SOLUTIONS

6.034 Final

December 15, 2020

You will have 4 hours total to complete and upload the quiz (3 hours to complete the quiz, 1 hour to upload) in a 24-hour submission window starting 8am EST, December 15, 2020 through 8am EST, December 16, 2020.

Note: If you open the quiz with less than 4 hours remaining in the submission window, you will only have the remaining time until 8am, December 16, 2020 to complete and upload the final.

This final is open book and open notes. You may NOT use the Internet **(including the recitation videos)** or other people. You can resubmit the final as many times as you want on Gradescope before the submission window closes.

Updates and clarifications will be posted to a pinned Piazza post for this final. **Send private Piazza posts** to the instructors to ask for clarification.

Quiz 1 includes pages 2-23. Quiz 2 includes pages 24-39. Quiz 3 includes pages 40-57. Quiz 4 includes pages 58-74. The tearoffs for each quiz are located at the beginning of each quiz.

Gradescope

Final submissions work exactly like Quiz 3 and 4, on Gradescope.

You will download a single PDF template of the final. Your quiz submission will be a single PDF or multiple camera images, submitted on Gradescope. **You must then designate which pages correspond to which questions on Gradescope.**

You will be given one extra-credit point for each quiz section if you correctly **orient** and designate your submission's pages to the corresponding questions.

1. Please orient all pages correctly.
2. Only designate the pages that have your written work i.e. the pages that need to be graded. **Please do not designate any blank pages and instruction pages to expedite grading.**

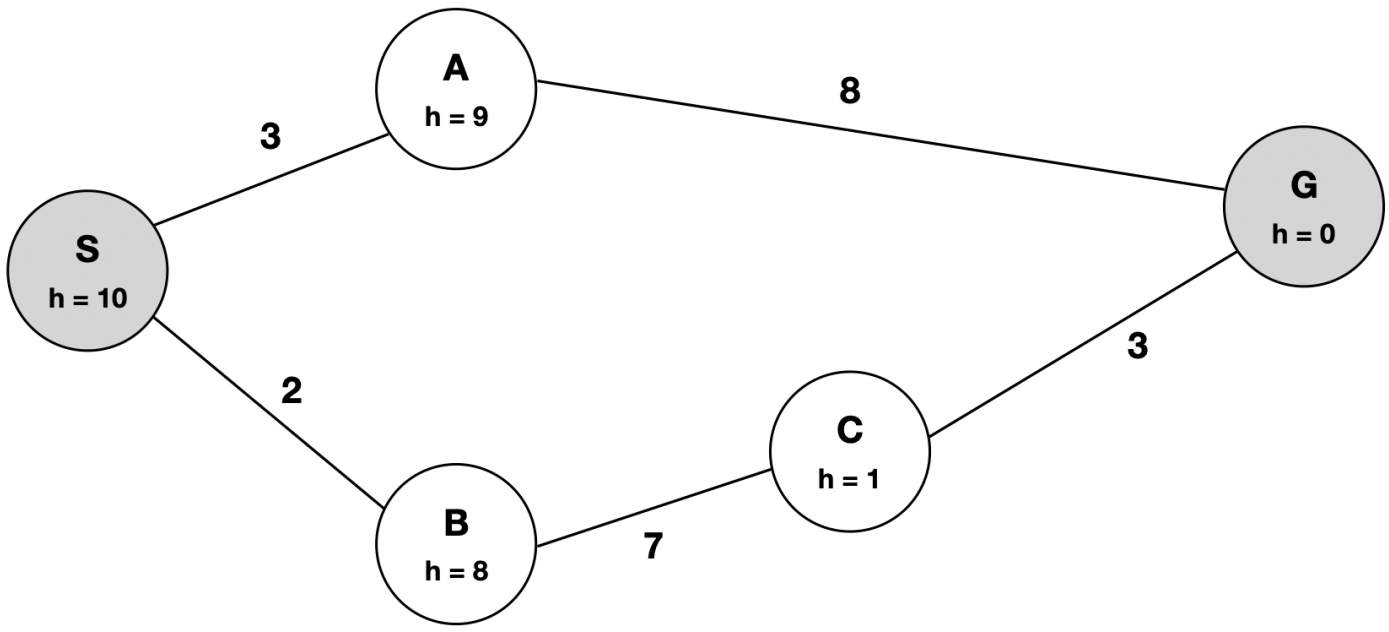
Quiz 1, Problem 1: Search (33 points)

Part A: A Dog's Journey (12 points)

After a long and tiring week of final exams at MIT, Kyla, who is sure she aced her 6.034 final, is ready to head home. Given how tired she is, she wants to get from the Stata Center (start node S) to her front Gate (goal node G) as soon as possible so she can finally curl up with a nice, warm blanket and sleep.

For your convenience, a copy of the graph is provided on a tear-off sheet.

On the map below, each location between Stata and her front Gate is a node labeled with a letter and a heuristic distance to the goal node (G). Each edge is labeled with its length, i.e., the distance between nodes. The start node (S) and goal node (G) are gray.

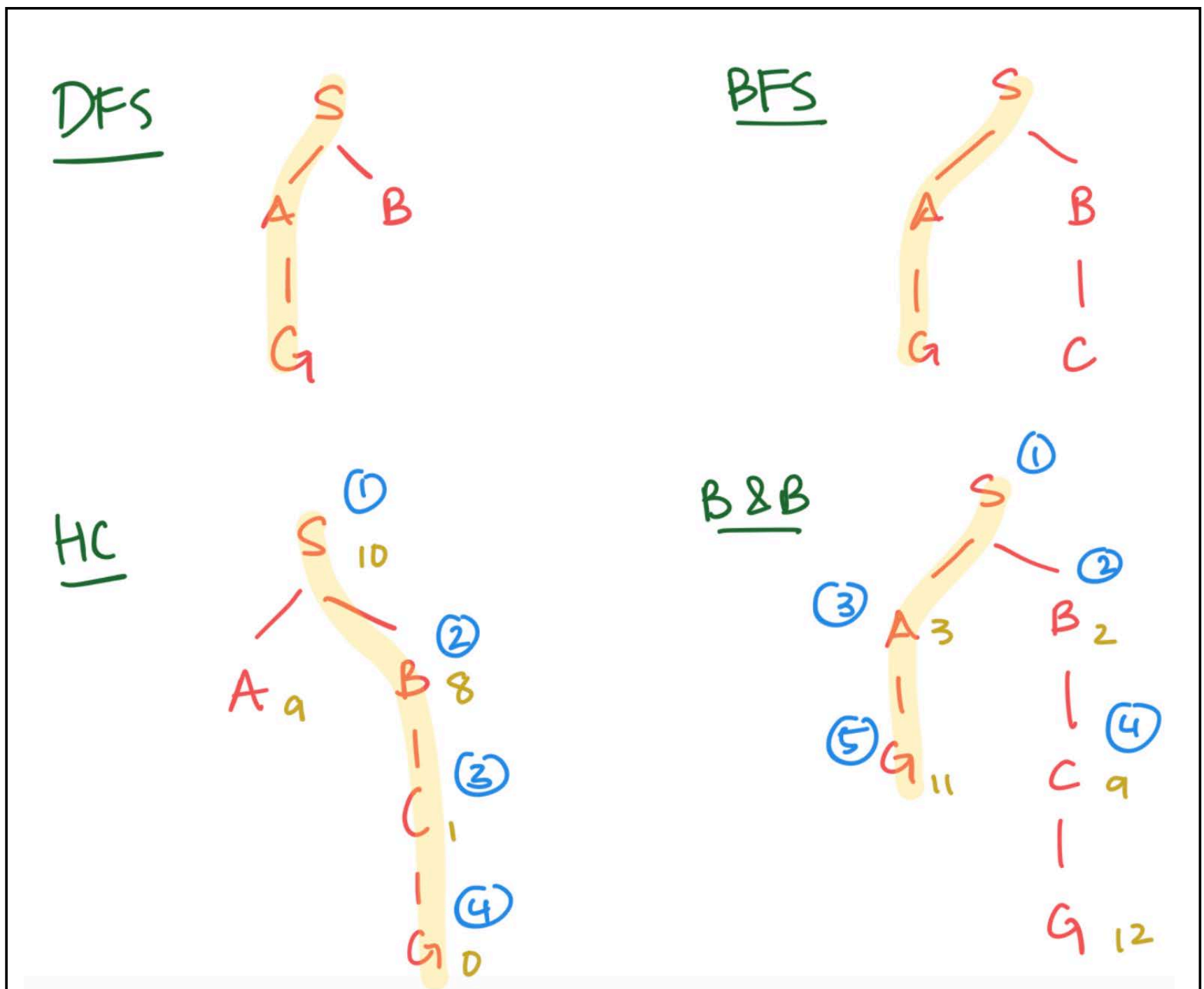


1.1 (8 points) Kyla has two different paths she can use to get to node G: **S-B-C-G** or **S-A-G**. Having just finished her 6.034 final exam, Kyla remembers four search algorithms that can be used to determine the path to choose. Circle which of the two paths is generated first by each of the following search algorithms. Break ties lexicographically.

You may show your work in the box provided on the next page for partial credit.

Branch and bound (no heuristic or extended set)	**S-B-C-G**	**S-A-G**
Depth-first search (with backtracking)	**S-B-C-G**	**S-A-G**
Breadth-first search	**S-B-C-G**	**S-A-G**
Hill climbing (with no backtracking)	**S-B-C-G**	**S-A-G**

For 1.1 partial credit, show your work in the box below. Clearly label which work belongs to which search algorithm.



1.2 (2 points)

Are the heuristics admissible? *Circle one* **Yes** **No**

Which node(s) are inadmissible? *Circle all that apply.* **S** **A** **B** **C** **G** **N/A**

For 1.2 partial credit, show your work in the box below.

Admissible: $H(N, G) \leq sp(N, G)$

A: $9 \not\leq 8$

S: $10 \leq 11$

B: $8 \leq 10$

C: $1 \leq 3$

1.3 (2 points)

Are the heuristics consistent? *Circle one* **Yes** **No**

Which edge(s) are inconsistent? *Circle all that apply.* **S-A** **S-B** **B-C** **C-G** **A-G** **N/A**

For 1.3 partial credit, show your work in the box below.

Consistent: $|H(N_1) - H(N_2)| \leq sp(N_1, N_2)$

S-A: $1 \leq 3$

S-B: $2 \leq 2$

B-C: $7 \leq 7$

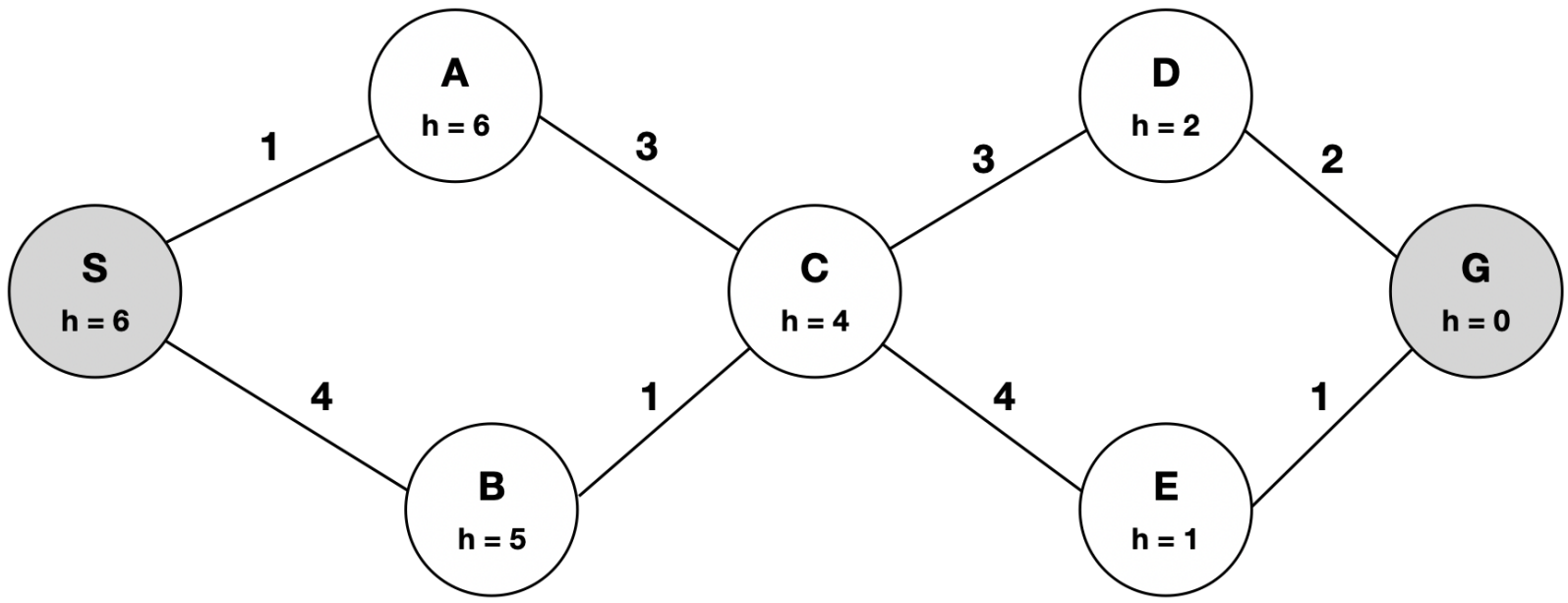
C-G: $1 \leq 3$

A-G: $9 \not\leq 8$

Part B: Tell the World I'm Coming Home (13 points)

Kyla suddenly remembers a search algorithm called **A* search** (with heuristic and extended set) that she can use to find her way from start node (S) to goal node (G). However, while she was thinking, she found even more locations between nodes (S) and (G). The new graph is drawn below. Each location is a node labeled with a letter and a heuristic distance to the goal (G). Each edge is labeled with its length, i.e., the distance between nodes. The start node (S) and goal node (G) are gray.

For your convenience, a copy of the graph is provided on a tear-off sheet.



2.1 (10 points) Draw the **A* search tree** (<u>with heuristic and extended set</u>) in the box below. Be sure to:

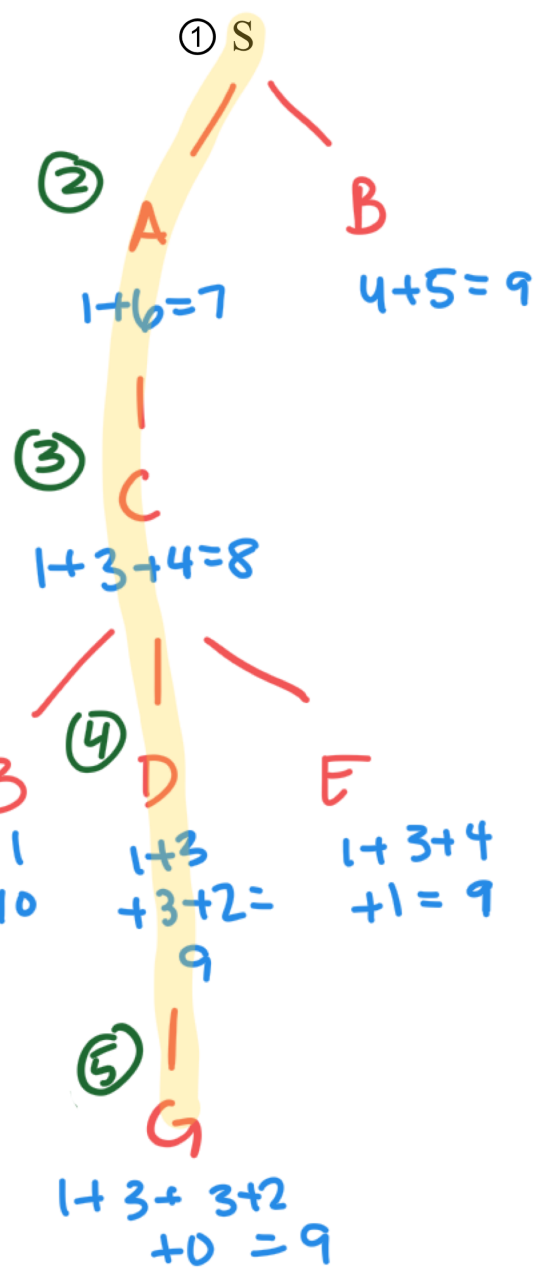
a. Draw the children of each node in lexicographical order (e.g., A < B < C).
b. Break any ties using alphabetical order of the entire path (e.g., S-A-E < S-B-E).
c. Write the queue at each step of the algorithm.
d. Clearly indicate the order in which you extended nodes by numbering the extended nodes in your search tree (①, ②, ③, …).

For credit, draw your search tree in the box below.

Queue:

1. [S]
2. [SA, SB]
3. [SAC, SB]
4. [SACD, SACE, SB, SACB]
5. [SACDG, SACE, SB, SACB]

Tree:



2.2 (1 point) List the **nodes** in the **extended set** in the order extended.

S, A, C, B, D, G

2.3 (1 point) What is the **final path** found? Write the nodes in order.

S-A-C-D-G

2.4 (1 point) For this graph, is it possible for Branch and Bound (with no heuristic and no extended set) to find a shorter path than A*? For full credit, please show your work and **explain why or why not**.

No, because the heuristics are all consistent, which mean the A* path will be the shortest/optimal

Part C: What's That Search? (8 points)

On the way home using the path found from **A* search**, Kyla decides to play a mental game to refresh her search memory. The following is a list of all of the search algorithms covered in class, followed by the corresponding abbreviation used in recitations throughout the semester:

Search Algorithm	**Abbreviation**	**Search Algorithm**	**Abbreviation**
A* search (**with Heuristic and Extended Set**)	A*	Beam Search	Beam
Branch and Bound	B&B	Hill Climbing	H.C.
Branch and Bound **with Extended Set** (*no heuristic*)	B&B + E.S.	Depth First Search	DFS
Branch and Bound **with Heuristic** (*no extended set*)	B&B + H	Breadth First Search	BFS
Best First Search	B.F.	British Museum	B.M.

For each of the following questions, list **all** the algorithms that apply. Please write the **abbreviation** of each algorithm that applies (ex: for Depth First Search, list **DFS**). If none apply, write **None**.

3.1 (1 point) Which of these algorithms sort **the agenda** after adding newly found paths?

B.F., B&B, B&B+E.S., B&B+H, A*

3.2 (1 point) Which of these algorithms are **guaranteed** to find the shortest path **based on edge length**, if it exists? (If an algorithm uses a heuristic, do not assume that the heuristic is admissible and/or consistent.)

B&B, B&B+E.S., B.M.
(B&B+H/A* requires admissible/consistent)

3.3 (1 point) Which of these algorithms are **not** guaranteed to find a path, if it exists?

Beam

3.4 (1 point) Which of these algorithms sort paths **before** adding them to the agenda?

Beam, H.C.

3.5 (1 point) Which of these algorithms can use a **consistent** heuristic to always find the shortest path to the goal **in terms of edge length**, if it exists?

B&B+H, A*
consistent implies admissible

3.6 (1 point) Which of these algorithms use **some** kind of cost function to determine which path to extend next?

H.C., Beam, B.F., B&B, B&B+E.S., B&B+H, A*

3.7 (1 point) Which of these algorithms have a built-in size limit on the agenda?

Beam

3.8 (1 point) Which of these algorithms add children to the **back** of the agenda?

