{5} - \frac{3}{5}\lg\frac{3}{5}\right) = \frac{2}{7} + \frac{5}{7}(0.97)$$

$$\text{disorder}_{powers} = \frac{4}{7}\left(\frac{-3}{4}\lg\frac{3}{4} - \frac{1}{4}\lg\frac{1}{4}\right) + \frac{3}{7}\left(\frac{-1}{3}\lg\frac{1}{3} - \frac{2}{3}\lg\frac{2}{3}\right)$$

$$= \frac{4}{7}(0.8) + \frac{3}{7}(0.9)$$

$-\left[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right] = 1$ $\quad -\left[\frac{2}{5}\log_2\frac{2}{5} + \frac{3}{5}\log_2\frac{3}{5}\right] \approx 0.97$ $\quad -\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right] \approx 0.9$

$-\left[\frac{1}{4}\log_2\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4}\right] \approx 0.8$ $\quad -\left[\frac{1}{5}\log_2\frac{1}{5} + \frac{4}{5}\log_2\frac{4}{5}\right] \approx 0.72$ $\quad -\left[\frac{1}{6}\log_2\frac{1}{6} + \frac{5}{6}\log_2\frac{5}{6}\right] \approx 0.65$

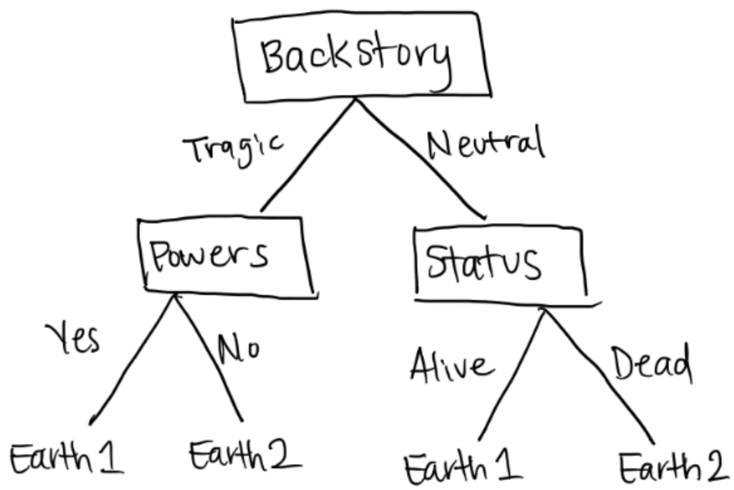
A2 (14 points) Cisco from STAR labs gets a vibe that you need advice, so he calls and tells you to draw the ID Tree starting with **BACKSTORY** as your first feature test. Before Cisco can give you a horrible nickname, you agree.

Regardless of your answers in part A1, follow Cisco's advice and design an ID Tree starting with the **BACKSTORY** feature test. Choosing from the four feature tests in the table above, construct the complete greedy disorder-minimizing ID Tree which classifies people as being from *Earth 1* or *Earth 2*. Break ties alphabetically:

BACKSTORY > EVIL > POWER > STATUS

Space is provided on the next page for you to show your work for partial credit.

For credit, draw your disorder-minimizing ID Tree in this box:



You can earn partial credit for part A2 by showing your work here:

Left branch

Status

A D

11 12

$d = \frac{1}{2}$

Powers

Y N

111 2

$d = 0$

Evil

Y N

1 112

$d = \frac{3}{4} \cdot 0.9 = 0.67$

Right branch

Status

A D

1 22

$d = 0$

Powers

Y N

2 12

$d = \frac{2}{3} \cdot 1 + 0 = \frac{2}{3}$

Evil

Y N

2 12

$d = 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$

A3 (3 points) Barry asks you to help classify Harrison Wells as being from *Earth 1* or *Earth 2* using the following **new** data point. Below, circle the **one** best answer indicating how your ID Tree from part A2 above classifies Harrison Wells.

Name	Earth	Status	Powers	Evil	Backstory
Harrison Wells	?	Dead	No	No	Tragic

Earth 1 ***Earth 2*** **CAN'T TELL**

Part B: Time Wraiths (5 points)

The producers of the show are concerned that having a high disorder for the **STATUS** feature test will cause reviews to fall. They decide to re-classify Iris West's STATUS so as to **make the average test disorder for STATUS as low as possible**. **Using the following simplified table of data**, how should Iris West's STATUS be classified in order to minimize the average test disorder for STATUS?

