6.034 Quiz 2

16 October 2019

Name	
Email	

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.

Sydney Gibson	**Rui Li**	**Allison Tam**
Udgam Goyal	**Jennifer Madiedo**	**Héctor Vazquez**
Jenna Hong	**Jack Murphy**	**Eric Wong**
Damon Jones	**Mira Partha**	

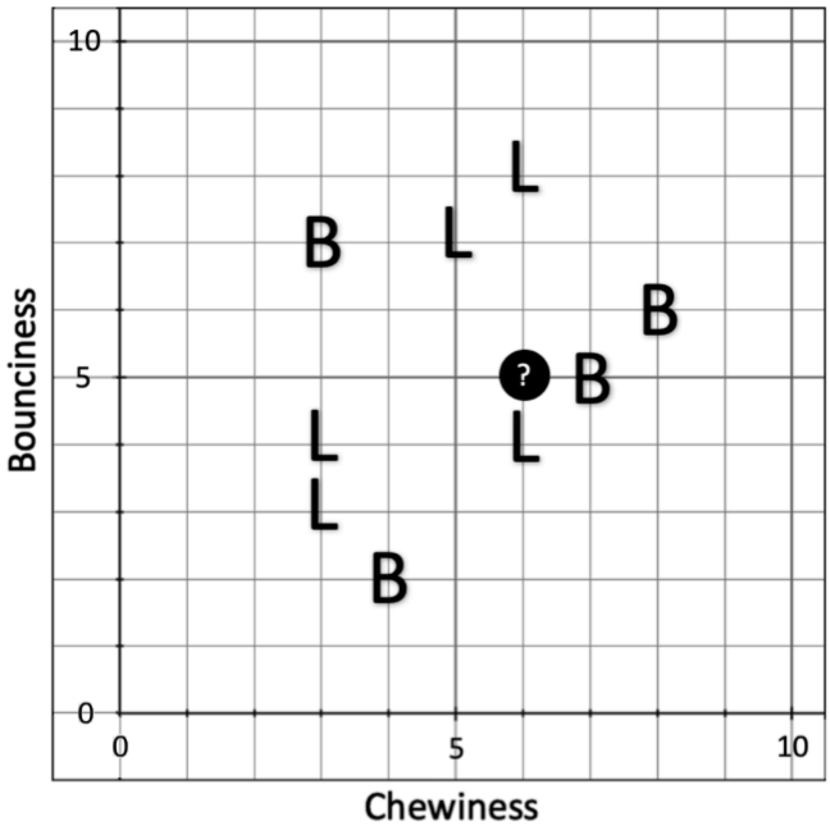
Problem number	Maximum	Score	Grader
1 - kNN	31		
2 - ID Trees	31		
3 - Constraints	32		
4 - SRN	6		
Total	100		

There are 18 pages in this quiz, including this one, but not including tear-off sheets. Tear-off sheets with duplicate drawings and data are located after the final page of the quiz. **We do not collect tear-off sheets, so please show your work on the quiz pages, not the tear-off sheets.**

As always, the quiz is open book, open notes, open just about everything, including a calculator, but no computers or cell phones.

Problem 1: k-Nearest Neighbors (31 Points)

Kyla, Dr. Koile's dog, is preparing for her weekly trip to the dog park and needs to choose which toys to bring with her. To help make this decision, she ranks each toy based on its chewiness and bounciness. In the graph below, each point labeled **B** or **L** represents a toy Kyla has already decided that she will **bring (B)** or **leave (L).** The mystery circle, labeled with a "?"**,** indicates the next toy Kyla is considering. (Assume the labels are on grid intersections.)