BFS, Beam

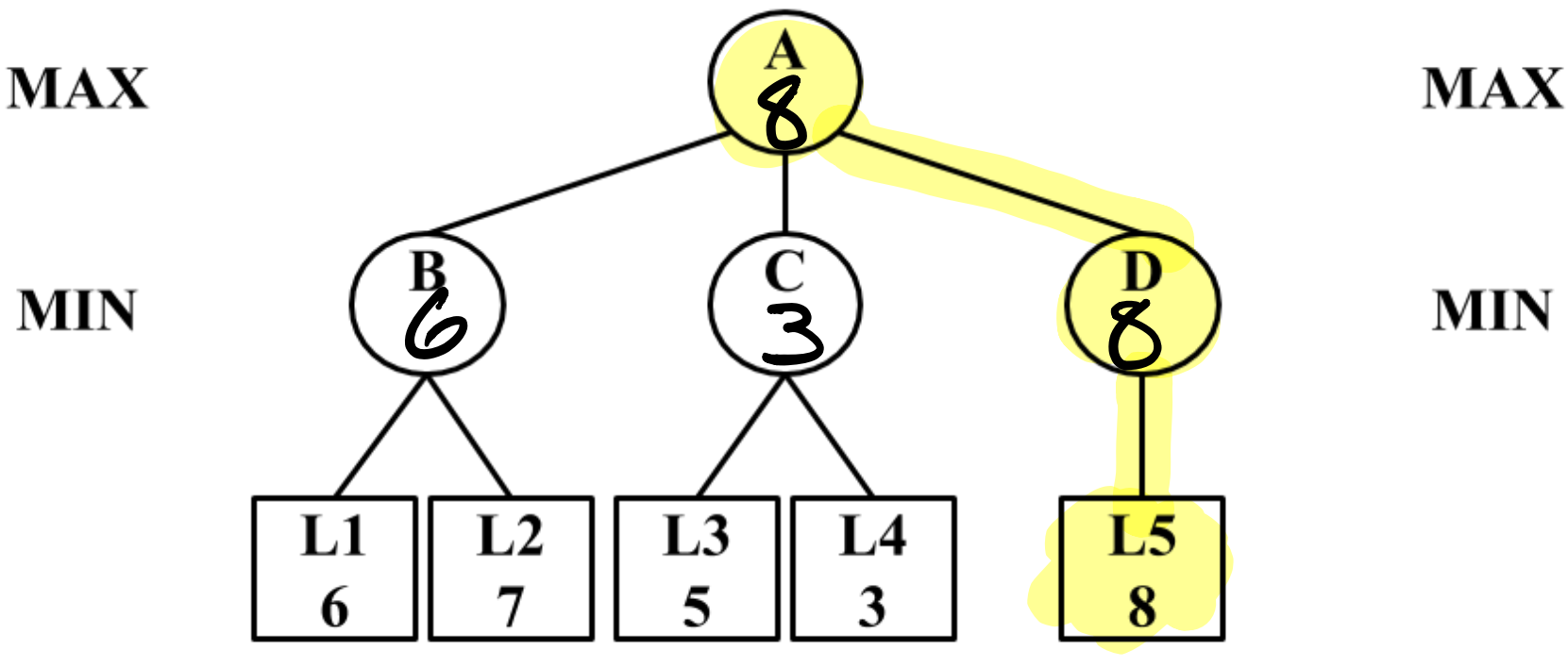
Quiz 1, Problem 2: Games (34 points)

Part A: An Influencer's Tug of War (10 points)

Your friendly neighborhood 6.034 YouTuber Victor Xavier is checking the status of his fanbase on social media before going to bed one evening. He stumbles upon a Piazza post in which students were complaining about the number of hours required to complete all the coursework. Victor, dismayed that his loyal fans were spending their free time studying instead of watching his content, devises a plan to remedy the situation.

Victor creates a game tree with circles depicting **decision nodes,** labeled A through D, and squares depicting **leaf nodes** labeled L1 through L5. The static evaluation at each leaf node represents the expected number of hours per week 6.034 students will spend completing all the required coursework, including Labs, SRNs and studying for the 6.034 Celebrations of Knowledge (a.k.a. Quizzes).

Victor aims to **minimize** the number of hours per week students take to complete all their coursework so that they can spend more time earning him advertisement revenue by watching his YouTube videos. His dissenting fans seek to **maximize** the number of hours they have available to work on their assignments and **make the first move**.



4.1 (8 points) Perform Minimax (without alpha-beta pruning) on the **tree above**, using the static evaluation values given by the tree's leaf nodes (L1 – L5). **Write each decision node's value inside the node**.

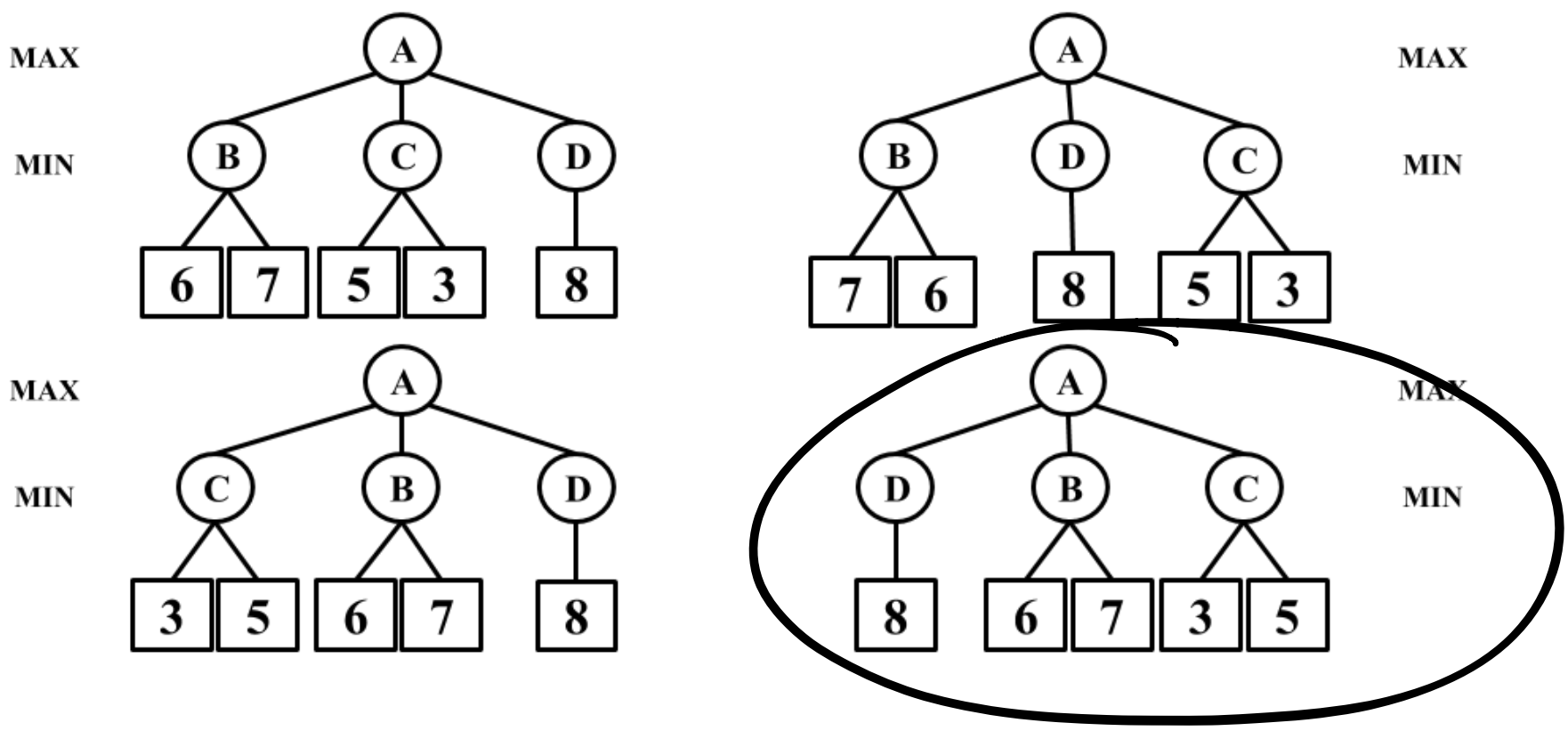
4.2 (2 points) What is the Minimax path? Write the nodes in order from start node on the left to leaf node on the right.

A - D - L5

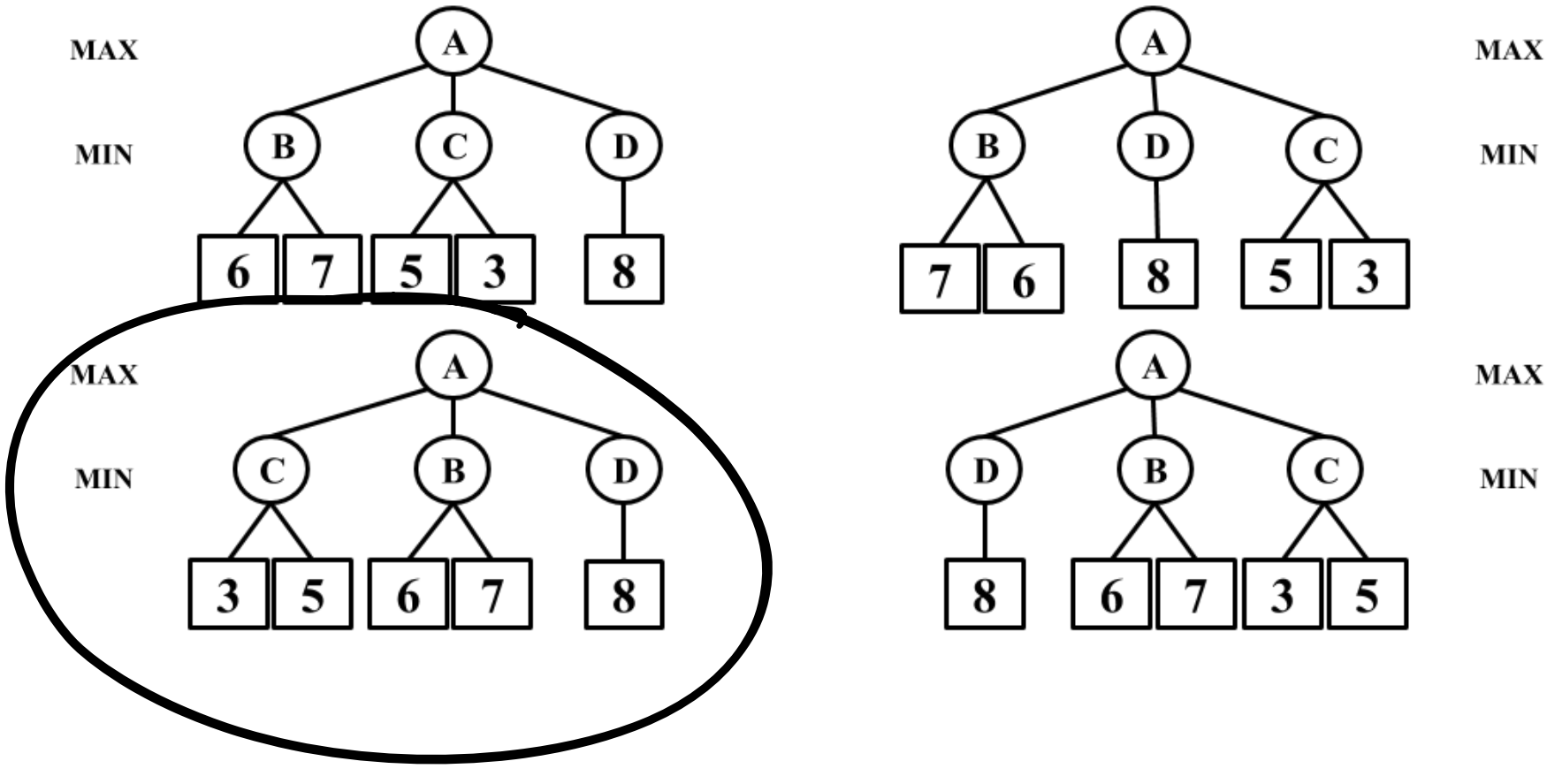
Part B: No time to watch YouTube (6 points)

Victor wants to expedite the tree evaluations from Part A in order to get back to making YouTube videos for his now not-so-loving fans. He decides to **reorder the tree** in order to maximize the number of branches he would be able to prune when performing alpha-beta pruning.

5.1 (3 points) From among the trees printed below, circle the **one tree that would <u>maximize the number of pruned branches (and sub-branches)</u> during alpha-beta pruning**.

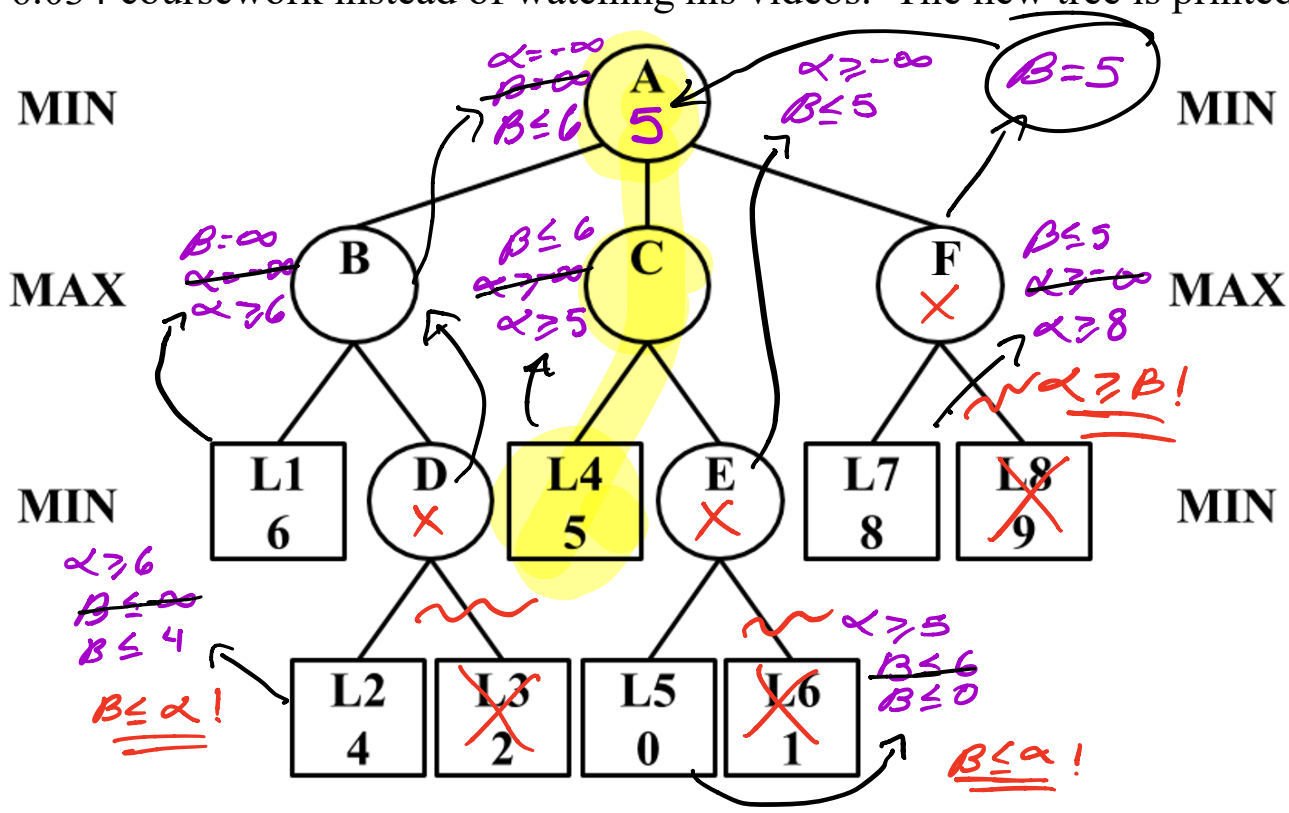


5.2 (3 points) From among the trees printed below, circle the **one tree that would cause <u>NO branches to be pruned</u> during alpha-beta pruning**.



Part C: Emergency Clickbait (18 points)

Victor does not like the prospect of losing viewership on any of his videos, so he decides to **search the game tree one level deeper** and to **take the first turn by minimizing** the expected number of hours per week his fans will spend on 6.034 coursework instead of watching his videos. The new tree is printed below.



6.1 (5 points) After performing alpha-beta pruning on the tree above, which **leaf nodes** would be pruned, i.e., not evaluated? (In the list below, circle **all** that apply. If no leaf nodes are pruned, circle NONE.)

Note: You may show your work on the tree above for partial credit.

L1 **L2** **L3** **L4** **L5** **L6** **L7** **L8** **NONE**

6.2 (5 points) Which **decision nodes** are pruned, i.e., not assigned values? (In the list below, circle **all** that apply. If no decision nodes are pruned, circle NONE.)

Note: You may show your work on the tree above for partial credit.

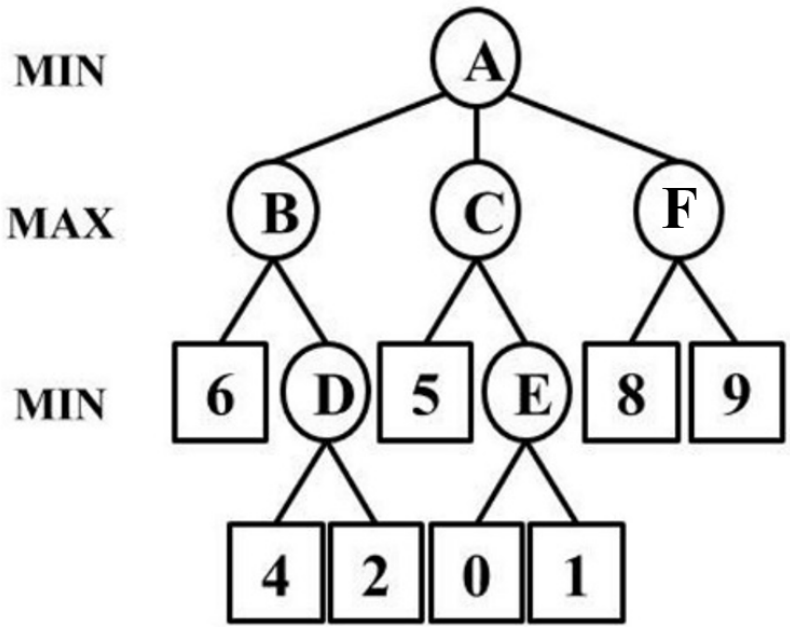
A **B** **C** **D** **E** **F** **NONE**

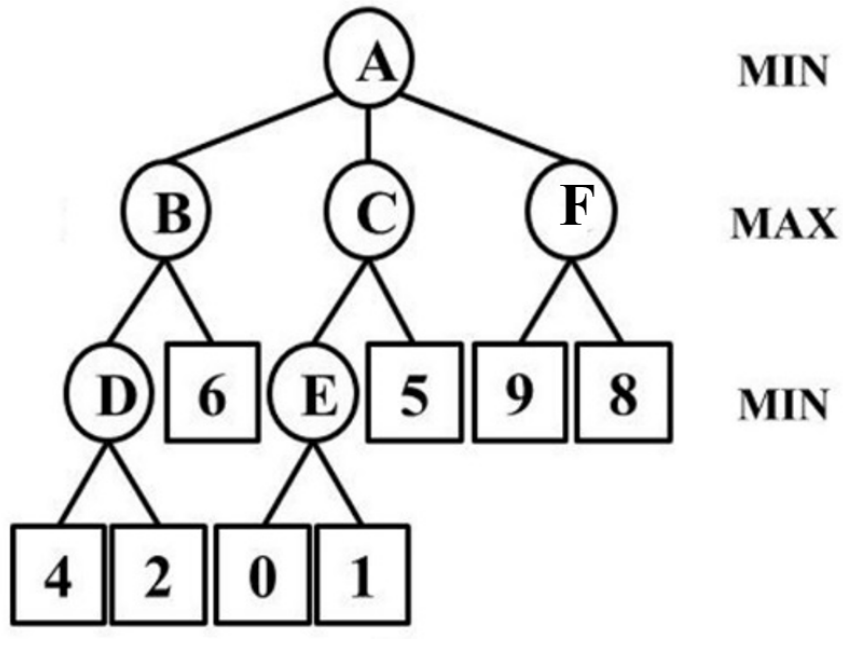
6.3 (2 points) What is the new Minimax path? Write the nodes in order from root node to leaf node.