Name	Earth	Status
Bartholomew	2	Alive
Iris West	1	?
Lord Savitar	1	Dead
Killer Frost	2	Dead

Below, circle the **one** best answer indicating what Iris West's STATUS should be to minimize the STATUS test disorder. If it doesn't matter if she is **Alive** or **Dead**, circle **DOESN'T MATTER**. If you can't tell without more information, circle **CAN'T TELL**.

Alive **Dead** **DOESN'T MATTER** **CAN'T TELL**

You may show your work below for partial credit:

Iris = Alive

$$\text{disorder}_{\text{status}} = \frac{2}{4}\left(\frac{-1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) + \frac{2}{4}\left(\frac{-1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) = 1$$

Iris = Dead

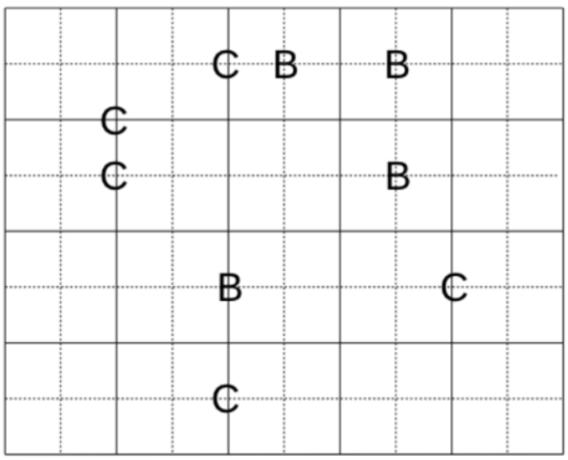
$$\text{disorder}_{\text{status}} = \frac{1}{4}(0) + \frac{3}{4}\left(\frac{-1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}\right) = 0.9$$

Problem 2: k-Nearest Neighbors (30 points)

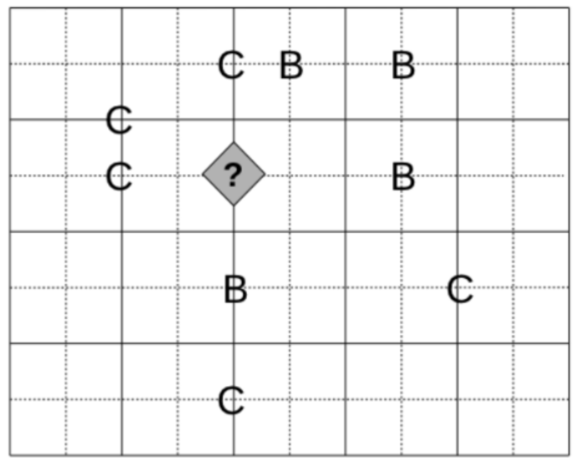
Part A: Unbeetable Stew (24 points)

Karen, Patrick Winston's wife, is making stew for dinner, and she asks Patrick to pick some vegetables from their backyard garden. She gives him a sketch of where she planted the **C**ARROTS and **B**EETS, insisting that he pick **all** of the **CARROTS** and **none** of the **BEETS**. Unfortunately, when Patrick goes to the garden, he finds that there is a mysterious vegetable that Karen didn't label on her sketch!

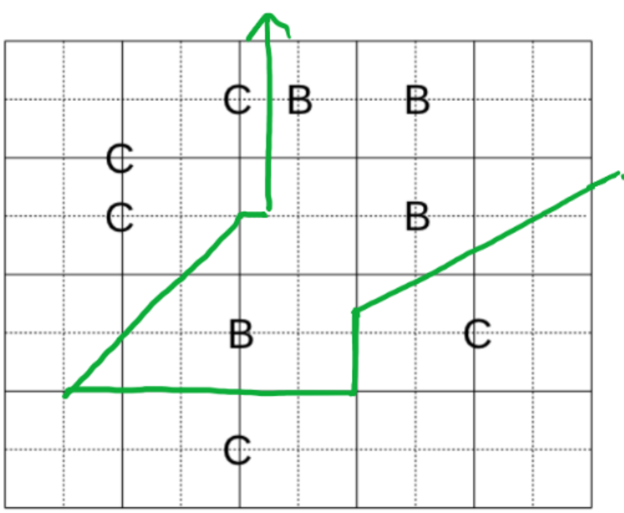
Karen's garden sketch:



What Patrick sees in the garden:

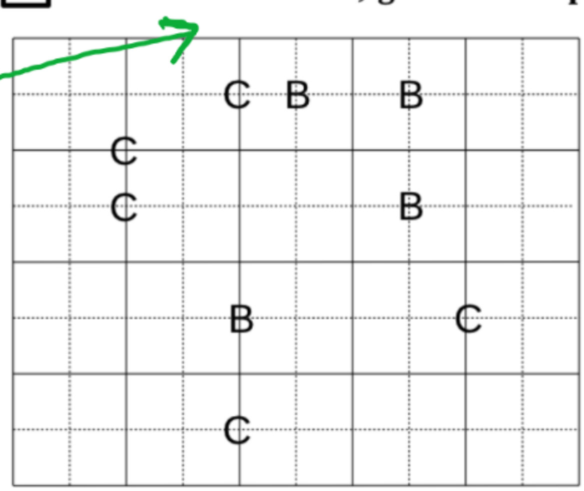


A1 (10 points) Patrick knows he can use 1-Nearest Neighbors (with Euclidean distance) to classify the mystery vegetable as either a CARROT or a BEET. Help him by **drawing the 1-nearest neighbors decision boundary** on the graph below:



This is a duplicate copy of the graph; if you want this copy to be graded instead, check the box:

☐ **I want to start over; grade this copy**



the boundary going off to the right will bend slightly due to the B in the top right, but you do not need to account for it to receive full credit.

A2 (2 points) Based on your decision boundary for k = 1 in **Part A1** above, how is the mystery vegetable classified? Circle the **one** best answer.

CARROT　　BEET　　CAN'T TELL

A3 (8 points) Now Patrick tries different k values with nearest neighbors and Euclidean distance. For each value of k, circle the **one** best classification for the mystery vegetable.

k = 5	CARROT	BEET	CAN'T TELL
k = 7	CARROT	BEET	CAN'T TELL

A4 (4 points) If Patrick instead uses **Manhattan distance with k = 7**, what will be the classification of the mystery vegetable? Circle the **one** best answer.