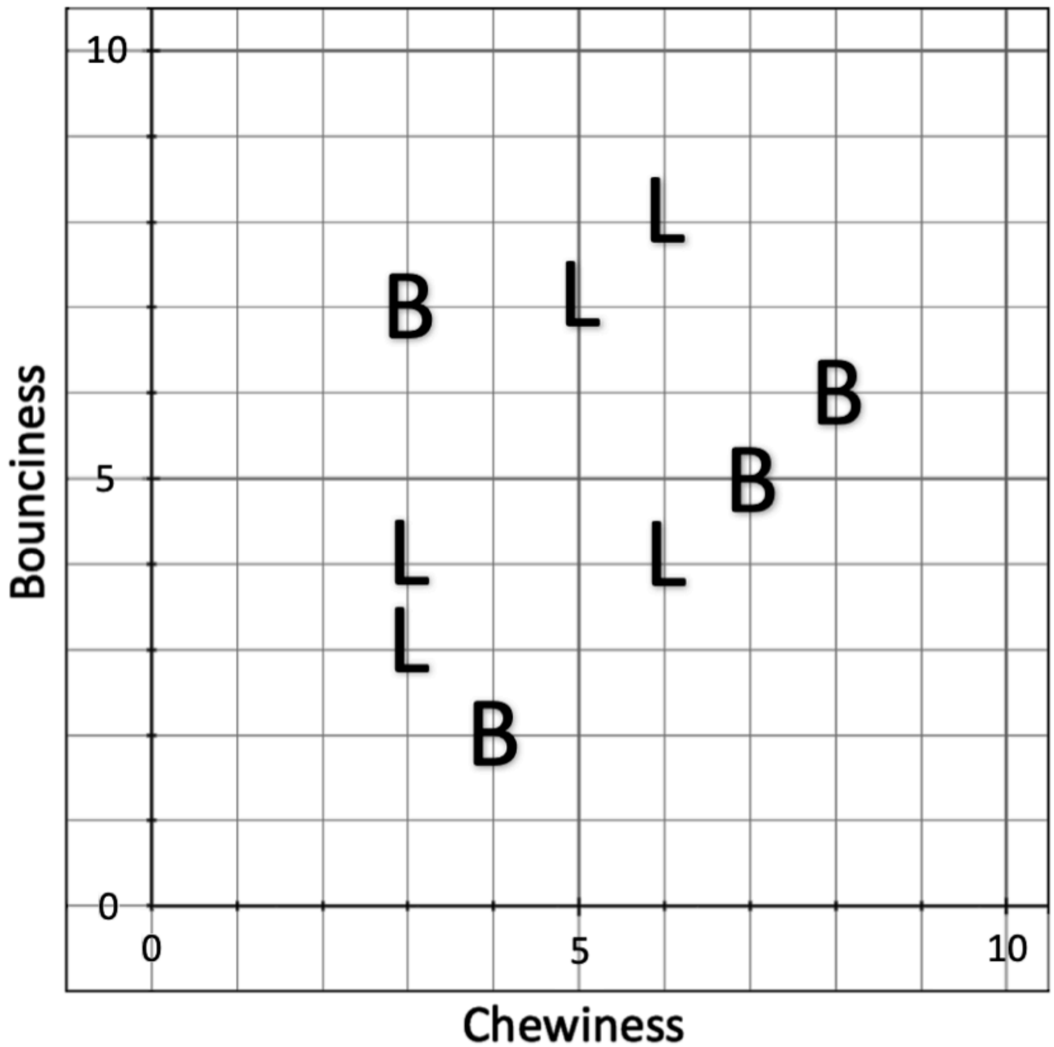


A1 (8 points) For each of the following k-values, use **k-Nearest Neighbors with Euclidean distance** to determine whether Kyla should **(B)**ring or **(L)**eave the mystery toy. Circle one for each k.

k=1	Bring	Leave	Can't Tell
k=3	Bring	Leave	Can't Tell
k=5	Bring	Leave	Can't Tell
k=7	Bring	Leave	Can't Tell