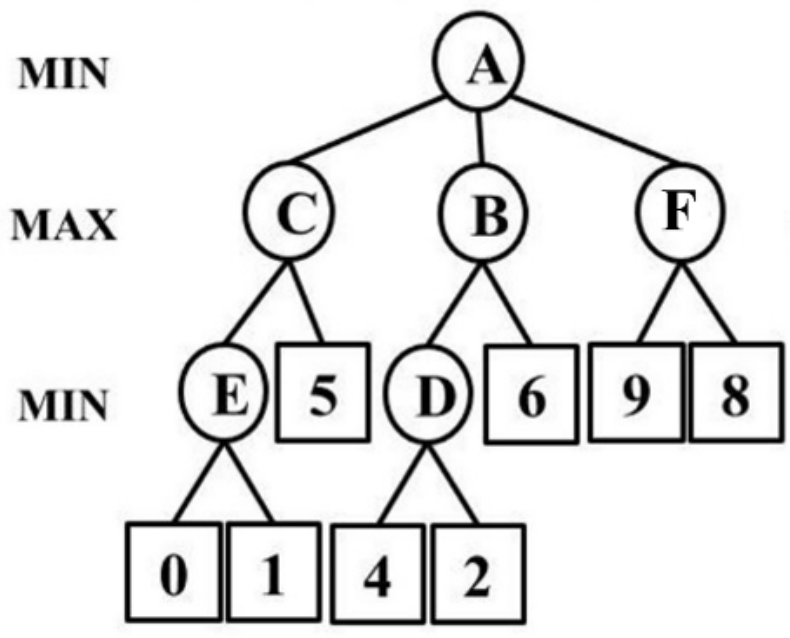
A - C - L4

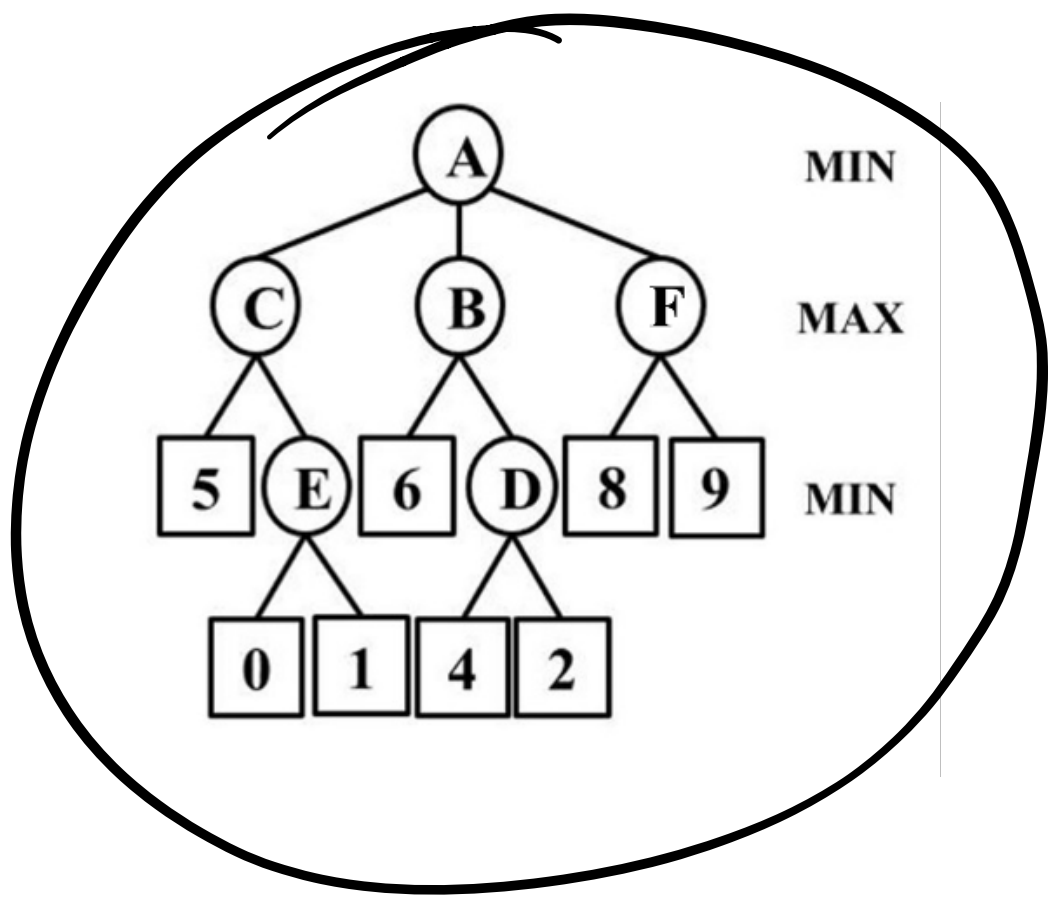
NOTE: Here, pruned decision nodes are those that have a pruned leaf node as a child

6.4 (3 points) Now that Victor has made the game tree deeper, he believes **progressive deepening (d=1) with reordering,** will be necessary once again in order to save himself some work by pruning more branches. Circle the one tree that one **would cause <u>more branches (and sub-branches) to be</u> <u>pruned</u> during alpha-beta pruning than the tree from Part 6.1**. For decision nodes without static values assigned, use the values they would acquire after Minimax as a heuristic estimate. (<u>Circle only one tree</u>.)

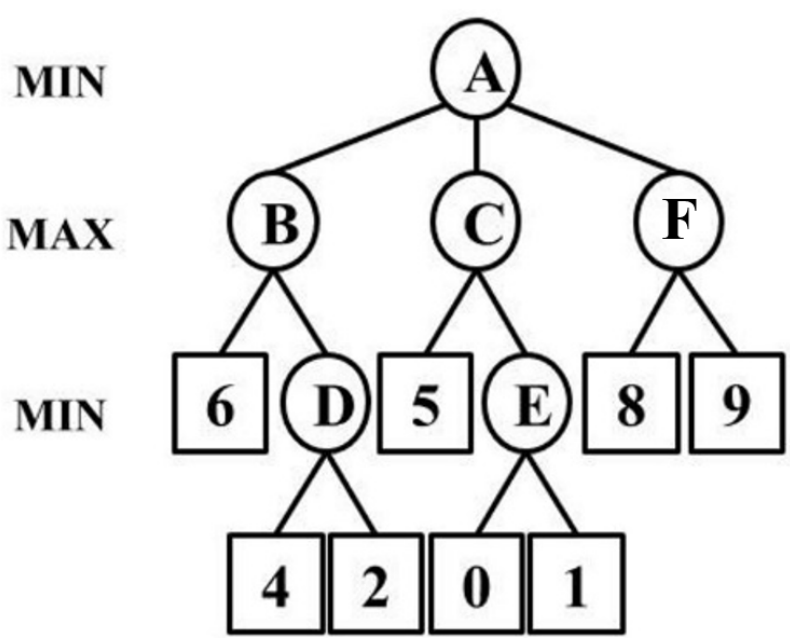


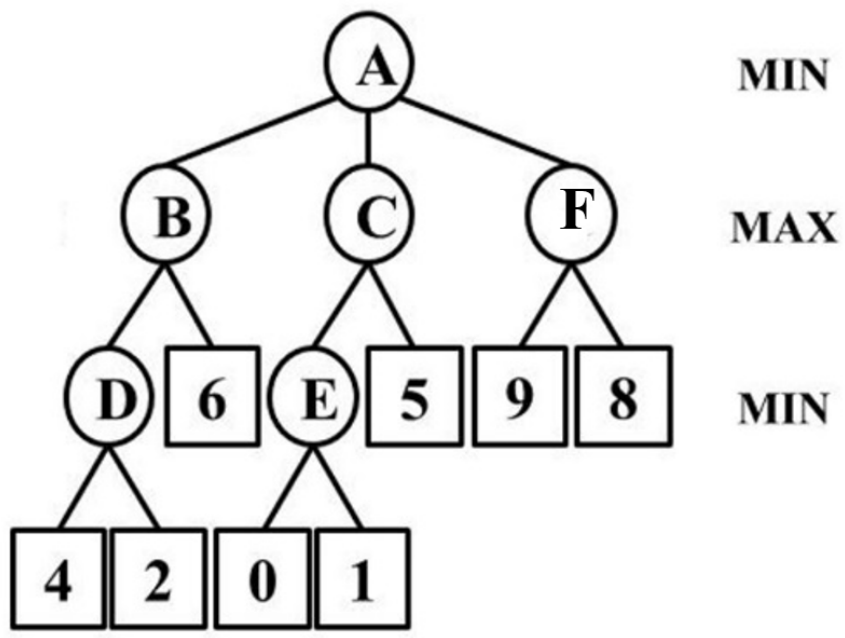


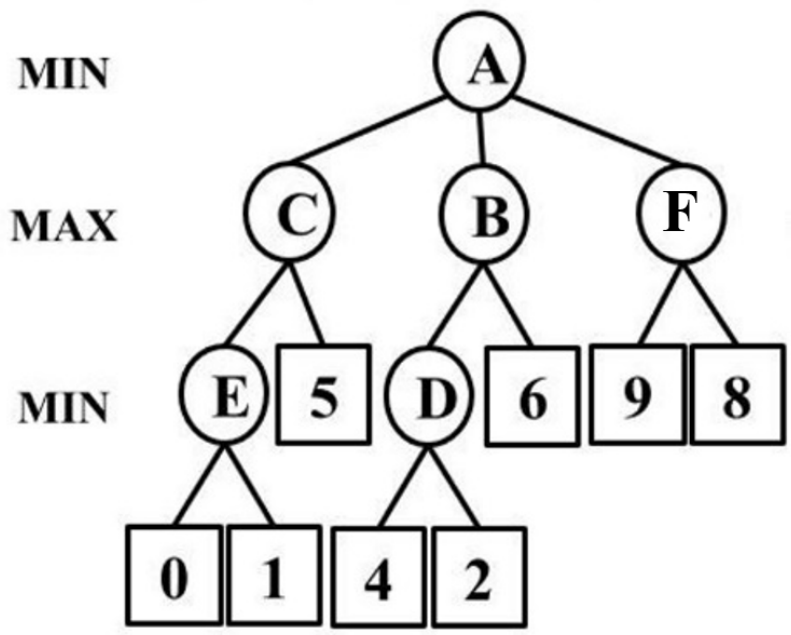


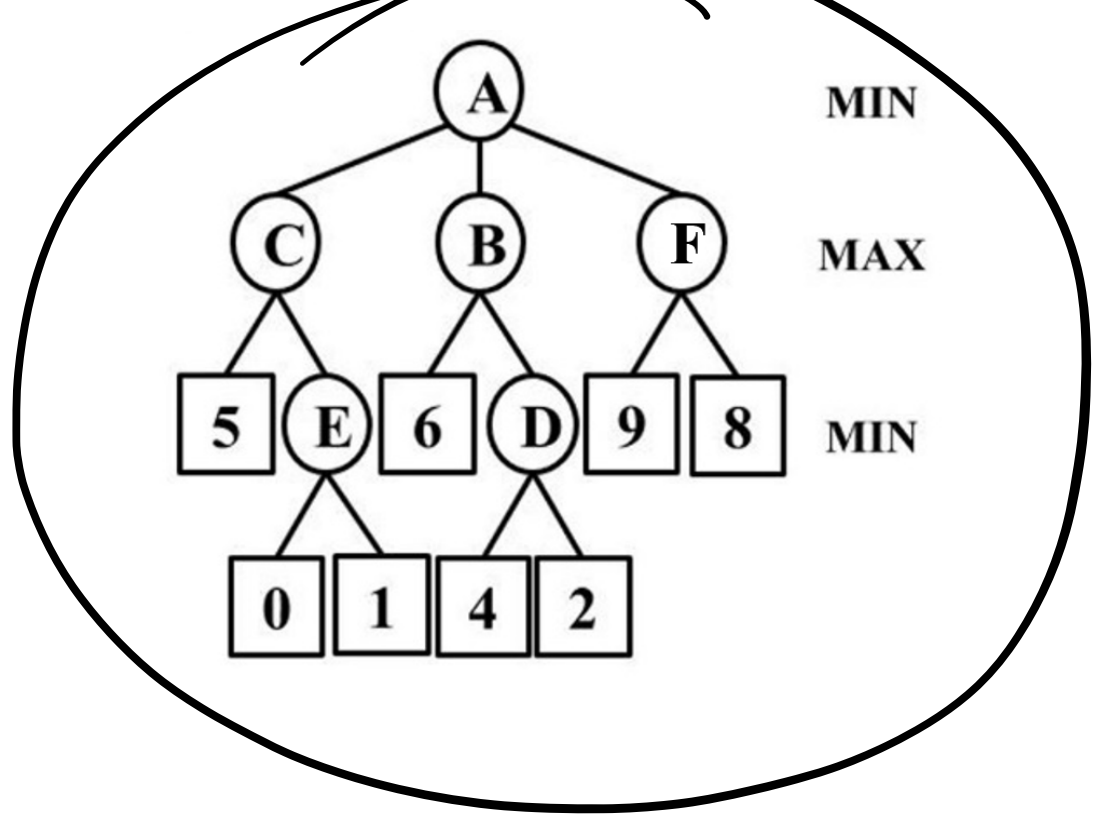


6.5 (3 points) Victor thinks a different value of d will work better, so he tries **progressive deepening (d=2) with reordering**. Circle the one tree that one **would cause <u>more branches (and sub-branches) to be pruned</u> during alpha-beta pruning**. For decision nodes without static values assigned, use the values they would acquire after Minimax as a heuristic estimate. (<u>Circle only one tree</u>.)









Quiz 1, Problem 3: Constraint Propagation (33 points)

Part A: Three Friends (14 points)

Three good friends Inigo **(I)**, Fezzek **(F)**, and Vizzini **(V)** are trying to protect Princess Buttercup **(B)** from the Dread Pirate Roberts. They decide to challenge the pirate to three tests: one of skill, one of strength and one of wits.

Inigo **(I)**, Fezzek **(F)**, and Vizzini **(V)** do not want to overlap their efforts and must all participate in different tests. Vizzini **(V)** decides that it would be best for Buttercup **(B)** to stay with him and participate in the same test.

For your convenience, a copy of the test table and constraint graph is provided on a tear-off sheet.

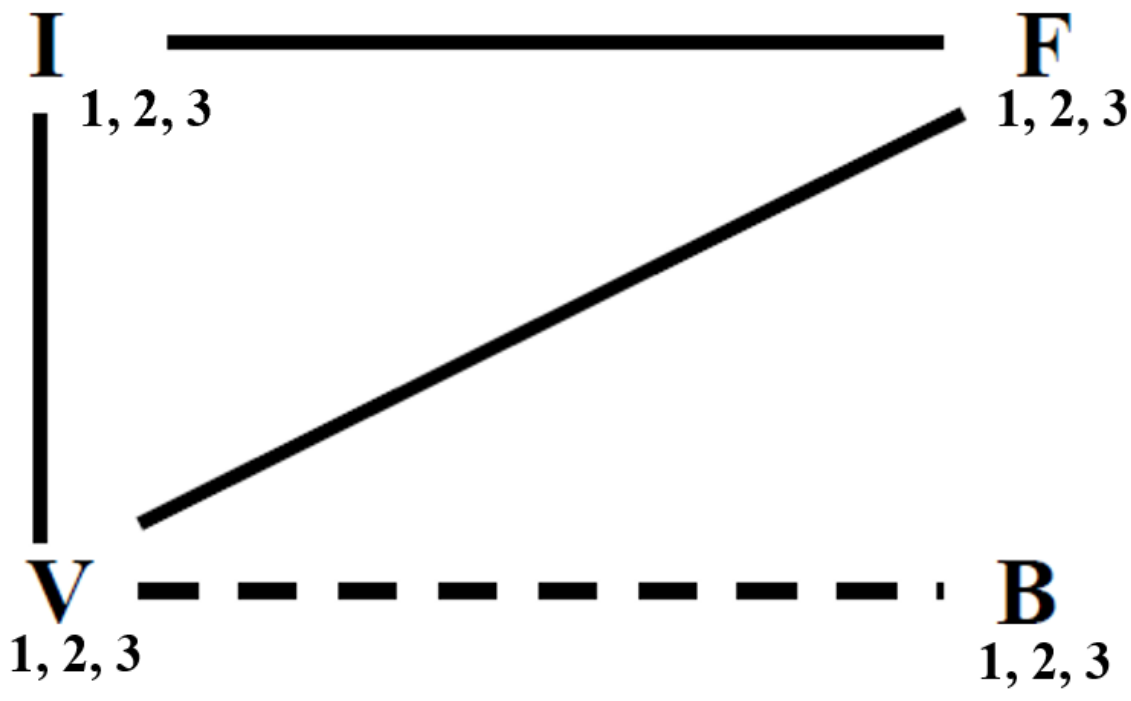
Test Number	Name
1	Test of Skill
2	Test of Strength
3	Test of Wits

Constraints

- **I**, **F**, **V** cannot be the same
- **V** and **B** must be the same

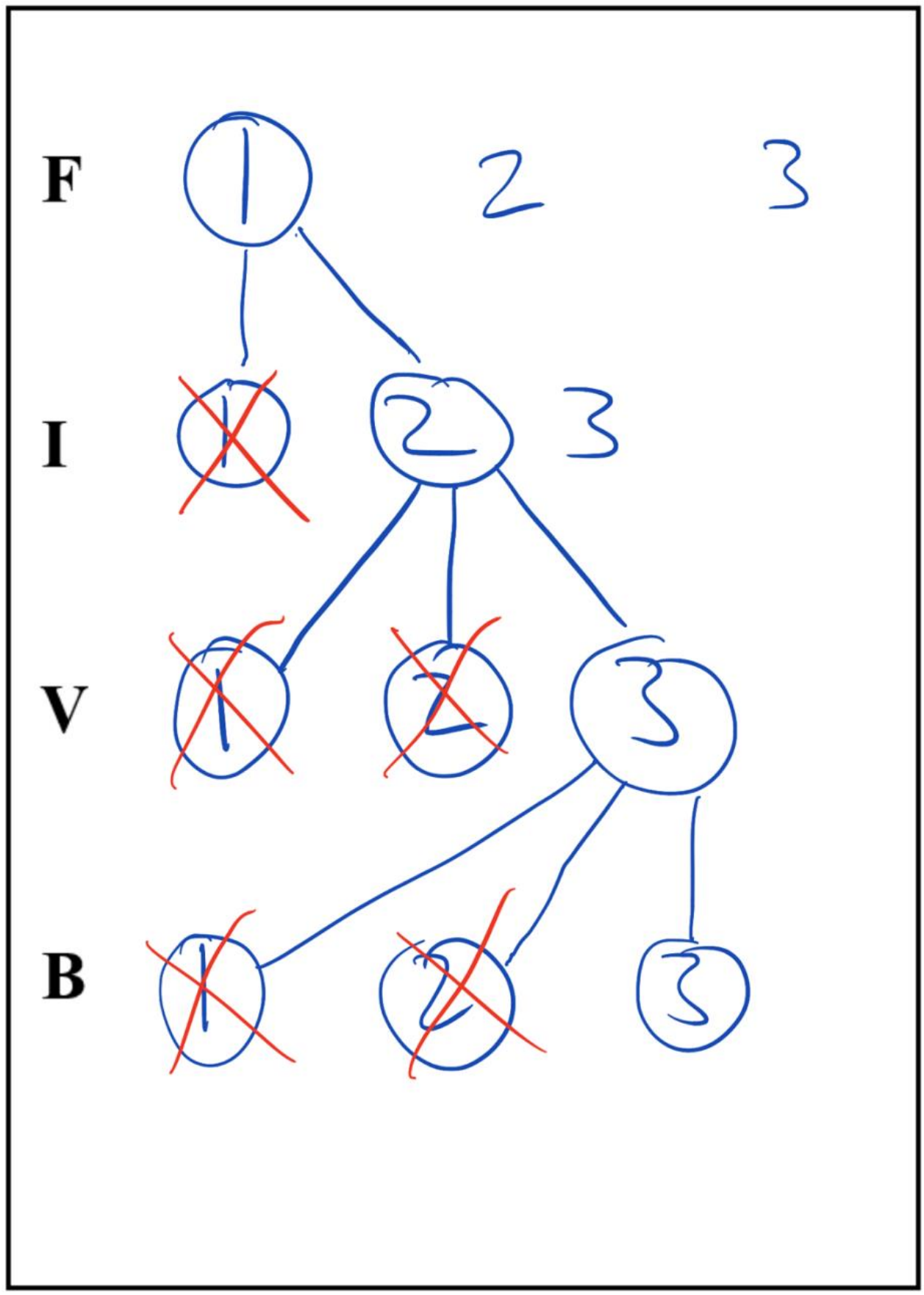
The constraint diagram is shown below.