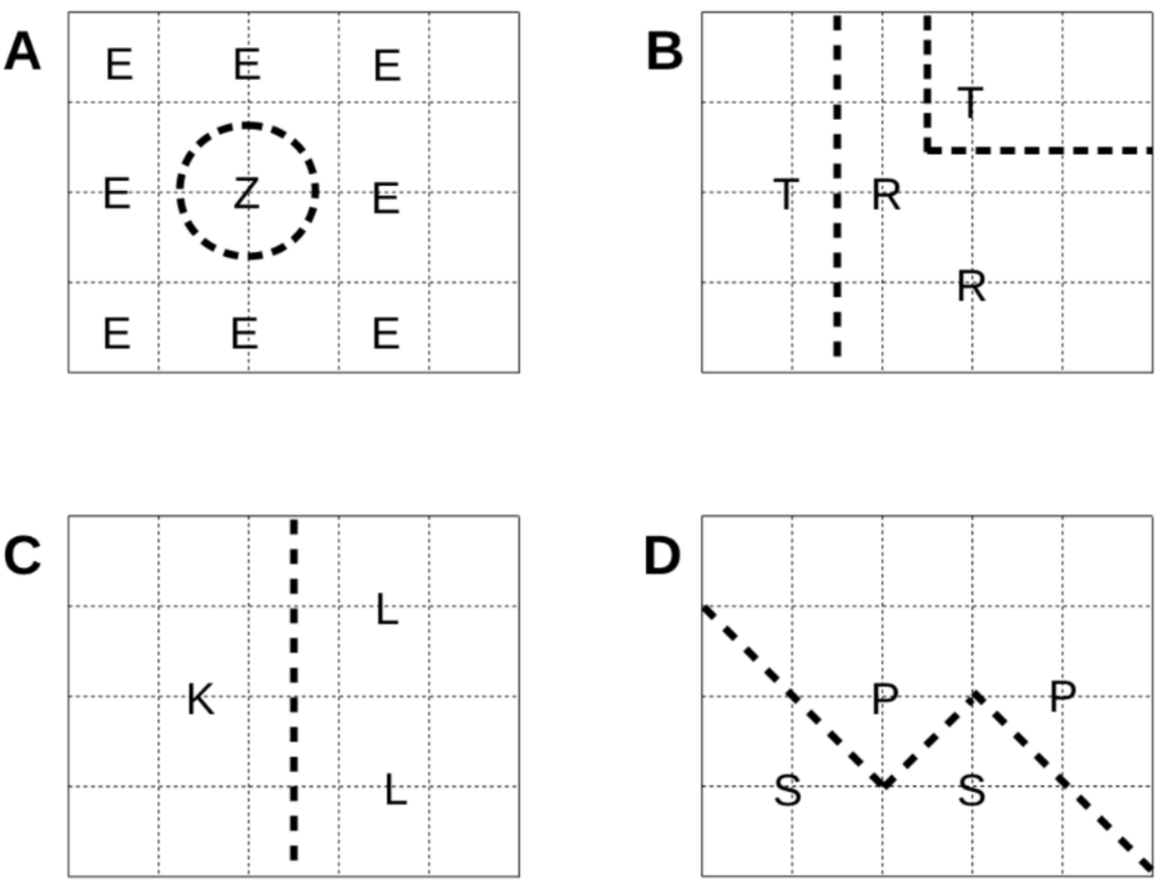
CARROT　　BEET　　CAN'T TELL

Part B: Back to our Roots (6 points)

Patrick saw how much fun you had drawing the decision boundary for carrots and beets—so he decides to do the same for other pairs of plants in the garden!

Which of the following graphs demonstrate accurate 1-Nearest Neighbor decision boundaries with Euclidean distance? Below, circle the letters corresponding to **all** of the graphs that have correct decision boundaries. If none of the boundaries are correct, instead circle **NONE**.

A **B** **C** **D** **NONE**



A is incorrect because
1. the boundaries should be an octagon shape

or

2. the 4 corner E's are too far out for the boundary to be a circle

Problem 3: Candyland Celebration (40 points)

★ **For credit on both parts A and B of this problem, show your work by both:**

(1) filling out the domain worksheet, and
(2) drawing the search tree.

Domain worksheet instructions (note: these are the same as in recitation):

1. When assigning values to a variable, do so in the order they appear in the domain.
2. Every time you **assign a value to a variable** or **propagate constraints through a variable**, fill out a new row in the table. (There may be more rows than you need.)
3. In the first column of that row, write **the name of the variable and the value you assigned** (e.g. X=2), or the **name of the variable through which you propagated constraints** (e.g. Y).
4. In the second column, list the **values that were just eliminated from neighboring variables' domains** as a result of the assignment or propagation (or "NONE" if no values were eliminated). Do not eliminate values from variables that have already been assigned.
5. If your search has to backtrack after assigning or propagating through a variable: First, **finish listing** all values eliminated from neighboring variables in the current row. Next, check the "backtrack" box in that row. Then, continue with the next assignment in the following row as usual.
6. If you have more than one choice of variable for propagation, break ties by propagating in **alphabetical order** (e.g. G before N).

Example row showing an assigned variable (with backtracking)

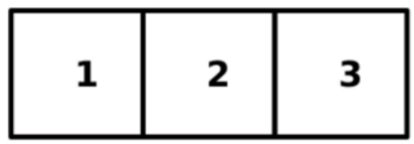
#	X = 3	Y ≠ 3, 4 Z ≠ 3	☑

Example row showing a propagated variable

#	X	W ≠ 1, 4	☐

Part A: Celebration at Candy Castle (18 points)