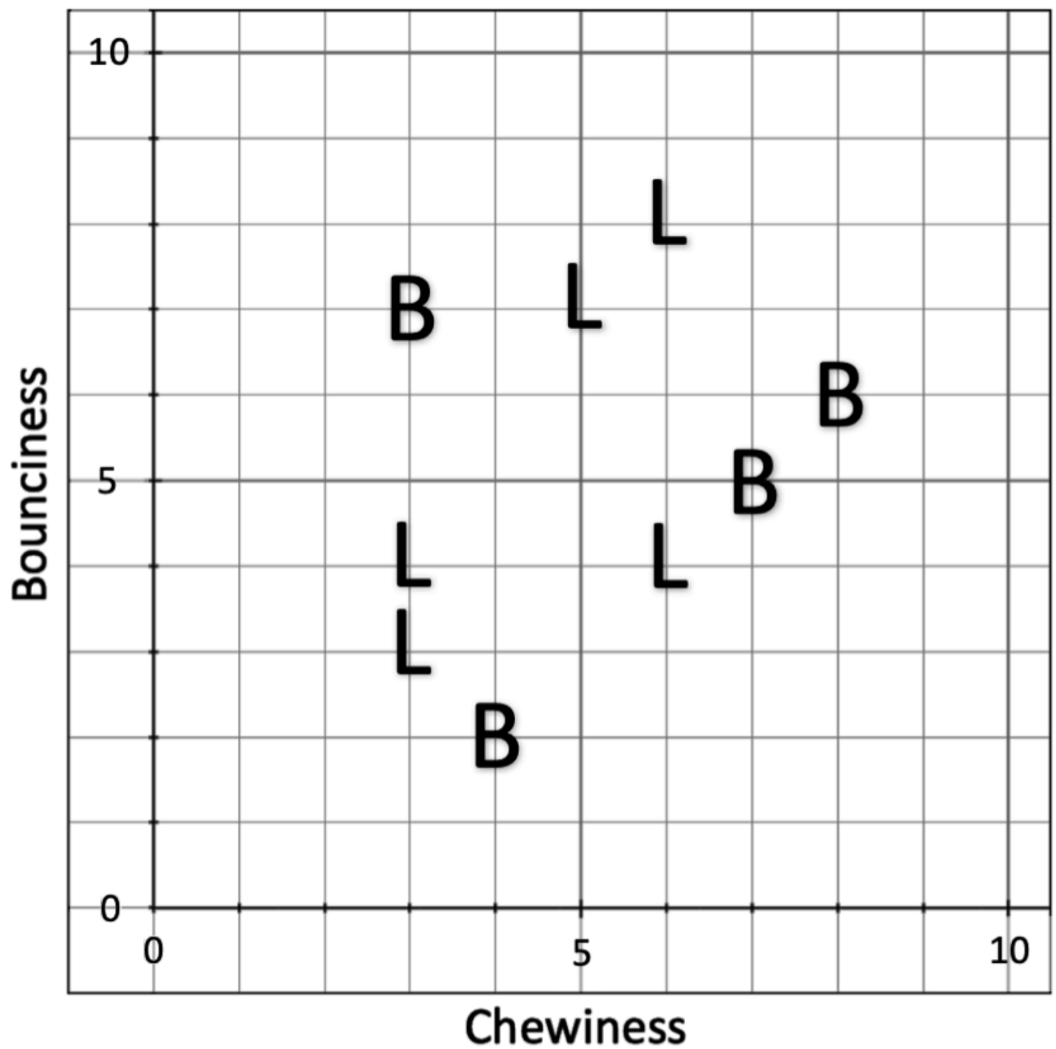
A2 (10 points) Draw the **1-Nearest Neighbors** decision boundary using the **Euclidian distance metric** on the graph below. Put arrows at the end of all decision boundaries that continue beyond the graph. Remember you can use test points to check your boundaries. *(If you want to start over, use the next page. If you write on both pages, clearly mark which page we are to grade.)*

Hint: Not all boundary lines must connect.