- Solid Line (——) represents "cannot be in the same test"
- Dashed Line (- - - -) represents "must be in the same test"



7.1 (6 points) Perform **DFS without Forward Checking** assigning in the order: **F → I → V → B**
Use the number of each test in making assignments (e.g. represent "Test of Skill" as "1") and break ties numerically by selecting the smallest number.

Draw your search tree for **DFS without Forward Checking.** Denote assignments by circling the value assigned. Denote backtracking by drawing an X on the assignment that leads to backtracking.



7.2 (2 points) Which of the below assignment orders would result in the least number of assignments?
Circle one.

I → F → V → B **F → I → B → V** **V → B → F → I**

7.3 (3 points) Assuming the same assignment order was used, would the assignments from 7.1 be the same or different if **DFS with Forward Checking** was used instead?
Circle one and explain why in the box below.

SAME **DIFFERENT**

Any sort of pruning covered does not prune solution branches. Since both methods use the same assignment order and DFS, both will return the same assignment.

7.4 (3 points) Assuming the same assignment order was used, would the assignments from 7.1 be the same or different if **DFS with Forward Checking and Propagation through Singleton Domains** was used instead?
Circle one and explain why in the box below.

SAME **DIFFERENT**

Any sort of pruning covered does not prune solution branches. Since both methods use the same assignment order and DFS, both will return the same assignment.

Part B: Prince Humperdinck (11 points)

Prince Humperdinck **(H)** arrives and the group decides to add a fourth test, a test of bravery. Humperdinck **(H)** is concerned for Princess Buttercup's **(B)** safety and wants to be in a test adjacent to Buttercup **(B)**, but not in the same test. Humperdinck **(H)** has already done the Test of Strength **(2)** before so he **cannot** participate in that test.

For your convenience, a copy of the test table and constraint graph is provided on a tear-off sheet.

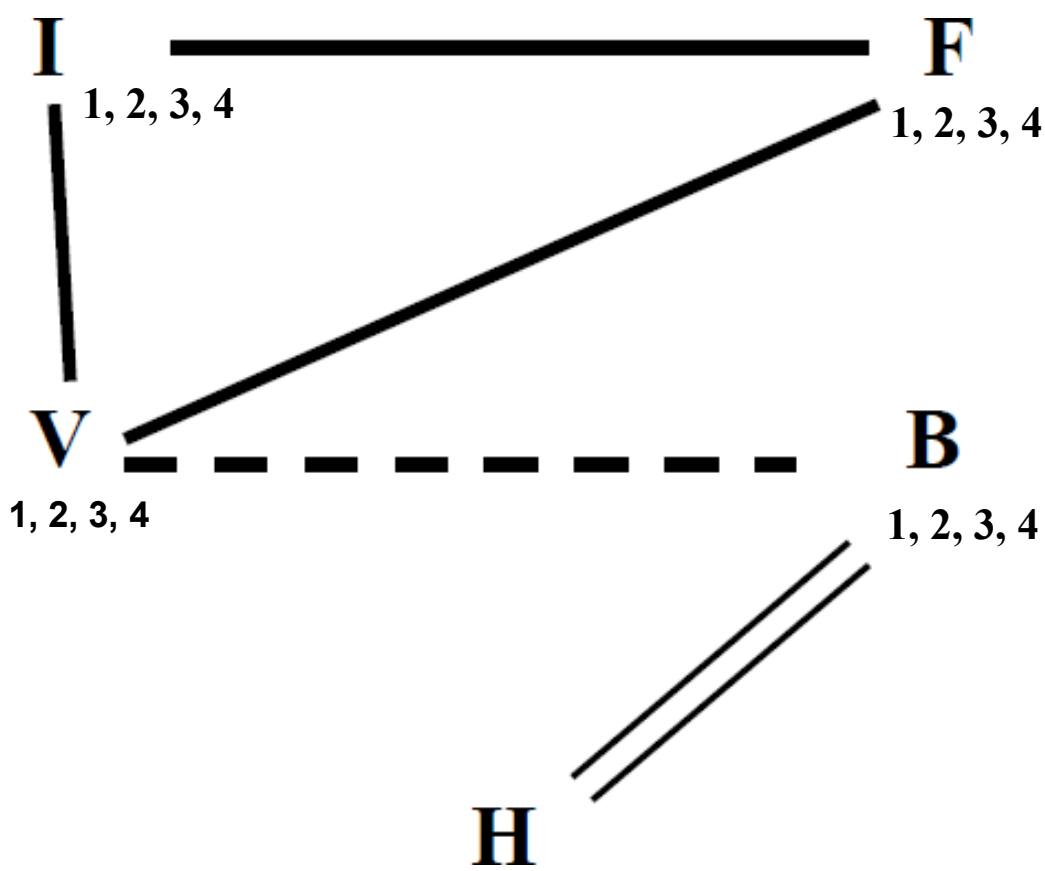
Test Number	Name
1	Test of Skill
2	Test of Strength
3	Test of Wits
4	Test of Bravery

Constraints

- **I**, **F**, **V** cannot be the same
- **V** and **B** must be the same
- **H** cannot be 2
- **H** and **B** must be adjacent (but not the same)

The constraint diagram is shown below.

- Solid Line (——) represents "cannot be in the same test"
- Dashed Line (- - - -) represents "must be in the same test"
- Double Solid Line (===) represents "must be adjacent (but not the same)"



8. (11 points) Perform **DFS + Forward Checking + Propagation Through Singleton Domains** to find a solution, by filling in the worksheet. Assign variables in this order: **I F V B H**
Break ties numerically by smallest test number when making assignments.
Break ties alphabetically when adding variables to the queue.

Show your work on the next page for partial credit.

	Variable de-queued /assigned	List all values just eliminated from neighboring vars or NONE	Back track?
1	**I = 1**	**F, V != 1**	
2	**F = 2**	**V != 2**	
3	**V = 3**	**B != 1, 2, 4 (B reduced to singleton)**	
4	**B**	**H != 1, 3 (H reduced to singleton)**	
5	**H**	**None**	
6	**B = 3**	**None**	
7	**H = 4**	**None**	
8			
9			
10			
11			

Domain worksheet instructions:

1. Every time you **assign a variable** or **remove a variable from the propagation queue**, fill out a new row in the table. There may be more rows than you need.
2. In that row, indicate **which variable you assigned or de-queued**; write its **assigned value** if it has one (e.g. X=x), otherwise just write its **name** (e.g. X). In the second column, list the **values that were just eliminated from neighboring variables** as a result (or "NONE" if no values were eliminated). Do not eliminate values from variables that have already been assigned.
3. If your search has to backtrack after assigning or de-queuing a variable: First, **finish listing** all values eliminated from neighboring variables. Next, check the "backtrack" box in that row.

Example row showing an assigned variable (with backtracking)

ex	X = 3	Y ≠ 3, 4 Z ≠ 3 (example)	☑

Example row showing a de-queued (propagated) variable

ex	X	W ≠ 1, 4 (example)	☐

Part C: Conceptual Questions (8 points)

For each of these questions about constraint satisfaction problems, circle one answer.

9.1 (2 points) By choosing a specific order for variables, you can control whether a constraint-satisfying solution exists.

TRUE (**FALSE**)

9.2 (2 points) Changing the order of variables can affect the number of assignments tried during depth first search (DFS) and forward checking.

(**TRUE**) **FALSE**

9.3 (2 points) Which one of the following is **NOT** true about DFS + Forward Checking + Propagation Through Reduced Domains? *Circle one.*

A. It reduces the number of potential assignments by maximizing propagation.
B. It has more overhead than other constraint satisfaction algorithms.
C. It never makes an assignment that violates a constraint.
(D.) It will never reduce a domain to size zero.

9.4 (2 points) Which of the following algorithms performs search such that backtracking is never necessary? *Circle one.*

A. DFS
B. DFS + Forward Checking
C. DFS + Forward Checking + Propagation Through Singleton Domains
D. DFS + Forward Checking + Propagation Through Reduced Domains
(E.) None of these

Quiz 2, Problem 1: Rules (50 points)

Part A: Backward Chaining (15 points)

Shirley wants to stay up to date with current events, but often gets overwhelmed by the sheer quantity (and variability in quality) of news sources. She does not have time to check all of them, and she doesn't want to be overly reliant on one source. So she opts to use a **rule-based system** to decide which news outlet to check every morning. **A tear-off sheet is provided for the below rules and assertions.**

Make the usual 6.034 backward chaining assumptions, including but not limited to:

- Rules are tried in the order they appear.
- Antecedents are tried in the order they appear in a rule.
- Short circuiting is in effect.
- The backward chainer never alters the list of assertions.

Rules:

P0	**IF OR**('**(?x)** has qualms with state-owned media' '**(?x)** doesn't speak Russian') **THEN** ('**(?x)** doesn't read Российская Газета')
P1	**IF NOT**('**(?x)** loves Glen Beck), **THEN** ('**(?x)** doesn't watch Fox News')
P2	**IF AND**('**(?x)** is feeling optimistic '**(?y)** believes the government is corrupt') **THEN** ('**(?x)** has qualms with state-owned media')
P3	**IF AND**('**(?x)** ate an egg salad sandwich for lunch yesterday', **NOT** ('**(?x)** doesn't speak Russian')) **THEN** ('**(?x)** watches Fox News')
P4	**IF AND**('**(?x)** doesn't read Российская Газета', '**(?y)** is a self-proclaimed "raging socialist"', '**(?x)** doesn't watch Fox News', '**(?x)** is feeling optimistic') **THEN** ('**(?x)** asks **(?y)** for the latest news')
P5	**IF AND**('**(?y)** is always on Twitter', '**(?x)** is always on Twitter')) **THEN** ('**(?x)** and (?y) check Twitter')

Assertions:

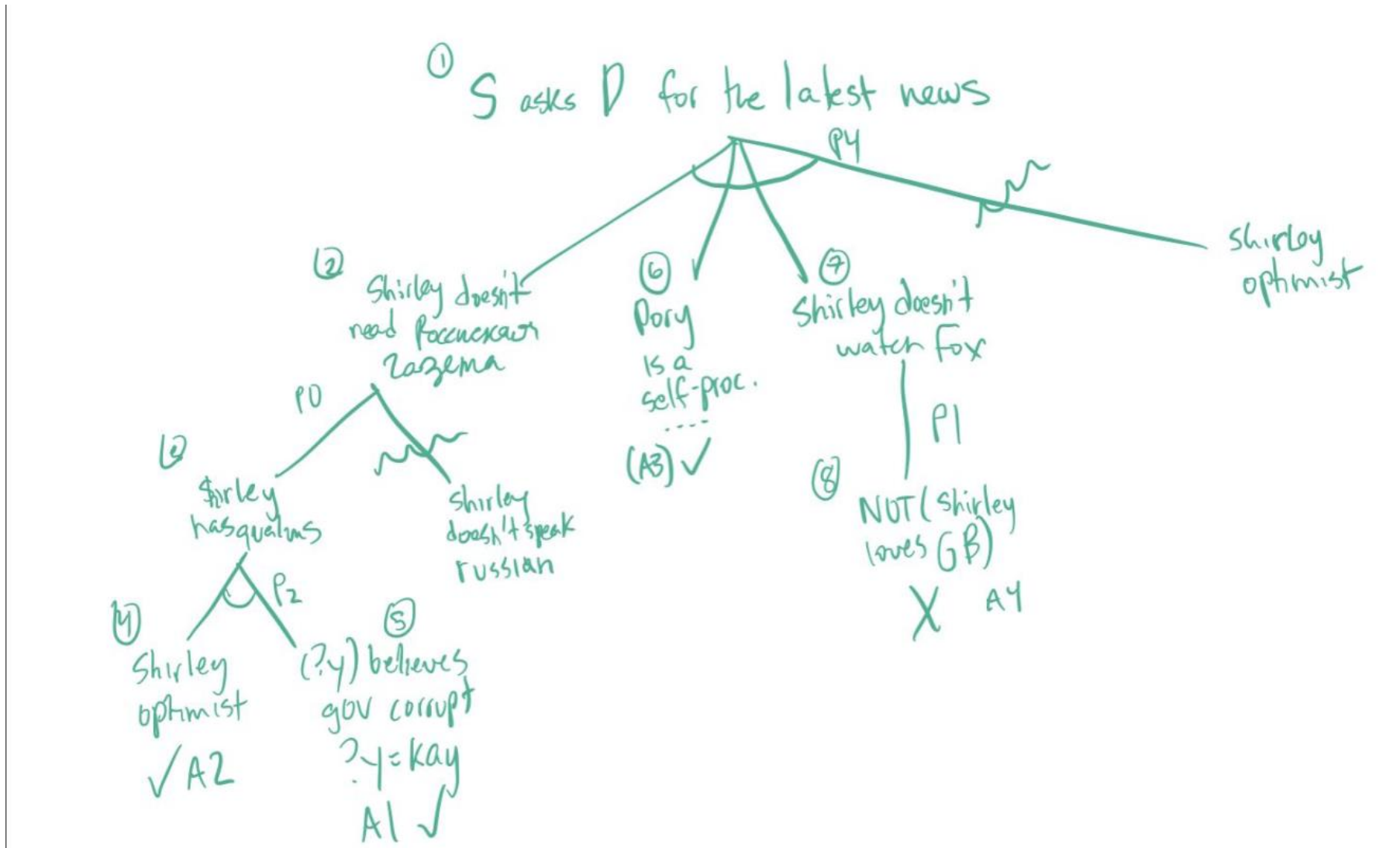
A0: Shirley speaks Russian
A1: Kay believes the government is corrupt
A2: Shirley is feeling optimistic
A3: Dory is a self-proclaimed "raging socialist"
A4: Shirley loves Glen Beck

Use backward chaining to decide whether the following hypothesis is true:

Shirley asks Dory for the latest news

10.1 (13 points) Perform backward chaining using the rules and assertions on the previous page. In the table on the next page, list all the hypotheses *in the order that they are checked* (there may be more lines than you need).

Draw the goal tree below for partial credit. Mark short-circuited branches with a squiggle (~).



H1.	Shirley asks Dory for the latest news
H2.	Shirley doesn’t read Российская Газета
H3.	Shirley has qualms with state-owned media
H4.	Shirley is feeling optimistic
H5.	**(?y)** believes the government is corrupt, (?y)=Kay
H6.	Dory is a self-proclaimed “raging socialist”
H7.	Shirley doesn’t watch Fox News
H8.	**NOT**(Shirley loves Glen Beck) XXX
H9.	

During backward chaining:

10.2 (1 point) How many times did you short-circuit? 2

10.3 (1 point) Is the hypothesis “Shirley asks Dory for the latest news” true?

YES **NO** CAN’T TELL

Part B: Reverse Engineering Forward Chaining (22 points)

You have recently come into an obsession with playing Sims and suspect that it uses a rule-based system to determine what the Sims do. Unfortunately, the code that contains the rules is well protected by a 256-bit encryption scheme. Fortunately, you were able to snag the table that got filled out that was helpfully labeled **forward chaining.**

11.1 (19 points) Assuming that the Sims game performed flawless forward chaining, following standard 6.034 assumptions and conventions to generate these new assertions, infer the original set of rules and assertions. For reasons unbeknownst to anyone, the table includes:

- the rules whose antecedents match the assertions
- the rule that fires
- the binding for the fired rule
- any new assertions added.

The table even has written None for you when no rules fire. **A tear-off sheet will be provided.**

On the next page, fill in the rules & assertions that make the below forward chaining possible.
You have two Sims, Libby and Ziyi. This table describes the events of an exhilarating Thursday night.

Step	Matched	Fired	Rule Instance Bindings	New Assertion(s) Added
1	P1, P2	P1	(?x) = Libby	A3: Libby is bored
2	P1, P2, P5	P2	(?z) = Libby	A4: Libby wants to eat
3	P1,P2,P4	P2	(?z) = Ziyi	A5: Ziyi wants to eat
4	P1-4	P3	(?x) = Ziyi	A6: Ziyi makes a snack
5	P1-4	P4	(?x) = Libby	A7: Libby orders thai food
6	P0-4	P0	(?x) = Libby	A8: Libby is content
7	P0-4	None	None	None

Fill in the rules & assertions that make the forward chaining on the previous page possible. Here is a helpful head start. There are five rules and 3 original assertions.

Rules:

P0	***IF*** ***'(?x)*** *orders thai food',* ***THEN*** *(* ***'(?x)*** *is content')*
P1	**IF** ***AND***(**'(?x)** finished work early', ***NOT*** (**'(?x)** has plans')) **THEN** (***'(?x)*** *is bored'*)
P2	**IF** ***OR***(***'(?z)*** *is hungry'*, ***'(?z)*** *is bored'*) **THEN** (***'(?z)*** *wants to eat'*)
P3	**IF** ***AND***(***'(?x)*** *wants to eat'*, **NOT** (**'(?x)** has money to spend')) **THEN** (***'(?x)*** *makes a snack'*)
P4	**IF** **AND**(***'(?x)*** *wants to eat'*, **'(?x)** has money to spend') **THEN** (***'(?x)*** *orders thai food'*)
P5	**IF** **AND**(**'(?y)** is bored', ***NOT*** *(****'(?y)*** *wants to eat')*) **THEN** (**'(?y)** scrolls through instagram')

Assertions:

A0: *Libby finished work early*

A1: Ziyi is hungry

A2: *Libby has money to spend*

11.2 (3 points) What is the maximum number of rules and assertions that would still yield the same forward chaining table?

∞

Part C: RBS IRL (13 points)

Rule Based Systems. In Real Life.
Raspberry Blueberry Scones. Iced Really Lovingly.
Ruth Bader. Strong Iconic Ruthless Lawful.

12.1 (1 point) What else could RBS IRL stand for? Give one clever example.

Ring bearer shenanigans illogical rueful love

You will be given two problems/situations that you should attempt to resolve **using an RBS approach.** For each situation, give the requisite number of example rules. And think about why RBS might be a good solution and what some of its limitations might be.

Example:

> **Task:** Bouncer at a bar
>
> **Example rule with OR:**
> ```
> IF  OR( '(?x)'s ID says they are under 21'
>         '(?x) is too inebriated')
> THEN ( 'Do not let (?x) in')
> ```
>
> **Limitations:**
> - What does "too inebriated" mean? There could be another rule that defines that based on BAC, but that varies from person to person.
> - Fake IDs. Would need another rule to determine authenticity of ID.

Note: This problem is intentionally open-ended, meant for you to demonstrate your understanding of RBS via the use of your imagination. Although we have some answers in mind for each situation, well-reasoned limitations and reasonable rules will receive full credit.

12.2 (3 points)
Task: College admissions

Provide an example rule with *at least one OR.*

```
IF  OR( '(?x) has below a 3.0 GPA'
        '(?x) got below a 800 on the SAT')
THEN ( '(?x) is not a good academic fit')
```

Provide an example rule with *a nested OR/AND*. (i.e. OR(x, AND(y,z)) or AND(y,OR(x,z))

```
IF  AND( NOT('(?x) is not a good academic fit')
         OR( '(?x)'s personal essay was super compelling',
             '(?x) has stellar recommendations'))
THEN ('Admit (?x)')
```

What is a limitation of relying just on a RBS in this context:

There isn't a hard and fast rule to who should get in. How do you have human judgement when RBS need hard, discrete cutoffs?