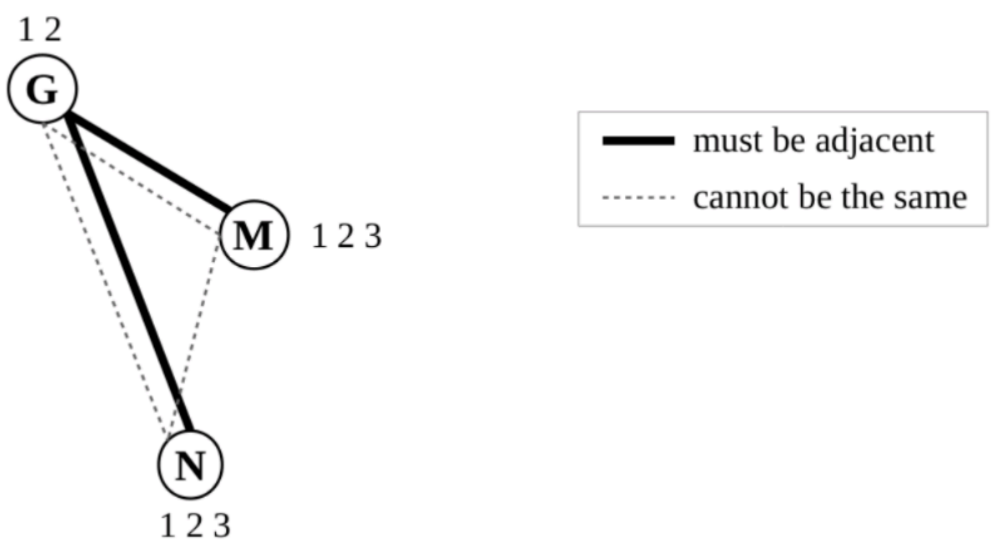
King Kandy has invited the citizens of Candyland to his home for a sweet candy celebration. The following guests have RSVPd: **N**anna Nutt, **M**r. Mint, and **G**loppy (the friendly chocolate monster). King Kandy needs to give each guest a room to stay in for the duration of the celebration, and he wants your help assigning guests to rooms! The layout and numbering of the rooms is show below:



King Kandy is working with the following constraints:

- Every guest must have their own room.
- **G**loppy and **M**r. Mint get along well, so they should be in adjacent rooms.
- **G**loppy and **N**anna Nutt get along well, so they should be in adjacent rooms.
- **G**loppy tends to melt in high temperatures. Rooms 1 and 2 have air-conditioning, so **G**loppy must stay in either Room 1 or Room 2.

To help you assign guests to rooms, we've drawn the constraint graph below. The domains have been initialized for you. **For your convenience, a copy of this graph is provided on a tear-off sheet at the end of the quiz.**



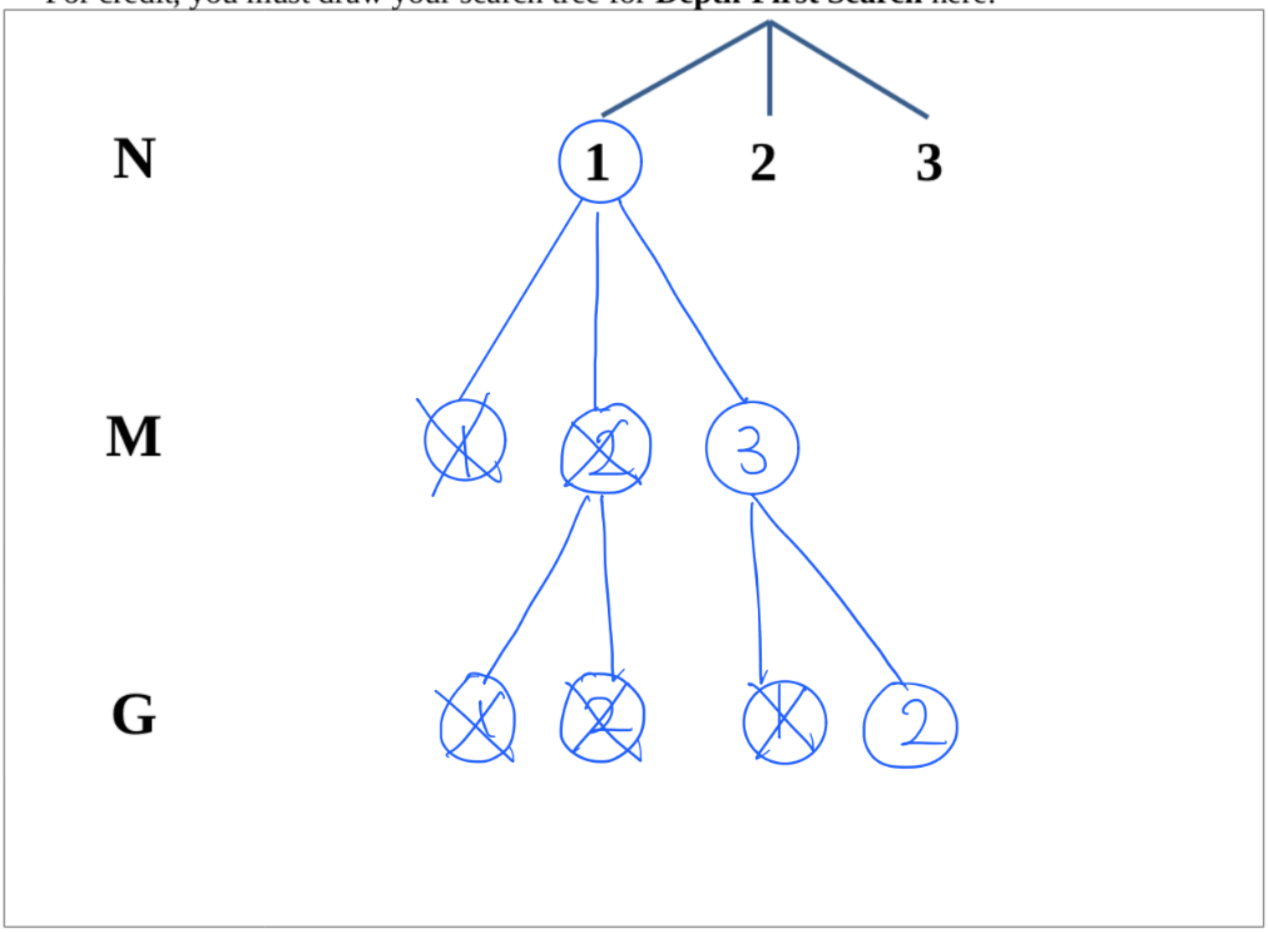
A1 (12 points) Help King Kandy make room assignments by performing a **<u>Depth-First Search (with NO Forward Checking, and NO Propagation)</u>** on the next page in order to find a set of room assignments.