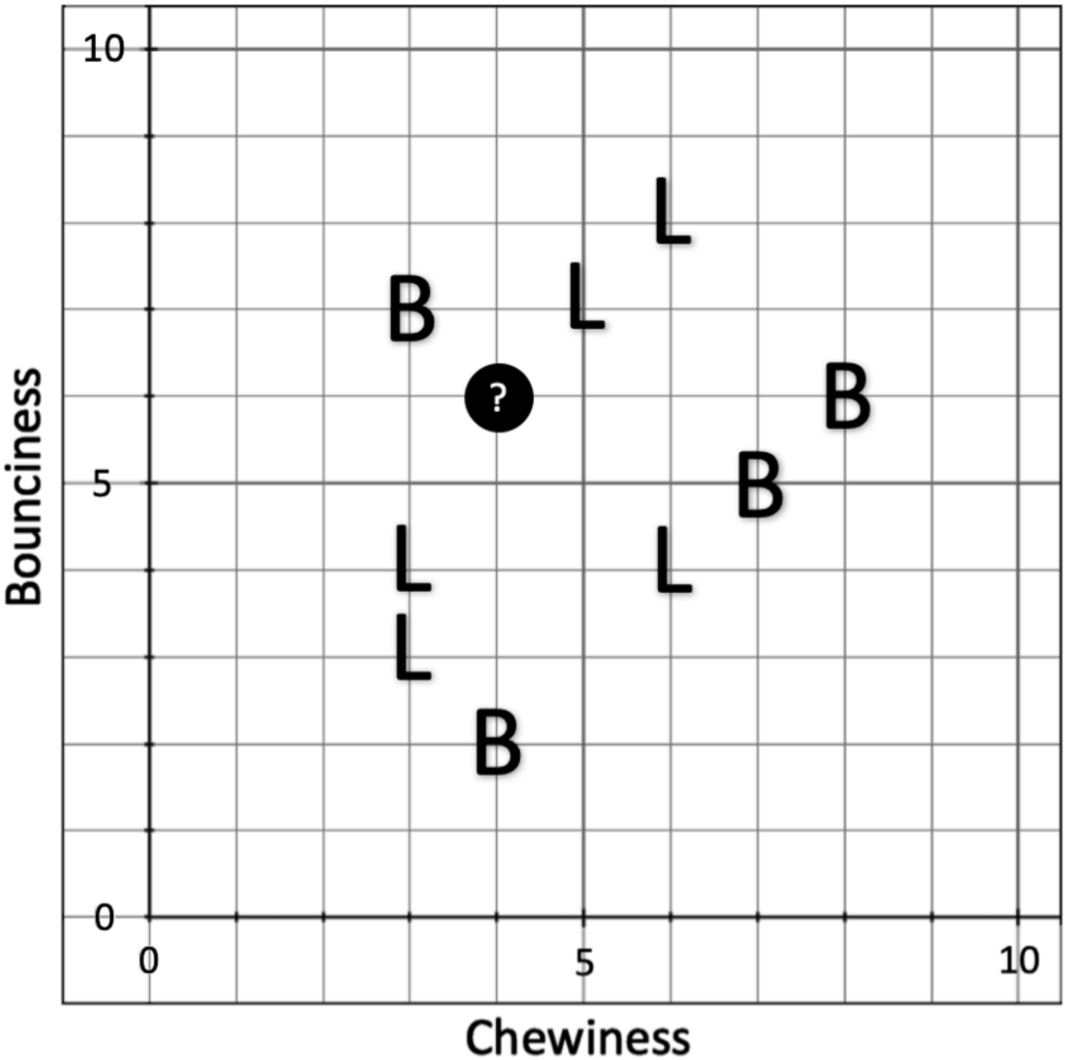
This is a duplicate copy of the graph for A3. If you want this copy graded instead, check the box.

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



A3 (6 points) Kyla is now considering a different toy, indicated by the new mystery circle, labeled with a "?"**,** in the graph below. Using **3-Nearest Neighbors**, which of the following distance metrics would have Kyla **Leave** (**L)** the new mystery toy? Circle **all that apply** or **None of These**.