12.3 (3 points)

Task: Emergency room triage: Who should the doctor see first?

Provide an example rule with *at least one NOT and one AND.*

```
IF  AND( NOT('(?x) is profusely bleeding'),
         '(?x) is conscious' )
THEN ('Make (?x) wait in the waiting room for hours')
```

Provide an example rule with *at least one OR*.

```
IF  OR( NOT('(?x) is conscious'),
        '(?x) has good health insurance' )
THEN ('Send a doctor to treat (?x) immediately')
```

What is a limitation of relying just on a RBS in this context:

You need to manually add new rules for new situations that arise. Case in point: COVID. The system would not be able to take that into consideration until related rules are hard-coded into the model.

12.4 (3 points)

What is a real-life situation (besides what was already mentioned) where a Rule-Based System would be a good solution? Give a brief explanation why.

- Deciding what type of data visualization to use (pie chart, bar graph, scatter plot, etc.)
- Clearly, well-defined tasks or the situation isn't high stakes

12.5 (3 points)

What is a real-life situation (besides what was already mentioned) where a Rule-Based System should **not** be used (e.g. for ethical reasons, just being ineffective, etc.)? Give a brief explanation why.

- Criminal justice system / court decisions
- Tinder swiping
- Subjective constraints / issues with ethics

Quiz 2, Problem 2: Bayesian Inference (50 points)

Several characters from Shakespeare's plays have landed themselves in baffling predicaments. Your task is to use your powers of Bayesian inference to help them sort out their problems.

Part A: To Bayes, or not to Bayes (5 points)

Hamlet, well-known for his indecisiveness, has several decisions to make:

- **B**: To be (**B**=true) or not to be (**B**=false) (<u>two</u> possible values)
- **F**: Whom to befriend? (**F** has <u>three</u> possible values: Horatio, Rosencrantz, and Guildenstern)
- **L**: Whom to be in love with? (**L** has <u>three</u> possible values: Ophelia, Horatio, and Gertrude)
- **K**: Whom to kill? (**K** has <u>four</u> possible values: Claudius, Polonius, Laertes, and Fortinbras)
- **G**: Whether to believe the ghost of his father? (**G**=true or **G**=false, <u>two</u> possible values)

13.1 (2 points) If nothing is known about the independence of Hamlet's decision variables, how many parameters would be needed to fully specify the joint probability?

143

$$\#\text{params} = \left[\prod_i (\#\text{possible values of } i)\right] - 1$$

$$= (2)(3)(3)(4)(2) - 1$$

$$= 143$$

13.2 (3 points) If all of Hamlet's decision variables are independent of each other, how many parameters would be needed to fully specify the joint probability?

9

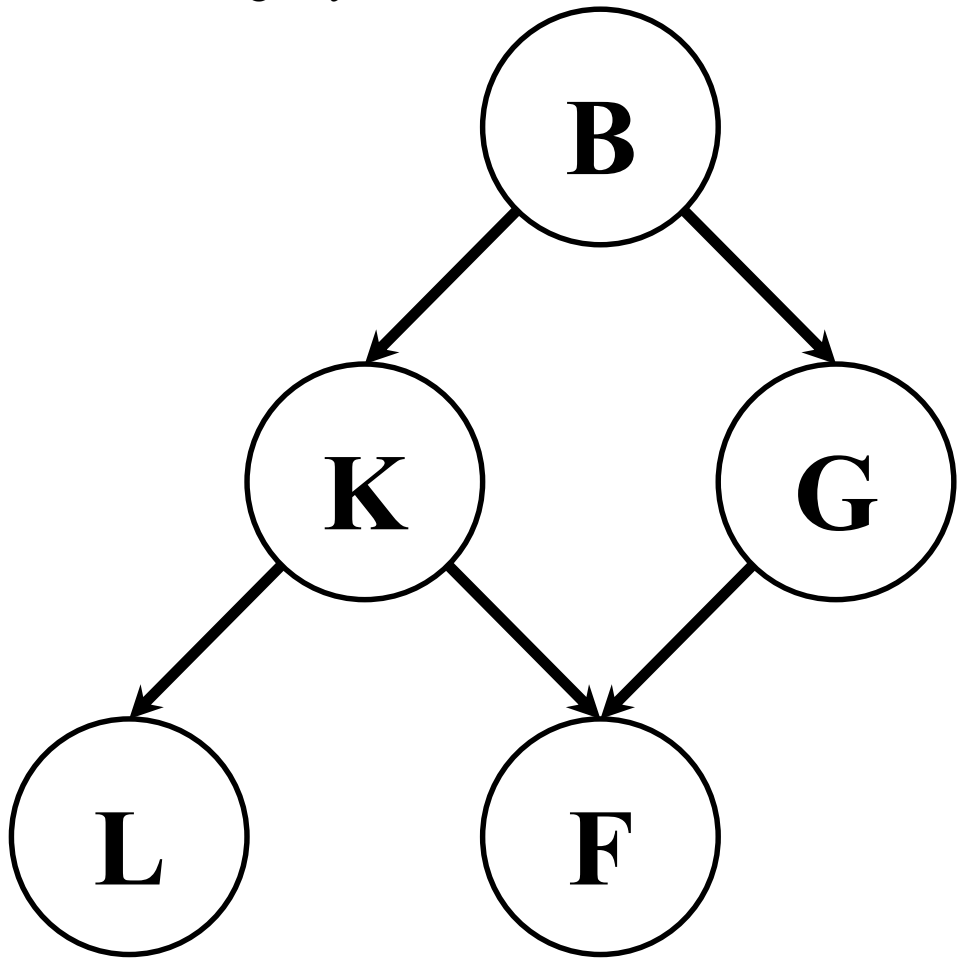
$$\#\text{params} = \left[\sum_i (\#\text{possible values of } i - 1)\right]$$

$$= (2-1)+(3-1)+(3-1)+(4-1)+(2-1)$$

$$= 9$$

Part B: Something is rotten in the state of 6.034 (25 points)

Consider the same decision variables from Part A/Problem 13. To help Hamlet make the right choices, you draw him the following Bayes Net:



For your convenience, a tear-off sheet of the Bayes Net will be provided.

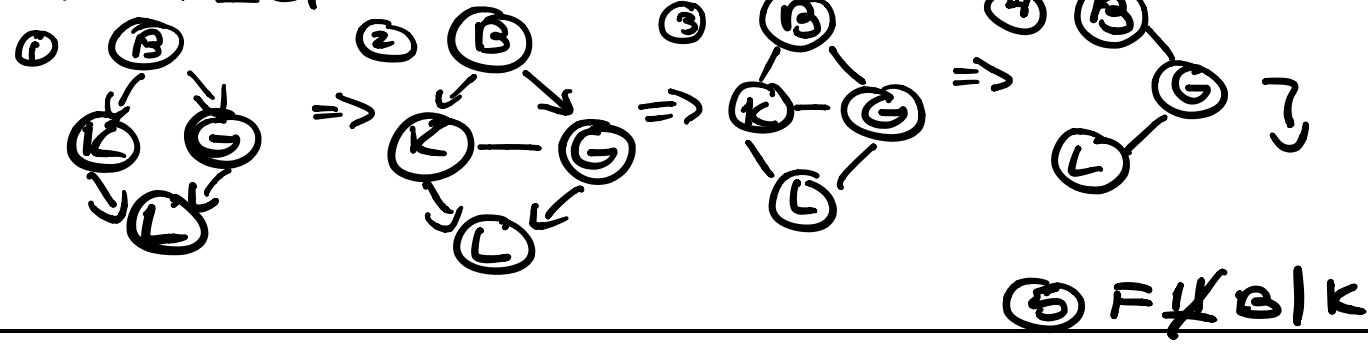
14. (25 points) Ascertain the validity of the following claims regarding **structural** independence. *Hint*: You should *not* need to use d-separation for every claim listed below. (What simplifying assumptions do you know?)

Claim 1:	$\mathbf{F} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{K}$	True	False	Can't Tell
Claim 2:	$\mathbf{K} \perp\!\!\!\perp \mathbf{G} \mid \mathbf{B}$	True	False	Can't Tell
Claim 3:	$\mathbf{L} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{K}$	True	False	Can't Tell
Claim 4:	$\mathbf{K} \perp\!\!\!\perp \mathbf{G} \mid \mathbf{F}$	True	False	Can't Tell
Claim 5:	$\mathbf{L} \perp\!\!\!\perp \mathbf{F} \mid \mathbf{K}$	True	False	Can't Tell

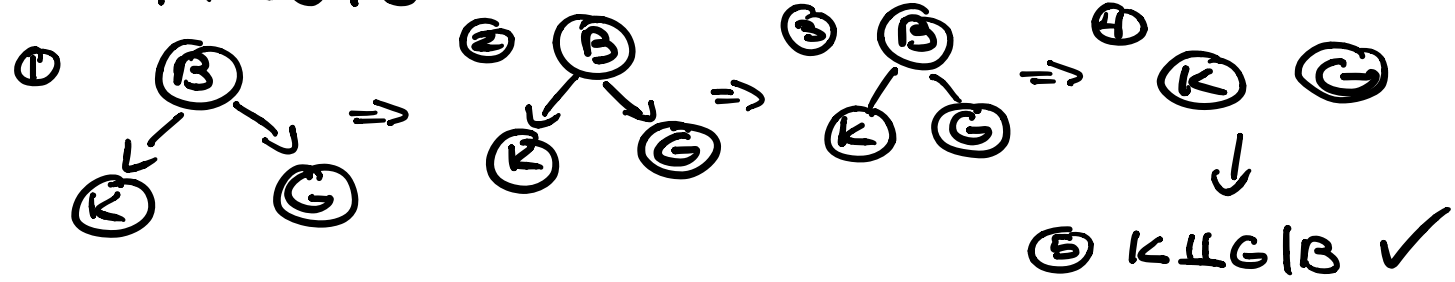
Show your work in the boxes on the following pages for partial credit.

You may show your work in the boxes below for partial credit. (D-separation shown)

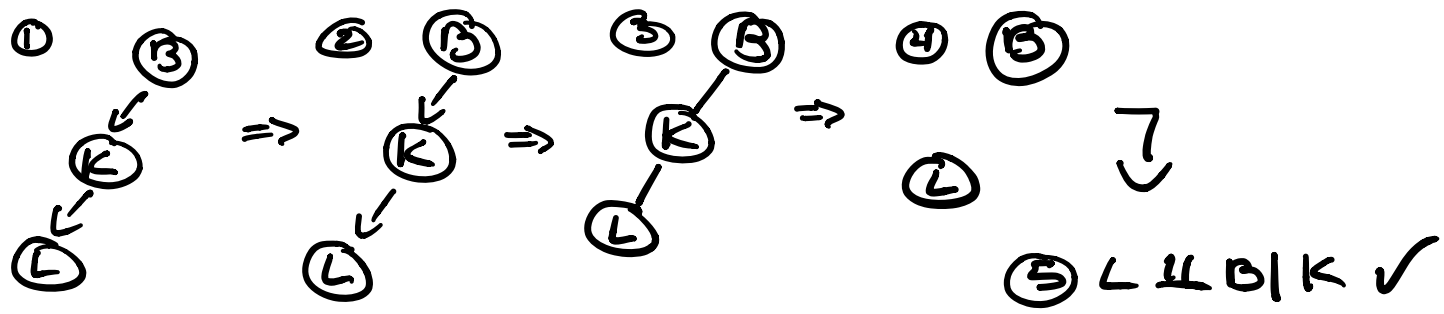
Claim 1: $F \perp\!\!\!\perp B \mid K$



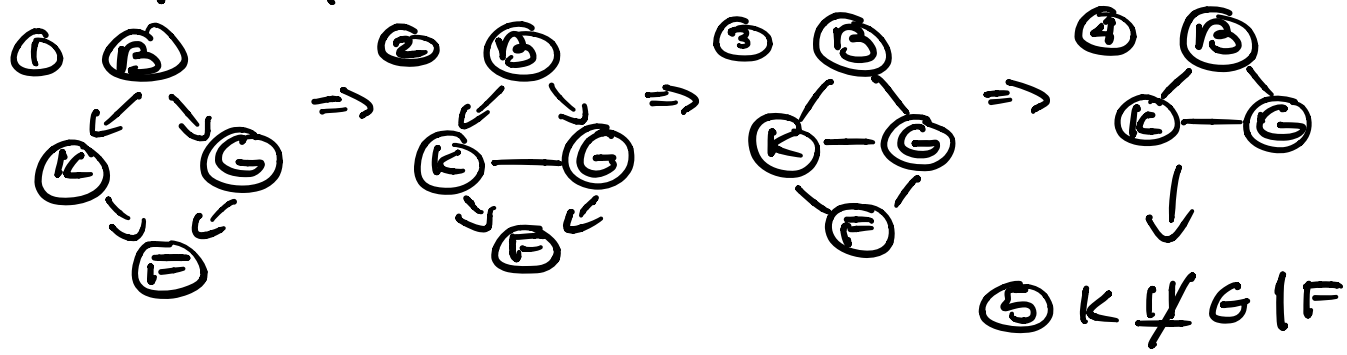
Claim 2: $K \perp\!\!\!\perp G \mid B$



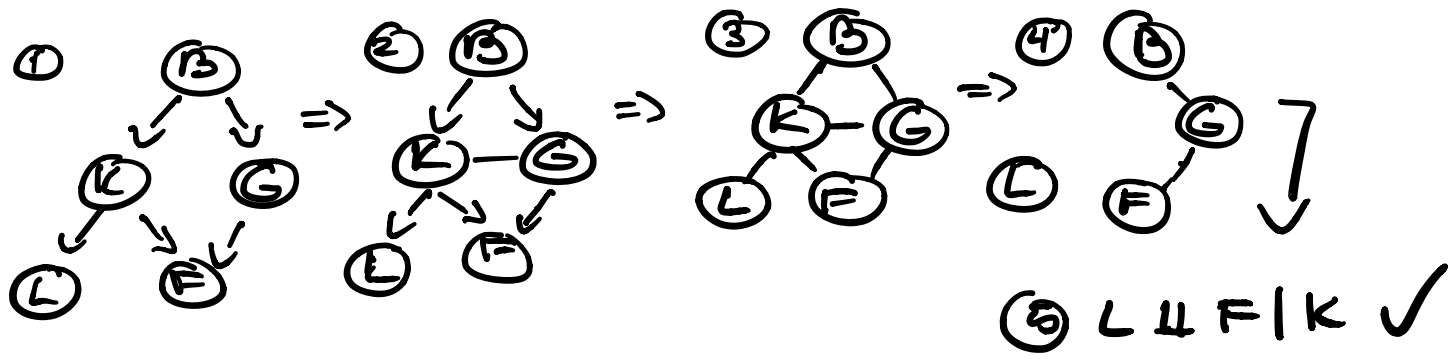
Claim 3: $L \perp\!\!\!\perp B \mid K$



Claim 4: $K \perp\!\!\!\perp G \mid F$



Claim 5: $L \perp\!\!\!\perp F \mid K$

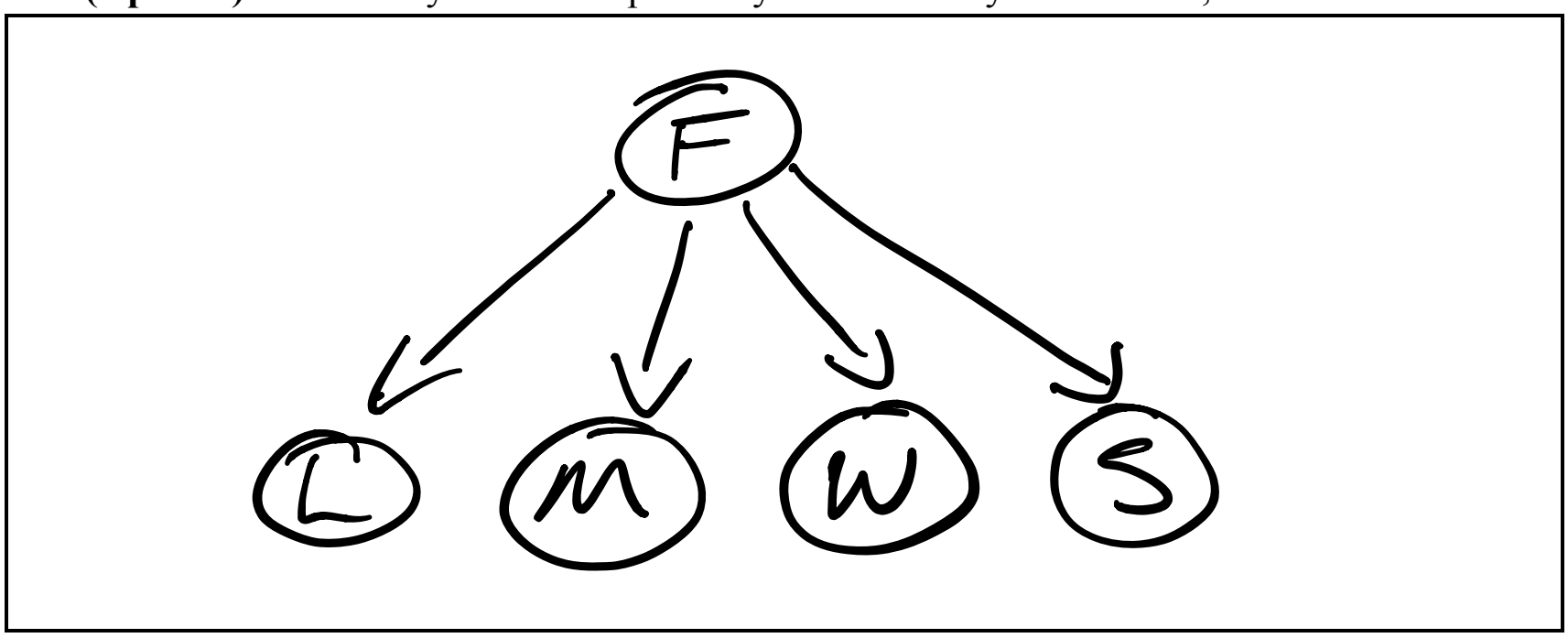


Part C: Romeo, Romeo, wherefore art thou Romeo? (20 points)

Prince Escalus wishes to create a **Naive Bayes Classifier** to determine which family a person belongs to (classification: **F**=Montague or **F**=Capulet), using the following (Boolean) features:

- **L**: Falls in **l**ove easily
- **M**: **M**arries young
- **W**: Is exceedingly **w**ealthy
- **S**: Is a skilled **s**wordsman

15.1 (4 points) Draw a Bayes Net to represent your Naive Bayes Classifier, in the box below.



15.2 (2 points) Given the independence relations encoded in your Bayes Net from Part 15.1, how many parameters would be needed to fully specify the joint probability of this system?

F: 1 parameter
L, M, W, S: 2 parameters each

9

Bayes Net Assumption:
=> (1) + (2) + (2) + (2) + (2)

15.3 (2 points) With your help, Prince Escalus determines the following probabilities associated with each feature, as well as the prior probabilities for each household:

F	*Prior Probability*	*P* (**L**=True | **F**)	*P* (**M**=True | **F**)	*P* (**W**=True | **F**)	*P* (**S**=True | **F**)
Montague	0.5	0.9	0.2	0.5	0.6
Capulet	0.5	0.5	0.8	0.5	0.4

Which, if any, feature can we ignore in building our Naive Bayes Classifier? (If none, write NONE.)

W

P(W | F = Montague) = P(W | F = Capulet)
so it does not affect the comparison between P(F = Montague | features) and P(F = capulet | features)

15.4 (12 points) Use your Naive Bayes Classifier and the probability table provided in 15.3 to classify the following characters:

1. Hermia: **L**=False, **M**=True, **W**=True, **S**=False

   Montague **Capulet** **Can't Tell**

2. Lysander: **L**=True, **M**=True, **W**=False, **S**=True

   Montague **Capulet** **Can't Tell**

3. Demetrius: **L**=True, **M**=False, **W**=True, **S**=True

   Montague **Capulet** **Can't Tell**

Show your work in the box below for partial credit.