Assign guests to rooms in the following order: **N, M, G**.

	Var assigned or propagated	Values just eliminated from neighboring variables' domains	Backtrack?
1	N=1	None	☐
2	M=1	None	☑
3	M=2	None	☐
4	G=1	None	☑
5	G=2	None	☑
6	M=3	None	☐
7	G=1	None	☑
8	G=2	None	☐
9			☐
10			☐

pure DFS will never use this column because it doesn't eliminate values from any domains

For credit, you must draw your search tree for **Depth-First Search** here:



A2 (2 points) What assignments did you make, based on your search from **Part A1**? Below, write the letter representing each guest (N, M, and G) in the room that you assigned them to; or, if your search yielded no solution, circle **NO SOLUTION**.

N **1**	G **2**	M **3**

NO SOLUTION

A3 (1 points) How many times did you backtrack? 4

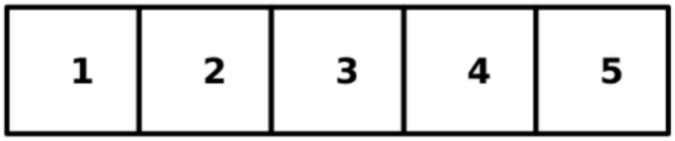
A4 (3 points) In this problem, we assigned guests to rooms in a certain order (N, M, G). Is there an order of assignment that would **decrease** the number of backtracks necessary during DFS? If so, circle **YES** and write such an ordering. If there is no such ordering, circle **NO**.

YES and such an ordering is: NGM or MGN

NO

Part B: Even More Guests! (22 points)

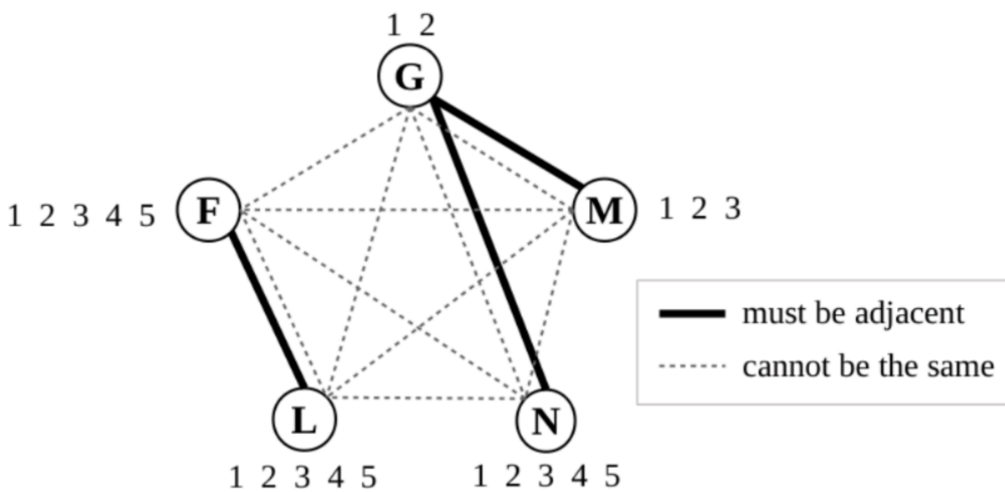
Two more guests RSVP to King Kandy's celebration: **F**rostine and **L**olly. King Kandy opens up two more rooms in his castle in order to accommodate all of the guests:



Because there are new guests, there are now more constraints for King Kandy to consider:

- Every guest must have their own room.
- **G**loppy and **M**r. Mint get along well, so they should be in adjacent rooms.
- **G**loppy and **N**anna Nutt get along well, so they should be in adjacent rooms.
- **G**loppy tends to melt in high temperatures. Rooms 1 and 2 have air-conditioning, so **G**loppy must stay in either Room 1 or Room 2.
- **L**olly and **F**rostine are cousins, so they should be in adjacent rooms.
- **M**r. Mint wants to stay in one of the original rooms, so he must be in either Room 1, 2, or 3.

As before, we have drawn the constraint graph below and initialized the domains for you. **For your convenience, a copy of this graph is provided on a tear-off sheet at the end of the quiz.**

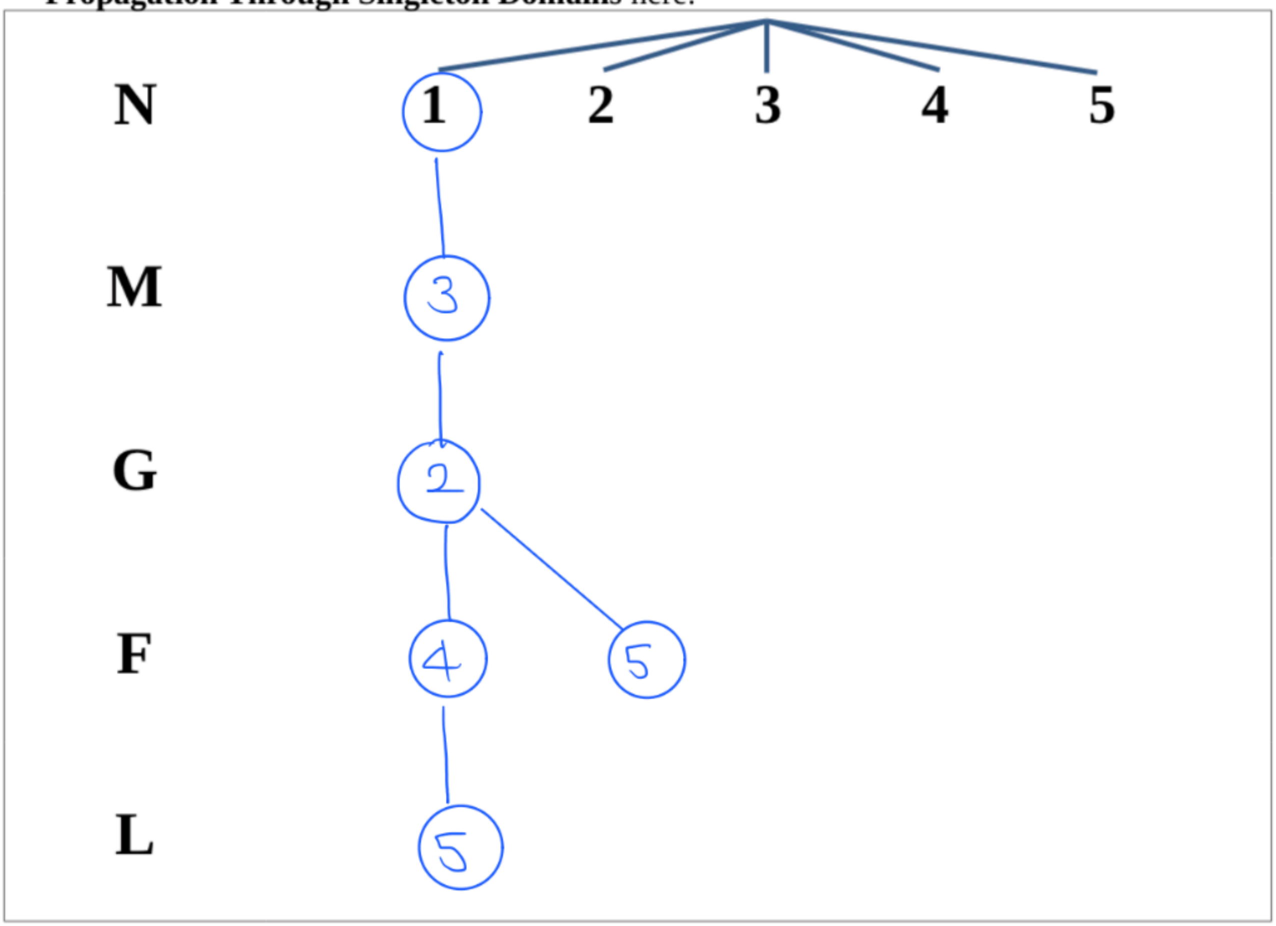


B1 (20 points) King Kandy wants to find room assignments for all of his guests, so he disregards the results from Part A. On the next page, perform **<u>Depth-First Search with Forward Checking and Propagation through Singleton Domains</u>** to find the new set of room assignments.

Assign guests to rooms in the following order: **N, M, G, F, L**.

	Var assigned or propagated	Values just eliminated from neighboring variables' domains	Backtrack?
1	N=1	G ≠ 1, F ≠ 1, L ≠ 1, M ≠ 1	☐
2	G	M ≠ 2, F ≠ 2, L ≠ 2	☐
3	M	F ≠ 3, L ≠ 3	☐
4	M=3	None	☐
5	G=2	None	☐
6	F=4	L ≠ 4	☐
7	L	None	☐
8	L=5	None	☐
9			☐
10			☐

For credit, you must draw your search tree for **DFS with Forward Checking and Propagation Through Singleton Domains** here:



B2 (2 points) What assignments did you make, based on your search from **Part B1**? Below, write the letter representing each guest (N, M, G, F, and L) in the room that you assigned them to; or, if your search yielded no solution, circle **NO SOLUTION**.

1	2	3	4	5
N	G	M	F	L

NO SOLUTION

Spiritual and Right Now (6 points)

For each of the following questions, circle the **one** best answer. There is **no penalty for wrong answers**, so it pays to guess in the absence of knowledge.