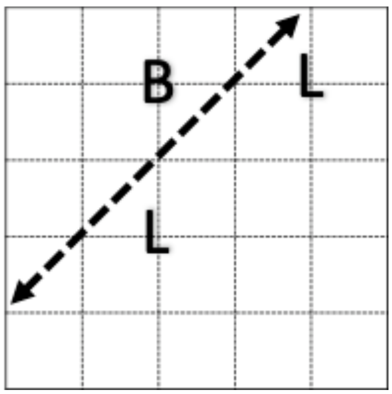
Manhattan **Hamming** **Cosine** **None of These**



A4 (8 points) Kyla is working hard to decide on each toy, but needs your help! She has drawn the following 1-Nearest Neighbor boundaries.

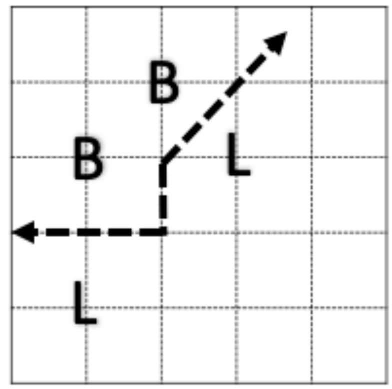
For each graph, circle whether the boundary is CORRECT or INCORRECT.

Graph A



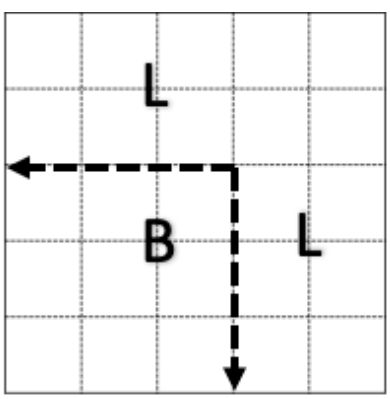
CORRECT **INCORRECT**

Graph B



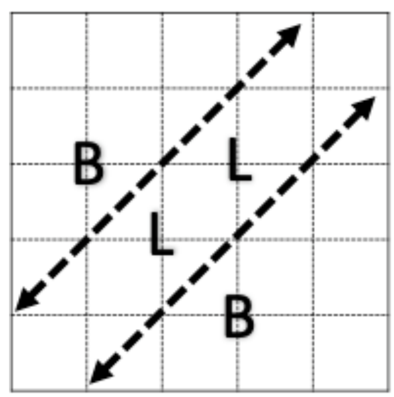
CORRECT **INCORRECT**

Graph C



CORRECT **INCORRECT**

Graph D



CORRECT **INCORRECT**

Problem 2: ID Trees (31 points)

Part A: Messy Merger (27 points)

As a parting gift before her retirement, Anne Hunter negotiated the merger of another MIT department with MIT EECS. Unfortunately, the merger was a mess, and now the department doesn't know which students are EECS students. You and your friends propose three attributes that can be used to identify the EECS students: their **operating system (OS)**, their programming **editor**, and the **city** where they interned last summer.

The table below lists the information you and your friends compiled to train the ID tree to identify EECS students (EECS? = Yes) and Non-EECS students (EECS? = No).

EECS?	**OS**	**Editor**	**City**
Yes	Mac	Vim	SF
No	Windows	Sublime	NYC
Yes	Mac	Vim	NYC
No	Mac	Sublime	BOS
No	Mac	Notepad	SF
No	Windows	Vim	NYC
No	Windows	Notepad	BOS

A1 (2 points) Order the three feature tests (**OS**, **Editor**, **City**) from least to greatest disorder. (You don't need to show your calculations; you may use your intuition.)

If two tests have the same disorder, break the tie lexicographically by feature test (i.e., prefer **City** over **Editor** and **Editor** over **OS**).