General equation: $P(F \mid features) = \frac{P(F)P(L \mid F)P(W \mid F)P(S \mid F)}{P(features)}$

NOTE: we can ignore $P(F)$, $P(W \mid F)$ and $P(features)$

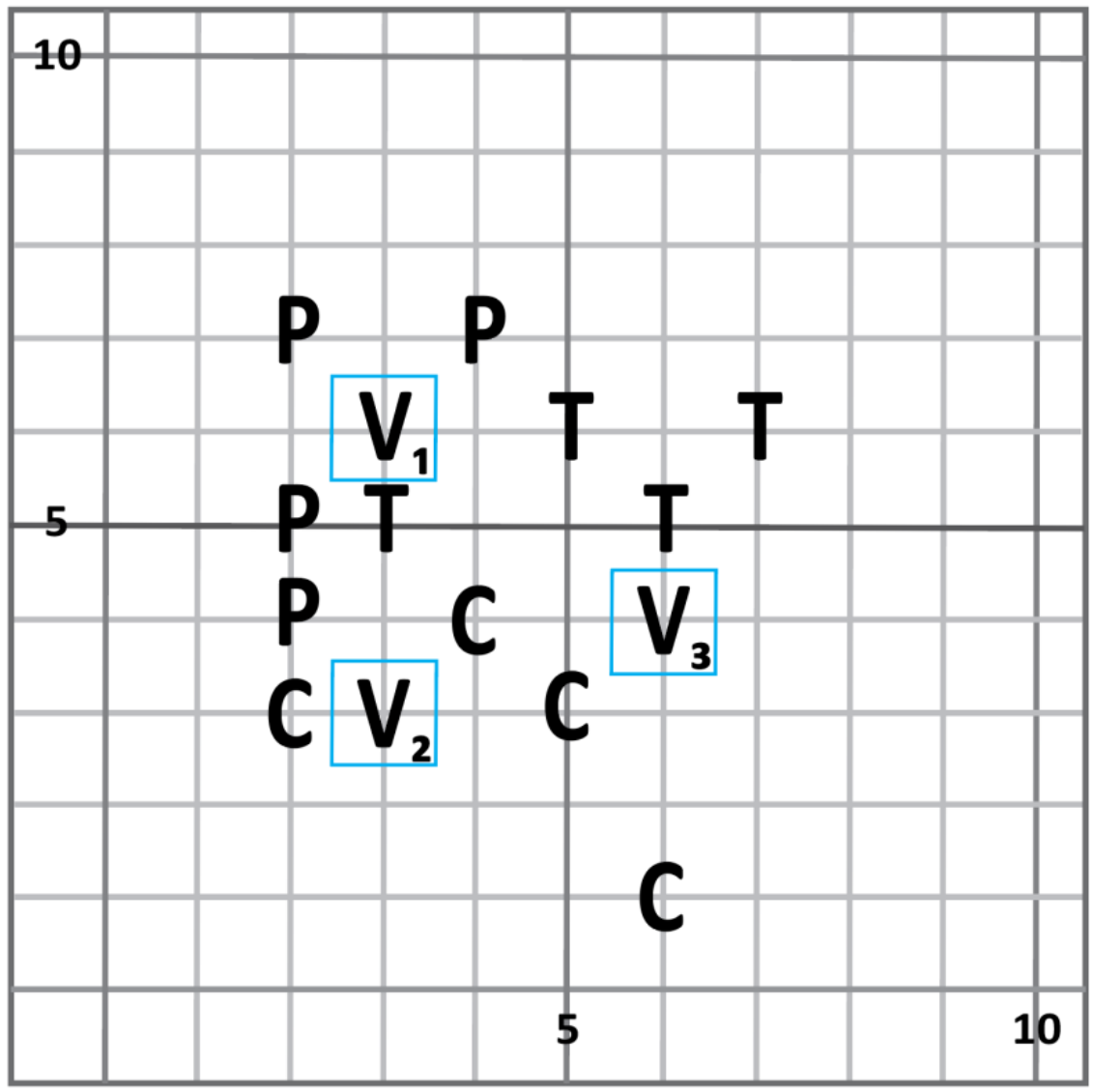
$P(Montague \mid features)$	$P(Capulet \mid features)$
Hermia	
$= \frac{(0.5)(0.1)(0.2)(0.5)(0.4)}{P(features)}$	$= \frac{(0.5)(0.5)(0.8)(0.5)(0.6)}{P(features)}$
Lysander	
$= \frac{(0.5)(0.9)(0.2)(0.5)(0.6)}{P(features)}$	$= \frac{(0.5)(0.5)(0.8)(0.5)(0.4)}{P(features)}$
Demetrius	
$= \frac{(0.5)(0.9)(0.8)(0.5)(0.6)}{P(features)}$	$= \frac{(0.5)(0.5)(0.2)(0.5)(0.1)}{P(features)}$

Note: the point of the calculations is to find out which classification has a higher likelihood. ($P(Montague \mid features)$ or $P(Capulet \mid features)$)

Quiz 3, Problem 1: k Nearest Neighbors (25 points)

Part A: The Vegetable Garden (17 points)

Kyla is enjoying a nice morning walk when she comes across a beautiful garden full of delicious food. Using her acute sense of smell, she determines that there are three types of vegetables planted: Potatoes **(P)**, Turnips **(T)**, and Carrots **(C)**. There are 3 mystery vegetables ($\mathbf{V_1}$, $\mathbf{V_2}$, $\mathbf{V_3}$) that Kyla can't identify. A map of the locations of the vegetables is shown below.



16.1 (9 points) For **k = 1, 3, and 5,** using the **Euclidean distance metric,** determine the classification of each mystery vegetable ($\mathbf{V_1}$, $\mathbf{V_2}$, $\mathbf{V_3}$)

Fill in the table below with the classification **P**, **T**, **C** or **Can't Tell**.

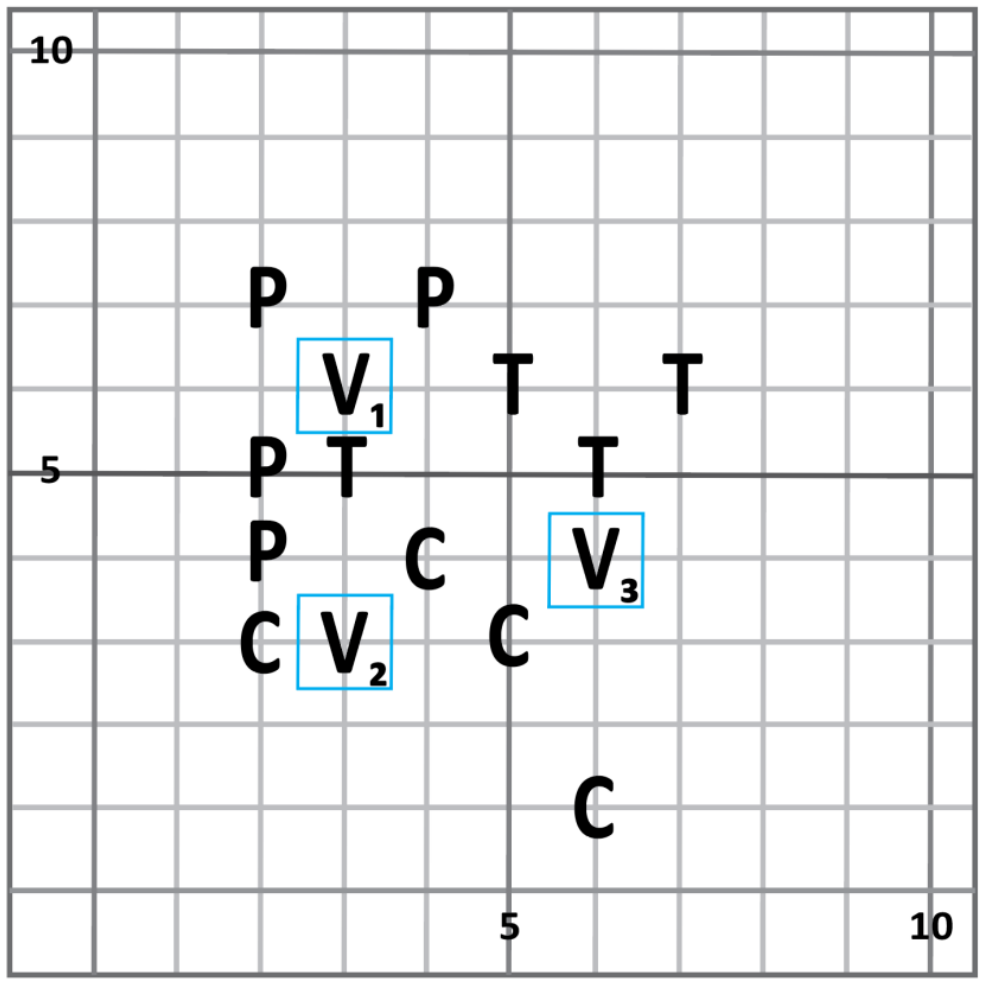
k-value	$\mathbf{V_1}$	$\mathbf{V_2}$	$\mathbf{V_3}$
k = 1	T	C	T
k = 3	P	C	C
k = 5	P	C	T

Dr. Koile tells Kyla the classification of each mystery vegetable.

V_1	V_2	V_3
P	C	T

16.2 (2 points) Given the classifications and the results from **A1,** what is the best k-value? Circle one.

k = 1 **k = 3** **k = 5**



16.3 (4 points) Kyla wants to choose the better distance metric: Euclidean or Manhattan. Using the **best k-value you determined in A2** and the **Manhattan distance metric,** determine the classification of each mystery vegetable (V_1, V_2, V_3)

Fill in the table below with the classification **P**, **T**, **C** or **Can't Tell**.

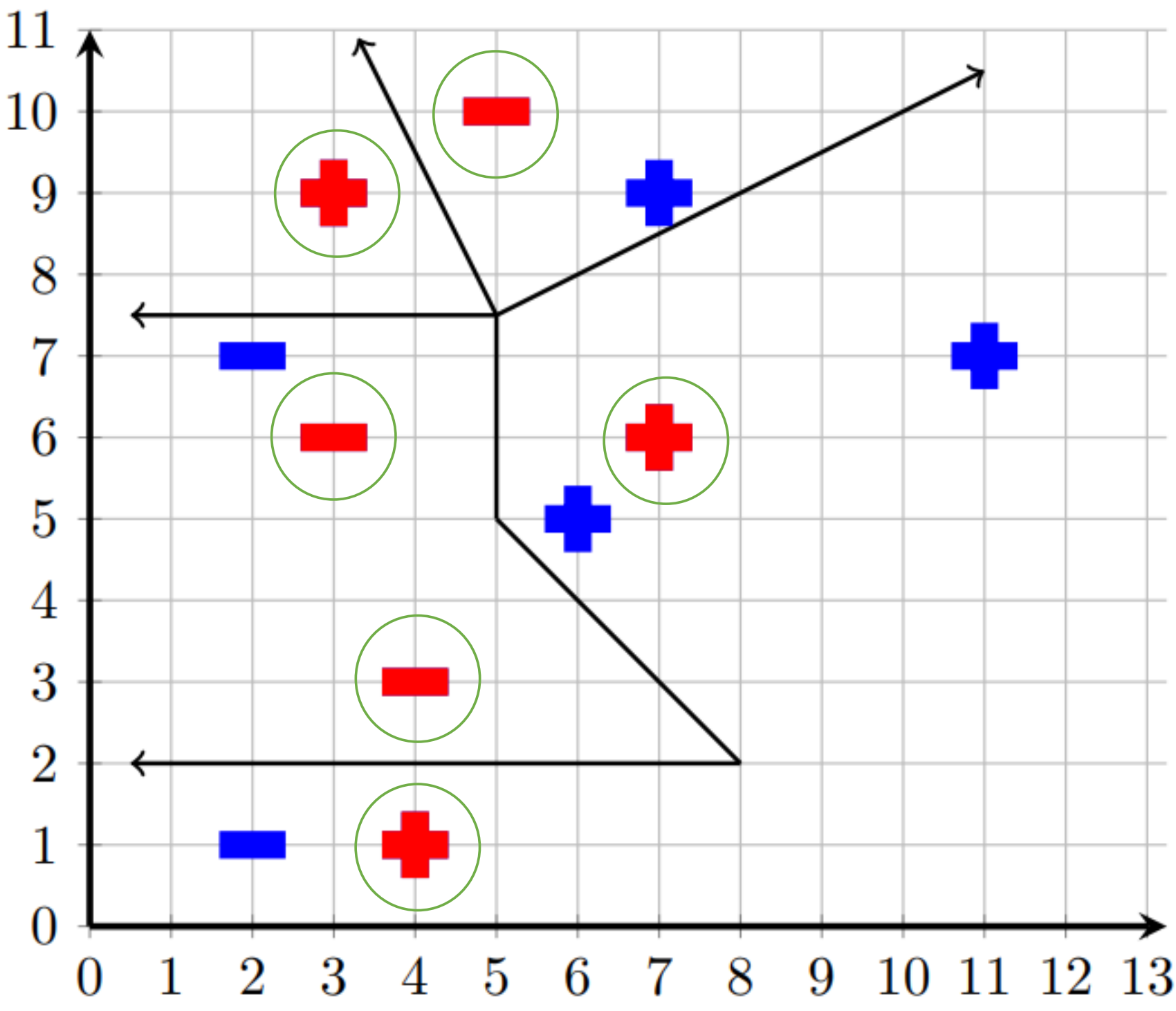
k-value (from A2)	Distance metric	V_1	V_2	V_3
5	Manhattan	P	C	Can't Tell

16.4 (2 points) Based on the results from **A1** and **A3**, which is the better distance metric? Circle one.

Euclidean **Manhattan** **Can't Tell**

Part B: Decision Boundary (6 points)

17. (6 points) The graph below shows a 1-Nearest Neighbors boundary line corresponding to exactly 6 training points that are classified as either + or -. Of the 11 potential training points, which set of 6 would result in this boundary line using a Euclidean distance metric? **Circle these 6 points in the graph below.**



Part C: True/False (2 points)

18. (2 points) For each of the general statements about k-NNs, circle **TRUE** or **FALSE**.

1. Excellent performance on training data but poor performance on test data can be caused by **underfitting**. **TRUE** **FALSE**

2. **Cross validation** can be used to estimate a k-NN model's generalization error. **TRUE** **FALSE**

Problem 2: Identification (ID) Trees (33 points)

Part A: To Piazza or not to Piazza (29 points)

You have been enlisted to recreate an ID tree to help next year's students decide what should be asked on Piazza or in Office Hours. We have identified four general questions that can help distinguish between Piazza posts and Office Hour help:

1. **Content?** Possible values: *Lab, Logistics, Private*.
2. **Bug?** Possible values: *Yes, No, Can't Tell.*
3. **Extension?** Possible values: *Yes, No.*
4. **New Question?** Possible values: *Yes, No, Can't Tell.*

From the 1000 posts on Piazza from the semester, 8 have been chosen to train the ID tree to identify whether a question should be asked on Piazza or in Office Hours. These 8 questions are classified based on **where they should be asked** (Piazza, or Office Hours):

Where to Ask	Content?	Bug?	Extension?	New Question?
OH	Lab	Yes	No	No
OH	Logistics	Can't Tell	Yes	No
OH	Lab	Yes	No	No
Piazza	Lab	Yes	No	Yes
Piazza	Private	No	Yes	No
Piazza	Logistics	No	Yes	Yes
OH	Lab	Can't Tell	No	No
Piazza	Logistics	No	No	Can't Tell

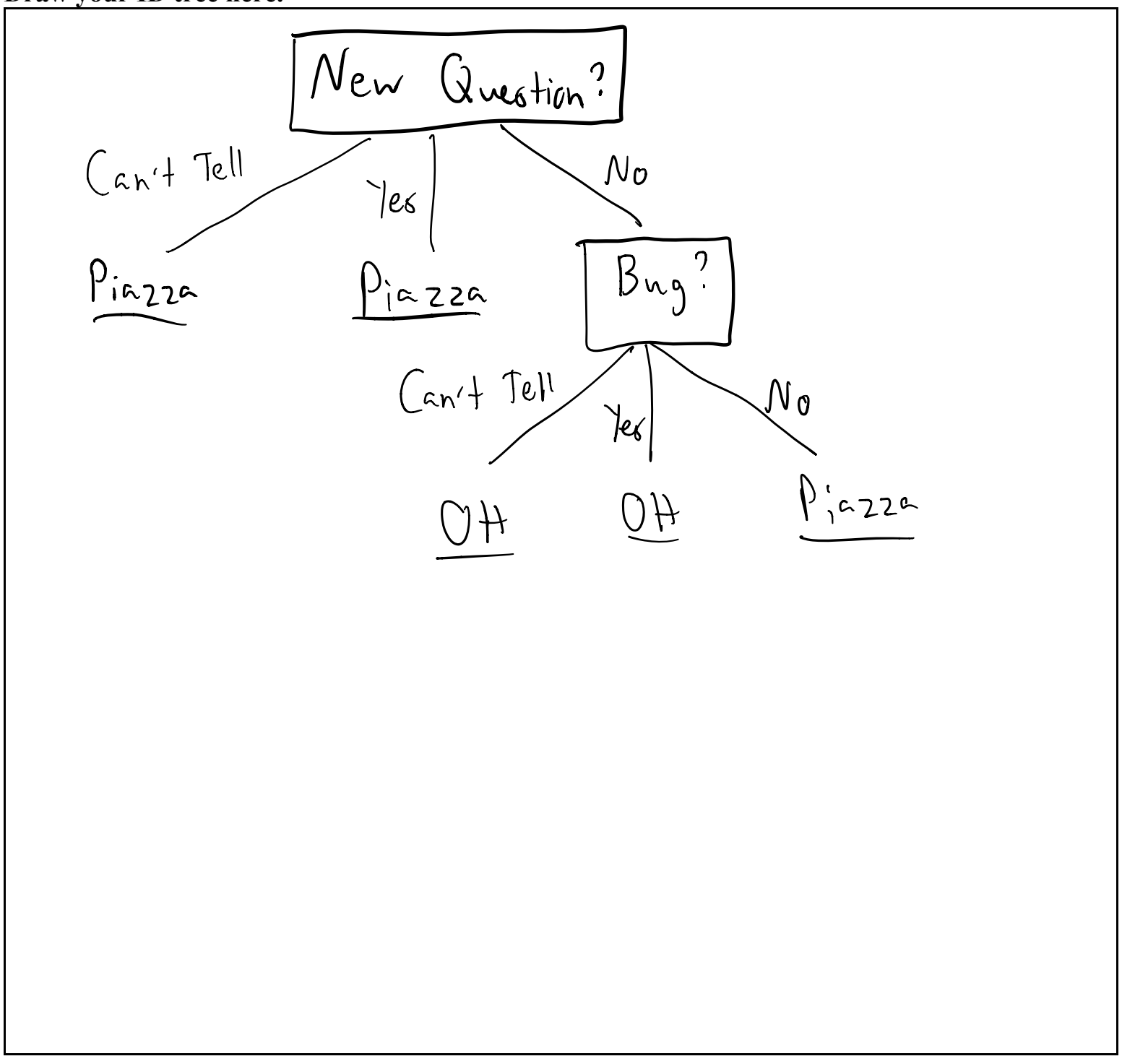
A1 (16 points) Compute the average test disorder for each of the four feature tests. **Please express your answer as a decimal (2 digits after the decimal point)**.