1. Kanwisher noted that the brain area specialized to *word forms*
   a. Is absent in non-human primates
   b. Responds vigorously to stick figures
   c **c.** Is at a location predicted by neural connection patterns
   d. Exhibits elevated activity when drawing pictures
   e. Has a diminished response in children who are dyslexic
2. To build a model of conflict-driven causal learning, Sussman described Stallman's and his approach for
   a. Goal-directed planning
   e **b.** Value-dependent truth maintenance
   c. Tree-based induction
   d. Neural-inspired propagation
   e. Dependency-directed backtracking
3. Sussman exemplified the idea of constraint propagators by describing
   a. How one can calculate the length of the Harvard bridge
   b. How one can measure the height of a building
   b **c.** How one can estimate the length of Kyla the Therapy Dog's tail
   d. How one can bound the distance between cities
   e. How one can derive mathematical equations
4. Boyden shared that a core motivation of his research is to
   a. Build better tools to study the brain
   b. Create an artificially intelligent being
   c. Manufacture the first brain-computer interface
   a **d.** Develop diagnostic methods for brain disorders
   e. Study the unique aspects of human intelligence using the Etruscan shrew
5. Describing his lab's expansion microscopy method, Boyden highlighted that his researchers could more clearly image the following structure:
   a. Blood-brain barrier membranes
   b. Glial cell insulation sites
   d **c.** Hippocampal place cell boundaries
   d. Presynaptic and postsynaptic connection sites
   e. Dorsal frontal cortex pathways
6. Pooling in deep neural nets refers to
   a. Arranging training sets so that similar samples are seen in groups
   b. The tendancy of weights to bounce around like billiard balls on a pool table
   c **c.** Taking the max value in a region to represent the region
   d. Research groups sharing ideas in specialized meetings
   e. Use of cloud computing to enable the training of extremely deep nets