Least disorder Most disorder

A2 (12 points) Calculate the average disorder for each of the three feature tests. Use the table of logarithms on the next page to express your answers as **sums and products of decimals and fractions only**. Your final answer should **not include logarithms**.

The original data table is repeated below and on a tear-off sheet for your convenience.

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

Feature Test	**Disorder (no logs)**
OS	
Editor	
City	

EECS?	**OS**	**Editor**	**City**
Yes	Mac	Vim	SF
No	Windows	Sublime	NYC
Yes	Mac	Vim	NYC
No	Mac	Sublime	BOS
No	Mac	Notepad	SF
No	Windows	Vim	NYC
No	Windows	Notepad	BOS

For partial credit, show your disorder calculation work here.

OS

Editor

City

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

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

A copy of this table of logs is available on a tear-off sheet.

A3 (11 points) Construct a **greedy, disorder-minimizing identification tree** to correctly classify all the students as **EECS? =Yes** or **EECS? =No** using the three feature tests **OS**, **Editor**, **City**. *(If you want to start over, use the next page. If you write on both pages, clearly mark which page we are to grade.)*

If two tests have the same disorder, break the tie lexicographically by feature test (i.e., prefer **City** over **Editor** and **Editor** over **OS**).

Draw your ID tree here.

A4 (1 point) The 6.034 staff members heard about your ID tree and are curious to see how the tree would classify the 6.034 TAs. How would your ID tree classify Eric given the data below?

Eric's Data:

EECS	OS	Editor	City
?	Windows	Notepad	NYC

Circle the classification below

EECS? =Yes **EECS? =No** **Can't Tell**

This is a duplicate for part A3 (Construct ID tree). If you want this copy graded instead, check the box.

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

Draw your ID tree here.

This is a duplicate for part A4. If you want this answer graded instead, check the box.

☐ **I want this answer graded.**

A4 (1 point) The 6.034 staff members heard about your ID tree and are curious to see how the tree would classify the 6.034 TAs. How would your ID tree classify Eric given the data below?

	EECS	OS	Editor	City
Eric's Data:	?	Windows	Notepad	NYC

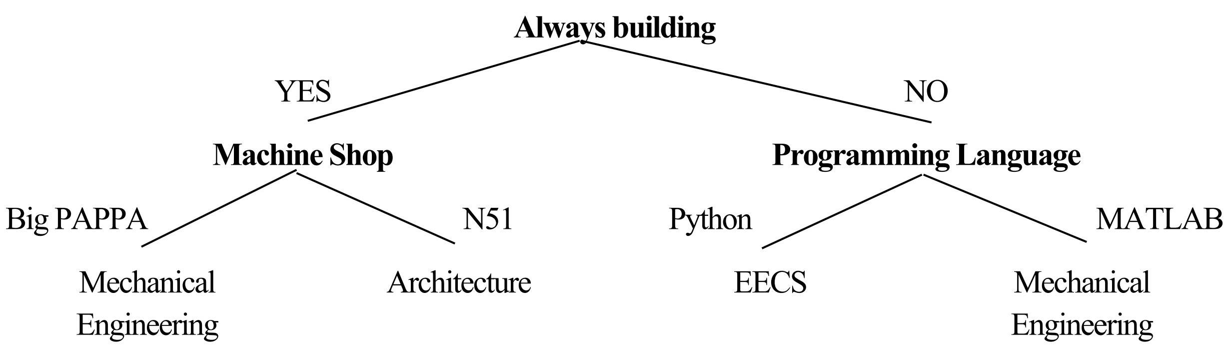
Circle the classification below

EECS? =Yes **EECS? =No** **Can't Tell**

Part B: To MechE or Not to MechE (4 points)

B1 (4 points) The Mechanical Engineering Department is inspired by the work you did for the EECS Department, so they decide to build their own ID tree to identify students. The tree they developed is shown below.

Write two **if-then rules** that will identify a Mechanical Engineering student based on the ID tree. (Feel free to abbreviate words.)