Feature Test	Average Test Disorder
A1.1: Content? (4 points)	Answer: 0.75 Calculations: $-\frac{1}{2}\left(\frac{1}{4}\log_2\frac{1}{4}+\frac{3}{4}\log_2\frac{3}{4}\right)-\frac{3}{8}\left(\frac{1}{3}\log_2\frac{1}{3}\quad\frac{2}{3}\log_2\frac{2}{3}\right)-\frac{1}{8}\left(1\log_2 1\right)$
A1.2: Bug? (4 points)	Answer: 0.69 Calculations: $-\frac{3}{8}\left(\frac{1}{3}\log_2\frac{1}{3}+\frac{2}{3}\log_2\frac{2}{3}\right)-\frac{3}{8}\left(\frac{1}{3}\log_2\frac{1}{3}+\frac{2}{3}\log_2\frac{2}{3}\right)-\frac{1}{4}\left(1\log_2 1\right)$
A1.3: Extension? (4 points)	Answer: 0.95 Calculations: $-\frac{5}{8}\left(\frac{3}{5}\log_2\frac{3}{5}+\frac{2}{5}\log_2\frac{2}{5}\right)-\frac{3}{8}\left(\frac{1}{3}\log_2\frac{1}{3}+\frac{2}{3}\log_2\frac{2}{3}\right)$
A1.4: New Question? (4 points)	Answer: 0.45 Calculations: $-\frac{5}{8}\left(\frac{4}{5}\log_2\frac{4}{5}+\frac{1}{5}\log_2\frac{1}{5}\right)-\frac{1}{4}\left(1\log_2 1\right)-\frac{1}{8}\left(1\log_2 1\right)$

A2 (12 points) Construct a **greedy, disorder-minimizing ID tree** to correctly classify whether a question should be asked on **Piazza** or during **Office Hours** using the four feature tests: **Content?**, **Bug?**, **Extension?**, **New Question?**.

If there are any ties between classifiers, choose the one that comes first in alphabetical order. A full credit tree must include: clearly labeled feature test nodes, branches, and leaves. There is space on the next page to show work for possible partial credit.

Draw your ID tree here.



B1 Partial Credit

Show your work and/or disorder calculations here for partial credit:

A3 (1 point) Your friend has a question about the 6.034 final exam and is not sure where to ask the question. With your ID tree from A2, where would you suggest that your friend ask their question?

Content?	Bug?	Extension?	New Question?
Logistics	No	No	Can't Tell

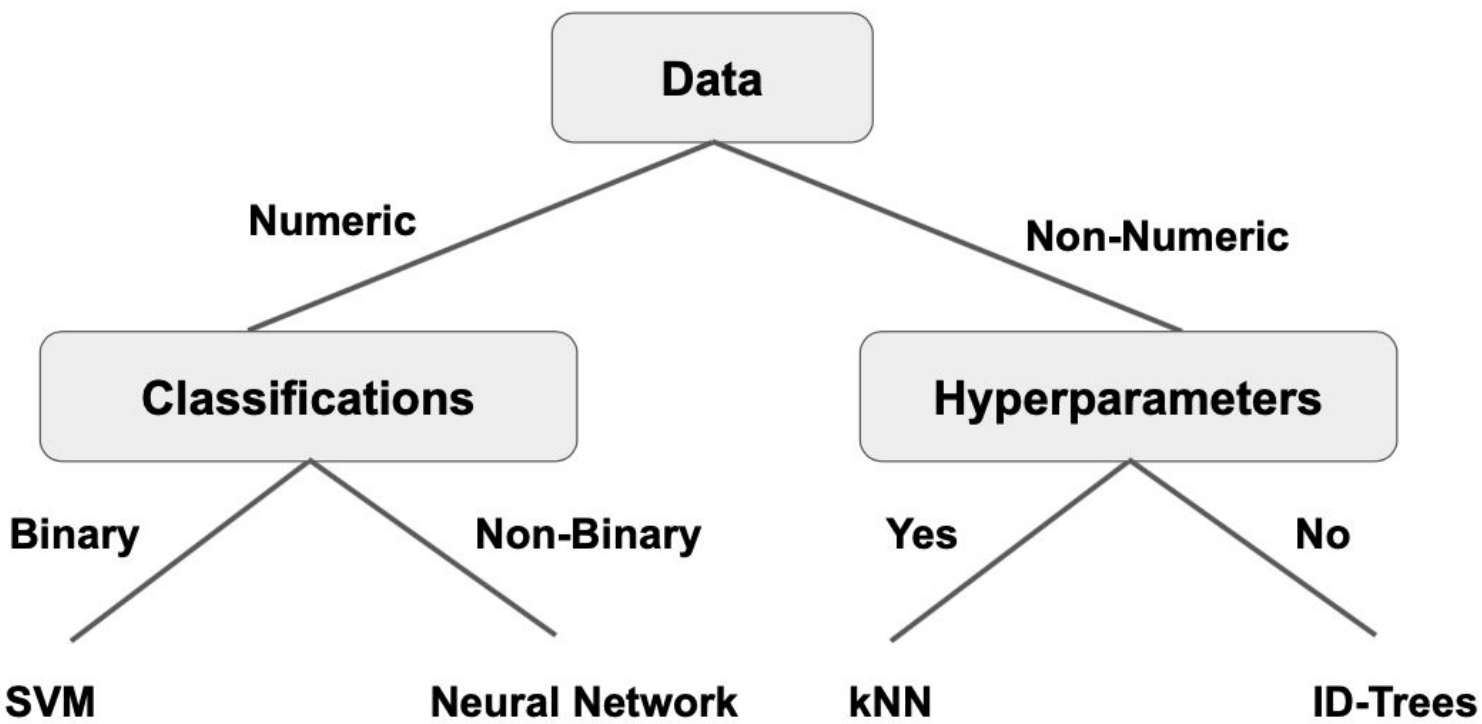
Circle the classification for your friend's question:

Where to Ask = **Piazza** **OH** **Can't Tell**

Part B: ID Trees: When to use them? (4 points)

B1 (4 points) After a semester of 6.034, the course staff hope that you have learned a lot about AI and its limitless capabilities. With all these new techniques up your sleeve, the TAs decided to create an ID tree to help you figure out which one you should use for future problems you face.

<u>Write the simplest if-then rule that will help you determine when to use ID Trees.</u> (Feel free to abbreviate words.)



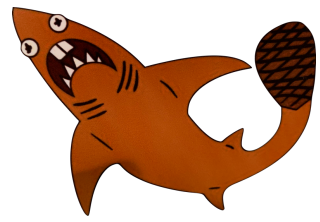
Rule (fill in the blank):

IF (The data is non-numeric and there are no hyperparameters

)

THEN ID-Trees

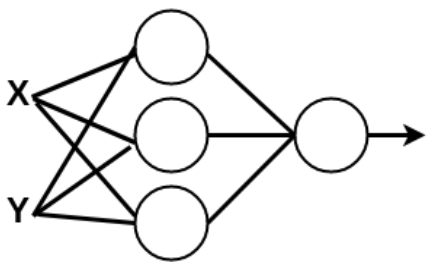
Quiz 3, Problem 3: Neural Nets (40 points)



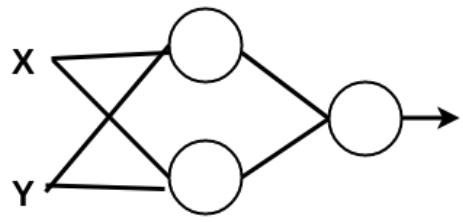
At Beaver Tank, 6.034's very own version of Shark Tank, investors use neural nets to help them decide whether to invest in students' inventions.

Part A: Meet the investors and their Nets (12 points)

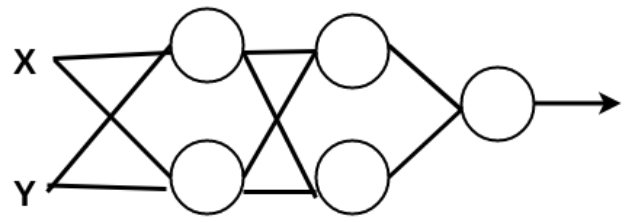
Juan



Kimberle



Theo



21. (12 points) For each of the following diagrams (a-d), **<u>circle all possible investors</u>** whose nets can produce that decision boundary. Assume that all neurons in all networks use the **stairstep activation function** and that **X and Y can be any real number.**

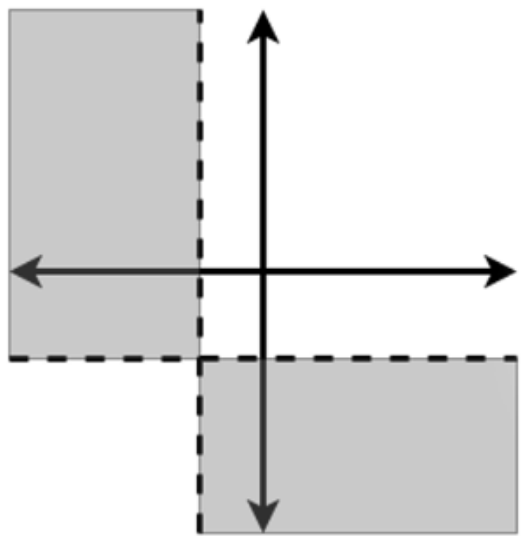
a.

Juan

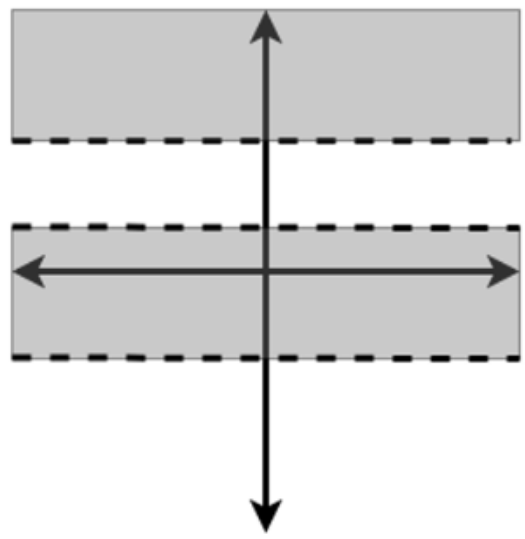
Kimberle

Theo

None



b.



Juan

Kimberle

Theo

None

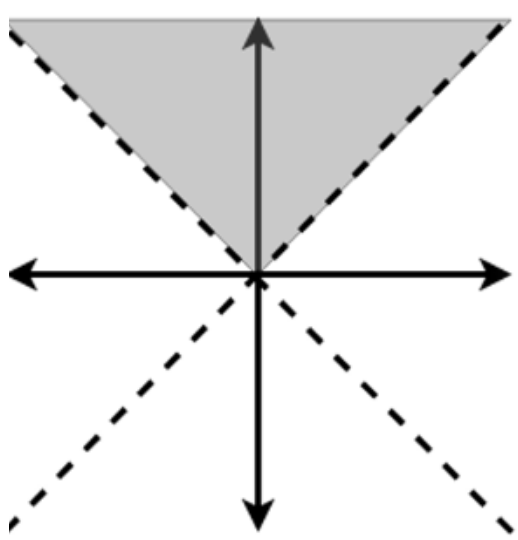
c.

Juan

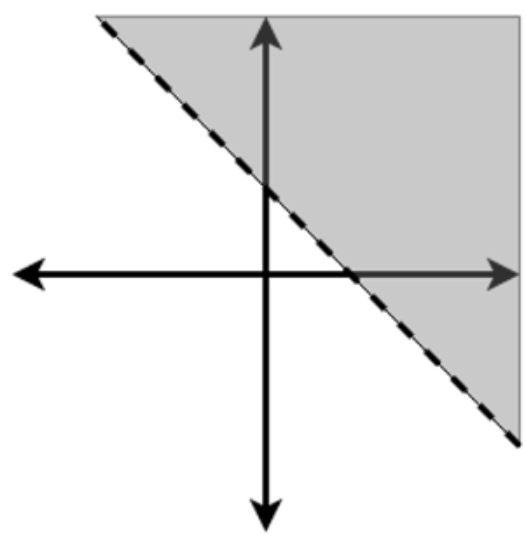
Kimberle

Theo

None



d.



Juan

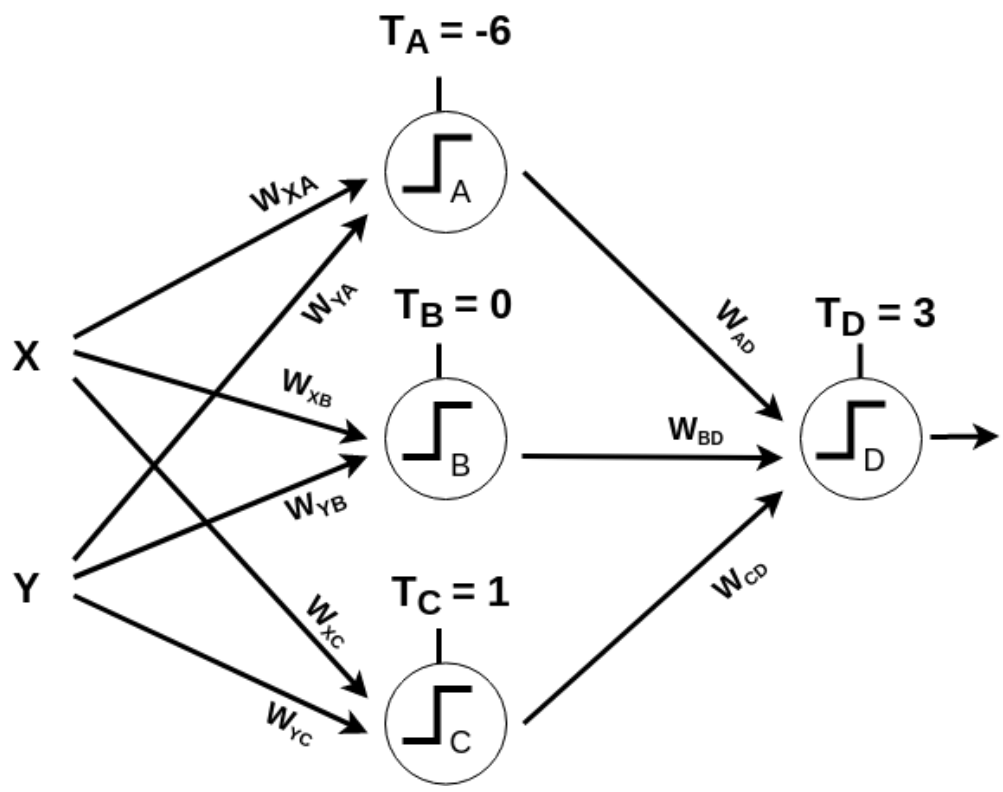
Kimberle

Theo

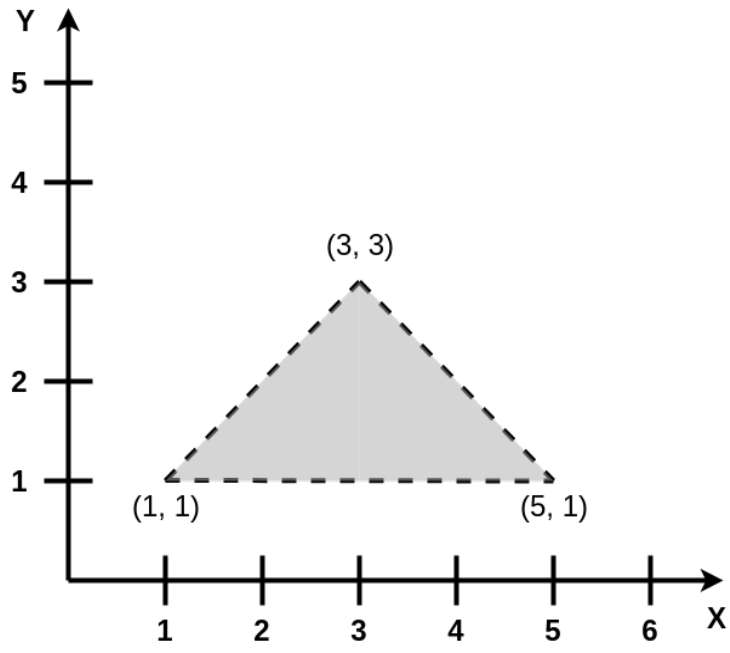
None

Part B: Zooming in on Juan's Net (14 points)

Let's take a closer look at investor Juan's net, shown below. All neurons use the **stairstep** activation function.



22. (14 points) Determine what the weights for each of the edges could be to produce the graph below. You can show your work on the next page for partial credit. *Note*: *There may be more than one possible set of answers. Any correct set of weights will get full credit.*



Note: Weights must be INTEGER values.

$W_{XA} =$ **-1**	$W_{XB} =$ **a, a> 0**	$W_{XC} =$ **0**
$W_{YA} =$ **-1**	$W_{YB} =$ **a, a> 0**	$W_{YC} =$ **1**
$W_{AD} =$ **1**	$W_{BD} =$ **1**	$W_{CD} =$ **1**

For partial credit, you can show your work in the box below.

line 1: $y = x$

$-x + y \geq 0$ → neuron B

$W_{xB} = -a$

$W_{yB} = a$

line 2: $y = -x + 6$

$-x - y \geq -6$ → neuron A

$W_{xA} = -1$

$W_{yA} = -1$

line 3: $y = 1$

$y \geq 1$ → neuron C

$W_{xC} = 0$

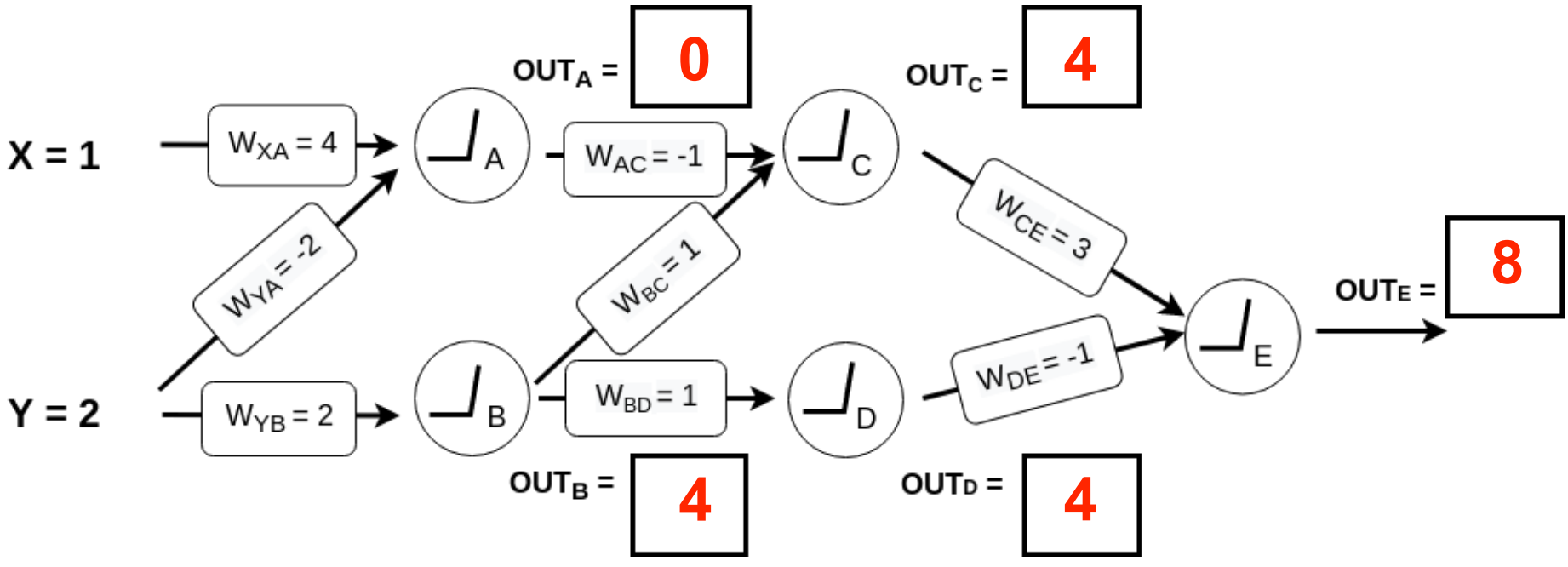
$W_{yC} = 1$

"AND" Together → neuron D which has $T_D = 3$

$W_{AD} = W_{BD} = W_{CD} = 1$

Part C: Theo's Upgrade (5 points)

Theo wants his neural network to predict how much he should invest, not just whether he should invest. The unit of the output of the network is in dollars. He modifies his network and uses the **ReLU activation function (ReLU(x) = max(0, x))** for all his neurons. Theo's new net is shown below.