Rule 1:

IF (

____

____

____

____

)

THEN Mechanical Engineering

Rule 2:

IF (

____

____

____

____

)

THEN Mechanical Engineering

Problem 3: Constraint Satisfaction (32 points)

Part A: The Lost Schedule (18 points)

The New England Women's and Men's Athletic Conference (NEWMAC) has lost the schedule it created for NCAA football games! Luckily, the quarterback of the MIT Engineers is none other than Udgam Goyal, a TA for 6.034. "No problem," he says. "I can have my students draft up a new schedule in no time."

There are 6 teams that need to play a game next week. They are:
Babson, **C**lark, **E**merson, **M**IT, **S**mith, and **W**heaton

Schedule constraints:

- Each team must play in **only one** game, either Game 1, 2, or 3
- **Exactly two** teams must play in each Game.
- **B**abson must play in Game 1.
- Last week's matchups, shown in the schedule below, cannot be repeated: i.e., **B** vs **C**, **E** vs **M**, **S** vs **W** are not allowed in the new schedule.

Last week's schedule:

Game 1	**B**abson	**C**lark
Game 2	**E**merson	**M**IT
Game 3	**S**mith	**W**heaton

A1 (2 points) Fill in the table below with the initial domain values, i.e., game numbers, for each of the variables.

	Domain
B	
C	
E	
M	
S	
W	

A2 (14 points) Perform **<u>Depth-First Search with Forward Checking and no propagation</u>** (DFS+FC) to generate a schedule for the next week. **Draw your search tree in box below, using the order of variables provided.** *(If you want to start over, use the next page. If you write on both pages, clearly mark which page we are to grade.)*

B

C

E

M

S

W

A3 (1 point) Which team will play against **M** (MIT) next week?

B **C** **E** **S** **W**

A4 (1 point) Which teams were reassigned during backtracking? Circle **all that apply** or **None** if no backtracking occurred.

B **C** **E** **S** **W** **M** **None**

This is a duplicate for part A2 (DFS + FC). If you want this copy graded instead, check the box.

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

B

C

E

M

S

W

This is a duplicate for parts A3 and A4. If you want these answers graded instead, check the box.

☐ **I want these answers graded.**

A3 (1 point) Which team will play against **(M)**IT next week?

B **C** **E** **S** **W**

A4 (1 point) Which teams were reassigned during backtracking? Circle **all that apply** or **None** if no backtracking occurred.

B **C** **E** **S** **W** **M** **None**

Part B: Rivalry Week and Domain Reduction (14 points)

Clark and **E**merson review your proposed schedule and insist that they need to play each other in the new schedule. Udgam Goyal also noticed that Game 1 is scheduled during 6.034 lecture, and Game 3 is during recitation. Therefore, **M**IT needs to play in Game 2. (These two additional new constraints are noted below.)

Schedule constraints:

- Each team must play in **only one** game, either Game 1, 2, or 3.
- **Exactly two** teams must play in each Game.
- Last week's matchups cannot be repeated, i.e., **B** vs **C**, **E** vs **M**, **S** vs **W** are not allowed.
- **B**abson must play in Game 1.
- (New) **C**lark and **E**merson must play in the same Game.
- (New) **M**IT must play in Game 2.

	Domain
B	1
C	1 2 3
E	1 2 3
M	2
S	1 2 3
W	1 2 3

B1 (13 points) Perform **Domain Reduction through singleton domains** (i.e., Prop-1 without search) to reduce the potential schedule choices as much as possible by filling out the table below. (There may be more rows than you need.) Initialize your propagation queue in alphabetical order.

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

Note: You don't need to do assignments; just reduce the domains.

	Variable de-queued	List all values just eliminated from neighboring variables or NONE
1	**B**	
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		

For partial credit, show your work here.

B2 (1 point) Which teams can **M (**MIT) play against in Game 2? Circle **all that apply**.

B **C** **E** **S** **W**

Problem 4: Spiritual and Right Now (6 points)

For each question, **write in the box** provided the letter corresponding to the **one** best answer **and circle** the answer. There is **no penalty for wrong answers,** so it pays to guess in the absence of knowledge.

1. According to Raibert, the two main kinds of intelligence are:
   A. Scholarly Intelligence and Logical Intelligence
   B. Emotional Intelligence and Commonsense Intelligence
   C. Athletic Intelligence and Scholarly Intelligence
   D. Spatial Intelligence and Athletic Intelligence
   E. Intrapersonal Intelligence and Interpersonal Intelligence

2. In his Right-now talk, Boyden did **not** discuss:
   A. A robot that does evolution
   B. A method for harvesting electricity from jellyfish
   C. A biological solar panel
   D. Mice induced to sleep by light
   E. A cancer-diagnosing machine learning model

3. The three techniques Boyden described for understanding the brain are:
   A. Submerge, Stain, Analyze
   B. Map, Validate, Test
   C. Identify, Diagnose, Treat
   D. Map, Control, Observe
   E. Label, Train, Test

4. When discussing deep neural nets, Szolovits discussed *dropout* as a technique that:
   A. Discards a percentage of the test data
   B. Discards a percentage of the training data
   C. Optimizes the hyperparameters of the model
   D. Prevents overfitting of the training data
   E. Decreases the time spent testing the model

5. During his Right-now talk, Madry claimed that:
   A. Adversarial perturbations are meaningless aberrations.
   B. Machine learning models can't identify images using a human perspective.
   C. Machine learning models should not be used at all going forward.
   D. Robust machine learning models are not useful.
   E. Adversarial perturbations are features.

6. In his talk, Madry did **not** discuss as an example of machine learning brittleness:
   A. Person vs. robot
   B. Dog vs. cat
   C. Pig vs. airliner
   D. Turtle vs. rifle
   E. Tap vs. toc

This page is intentionally blank (except for this sentence).

Tear-off sheet
We do not collect tear-off sheets, so please show your work on the quiz pages, not the tear-off sheet.

Problem 2 (ID Trees)

EECS?	OS	Editor	City
Yes	Mac	Vim	SF
No	Windows	Sublime	NYC
Yes	Mac	Vim	NYC
No	Mac	Sublime	BOS
No	Mac	Notepad	SF
No	Windows	Vim	NYC
No	Windows	Notepad	BOS

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

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

Tear-off sheet
We do not collect tear-off sheets, so please show your work on the quiz pages, not the tear-off sheet.

Problem 3 (Constraint Satisfaction)

There are 6 teams that need to play a game next week. They are:
Babson, **C**lark, **E**merson, **M**IT, **S**mith, and **W**heaton

Part A

Schedule constraints:
- Each team must play in **only one** game, either Game 1, 2, or 3.
- **Exactly two** teams must play in each Game.
- **B**abson must play in Game 1.
- Last week's matchups, shown in the schedule below, cannot be repeated:
  i.e., **B** vs **C**, **E** vs **M**, **S** vs **W** are not allowed in the new schedule.

Last week's schedule:

Game 1	**B**abson	**C**lark
Game 2	**E**merson	**M**IT
Game 3	**S**mith	**W**heaton

Tear-off sheet
We do not collect tear-off sheets, so please show your work on the quiz pages, not the tear-off sheet.

Problem 3 (Constraint Satisfaction)

Part B

Schedule constraints:

- Each team must play in **only one** game, either Game 1, 2, or 3.
- **Exactly two** teams must play in each Game.
- Last week's matchups cannot be repeated, i.e., **B** vs **C**, **E** vs **M**, **S** vs **W** are not allowed.
- **B**abson must play in Game 1.

New:
- **C**lark and **E**merson must play in the same Game.
- **M**IT must play in Game 2

	Domain
B	1
C	1 2 3
E	1 2 3
M	2
S	1 2 3
W	1 2 3