23.1 (4 points) Given an input with features **(X = 1, and Y=2),** determine what the output of each neuron should be. **Fill in the corresponding blanks in the diagram shown above.**

For partial credit, you can show your work in the box below.

$OUT_A = ReLU(4 \cdot 1 + (-2) \cdot 2) = 0$

$OUT_B = ReLU(2 \cdot 2) = 4$

$OUT_C = ReLU(-1 \cdot 0 + 4 \cdot 1) = 4$

$OUT_D = ReLU(1 \cdot 4) = 4$

$OUT_E = ReLU(3 \cdot 4 + (-1) 4) = 8$

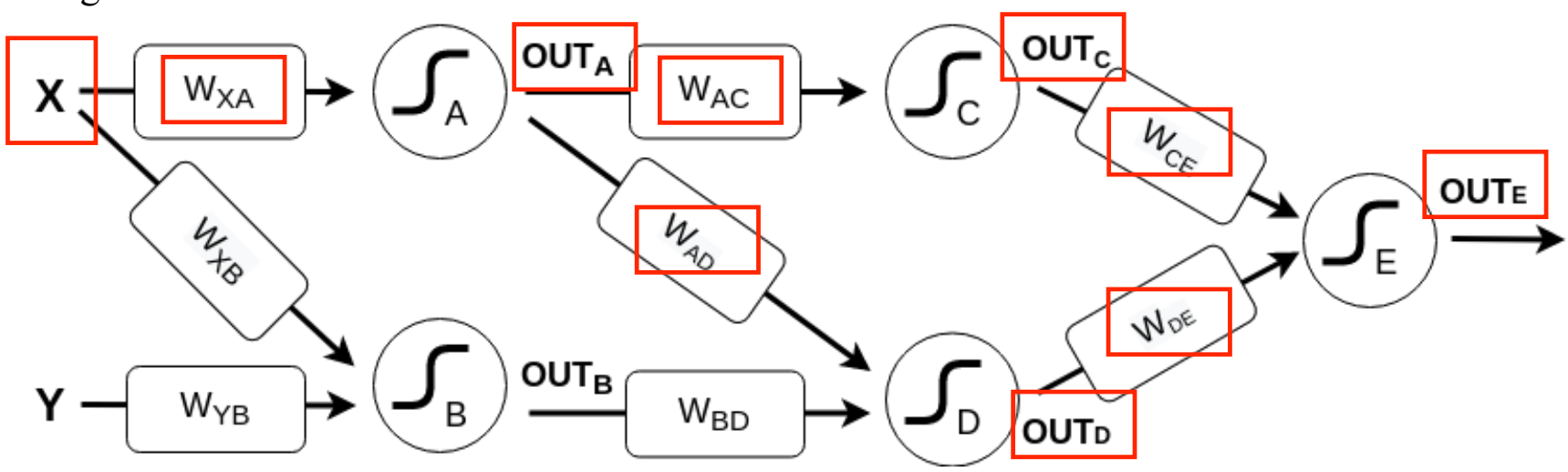
23.2 (1 point) According to his neural net, how much should Peter invest?

$8

Part D: It's a loss, let's backprop (9 points)

Investor Kyla decides to join. Kyla lost a lot of money from her first few investments, so she decides to tune her network by performing backpropagation. Kyla's network is shown below. Note that she uses the **sigmoid activation function for all neurons.**

24.1 (3 points) Circle all neuron outputs, weights, and inputs necessary to update weight W_{XA} on the diagram below.



24.2 (4 points) Write the update equation for W_{XA} in terms of the learning rate r, inputs, weights, and neuron outputs.

$$W_{XA,new} = W_{XA} + r \cdot X \cdot \delta_A$$

$$\delta_A = OUT_A(1-OUT_A)(W_{AC}\delta_C + W_{AD}\delta_D)$$

$$\delta_C = OUT_C(1-OUT_C)(W_{CE}\delta_E)$$

$$\delta_D = OUT_D(1-OUT_D)(W_{DE}\delta_E)$$

$$\delta_E = (OUT^* - OUT_E)OUT_E(1-OUT_E)$$

$$W_{XA,new} = W_{XA} + r \cdot X \cdot OUT_A(1-OUT_A) \cdot \Big[W_{AC}OUT_C(1-OUT_C) \cdot W_{CE}(OUT^*-OUT_E)OUT_E(1-OUT_E) + W_{AD}OUT_D(1-OUT_D) \cdot W_{DE}(OUT^*-OUT_E)OUT_E(1-OUT_E) \Big]$$

24.3 (1 point) Is it possible to run backpropagation if the activation function is changed to **ReLU**?
Circle one.

Yes **No**

24.4 (1 point) Is the network's prediction guaranteed to be better after one round of backpropagation?
Circle one.

Yes **No**

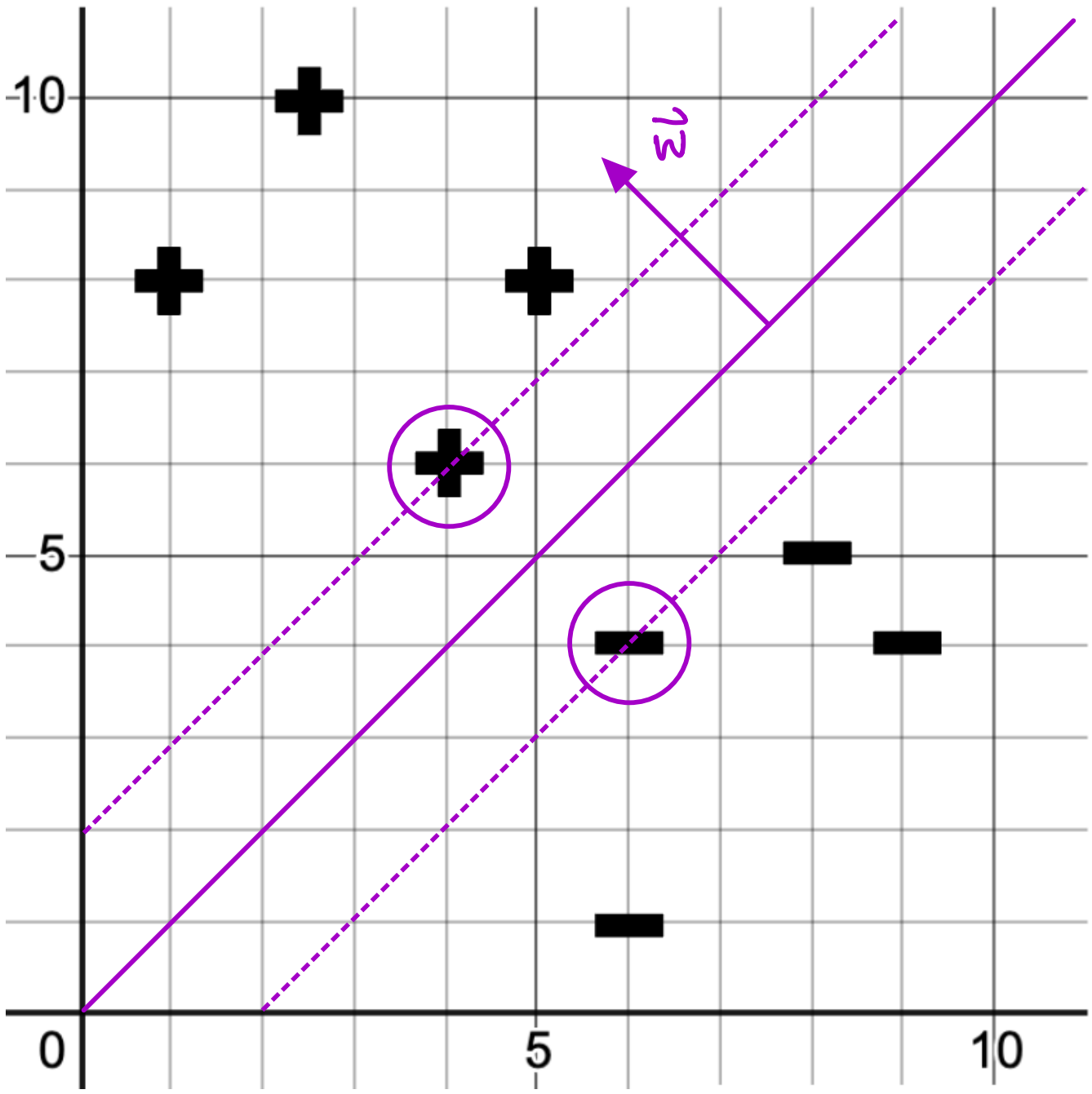
Quiz 4, Problem 1: SVMs (50 points)

Part A: Top Fan Status (16 points)

During finals week, Tim the Beaver is overwhelmed by the number of posts submitted to MIT Confessions. In exchange for top fan status, he asks for your help in building a classifier to identify quality confessions based on data points he has collected.

25.1 (6 points) On the graph below,

- Draw the **decision boundary** with a solid line.
- Draw the **gutters** with a dashed line.
- Circle the **support vectors**.
- Denote the *direction* of vector $\vec{w}$ with an arrow.



25.2 (10 points) Using your decision boundary in part A1, compute the values of the normal vector $\vec{w}$ and offset b.

$\vec{w} =$ $\begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$ $\quad b =$ 0

Show your work in the box below for partial credit.

$y = x$

$x - y = 0$

$1(x) - 1(y) = 0$

$\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$

$m = 2 - \sqrt{2}$

$\|\vec{w}\| = \frac{1}{\sqrt{2}}$

$c\left(\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix}\right) = 1$

$c(-2) = 1$

$c = -\frac{1}{2}$

$-\frac{1}{2}\left(\begin{bmatrix} 1 & -1 \end{bmatrix}\right)$

$\therefore \vec{w} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad b = 0$

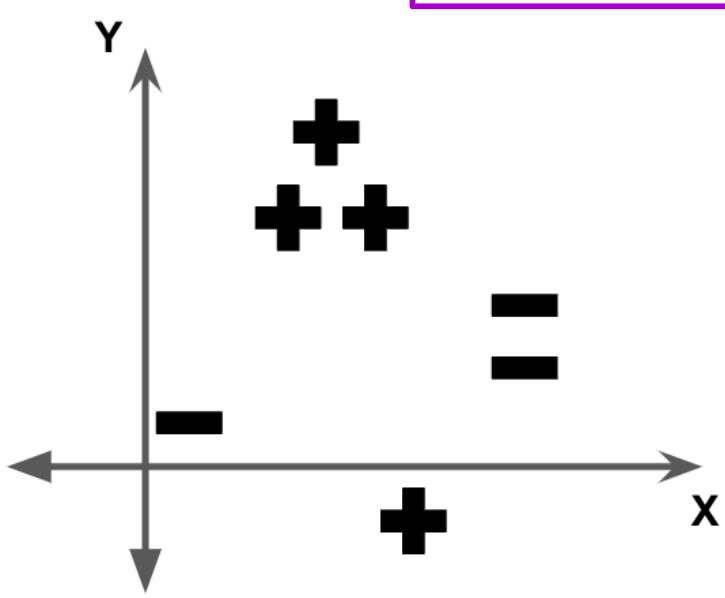
Part B: Iterations, iterations (11 points)

Tim the Beaver starts using the SVM to decide which confessions to post. He isn't satisfied with the results, so he collects more data. Help him pick a new kernel.

26.1 (8 points) For each kernel (a-d), **circle** if the points are SEPARABLE or are NOT SEPARABLE. If the data is SEPARABLE, also **draw a decision boundary that would classify the training data correctly using that kernel.**

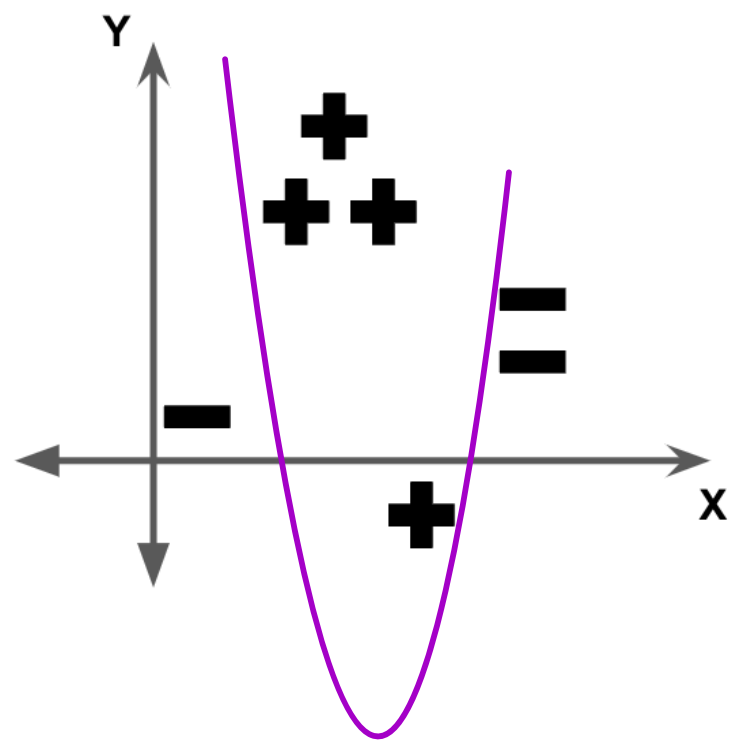
a. With a **linear kernel**, the data is:

SEPARABLE | NOT SEPARABLE



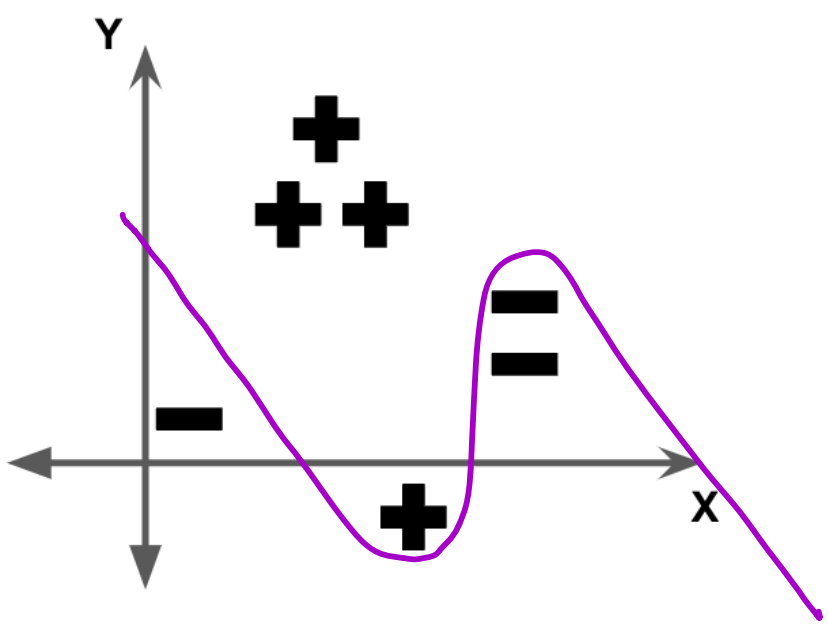
b. With a **quadratic kernel,** the data is:

SEPARABLE | NOT SEPARABLE



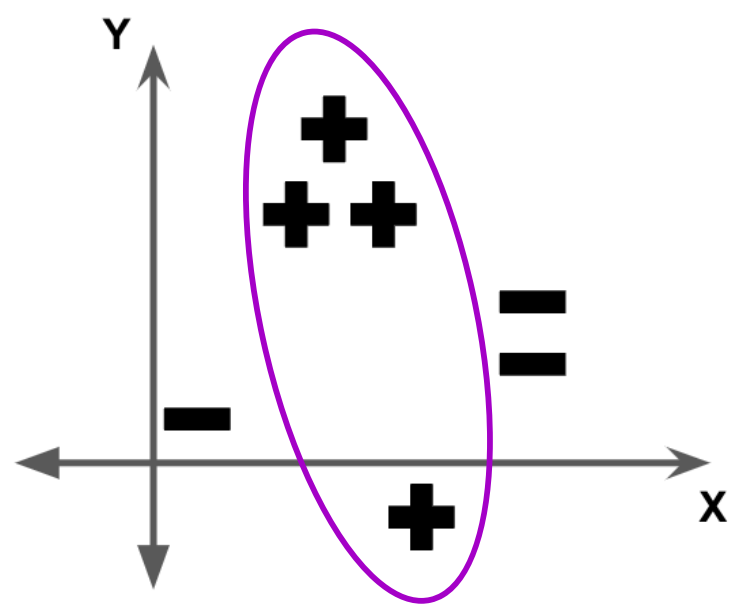
c. With a **cubic kernel,** the data is:

SEPARABLE | NOT SEPARABLE



d. With a **radial basis kernel,** the data is:

SEPARABLE | NOT SEPARABLE



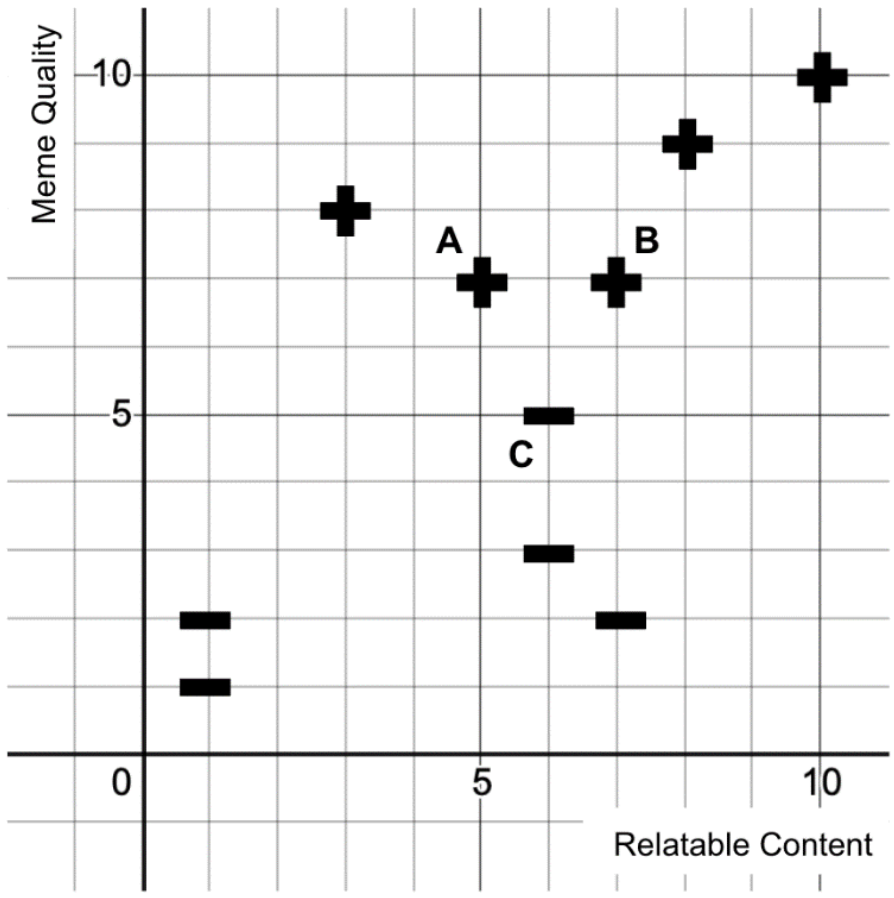
* Several other variations exist for drawn functions

26.2 (3 points) Tim the Beaver is considering using higher-dimensional features for the data. Which of the following describes the decision boundary for a 4-dimensional linear SVM?

a. Line
b. Hyperplane
c. Point
d. Half-plane

Part C: Changing Labels (10 points)

A different training set is shown below.

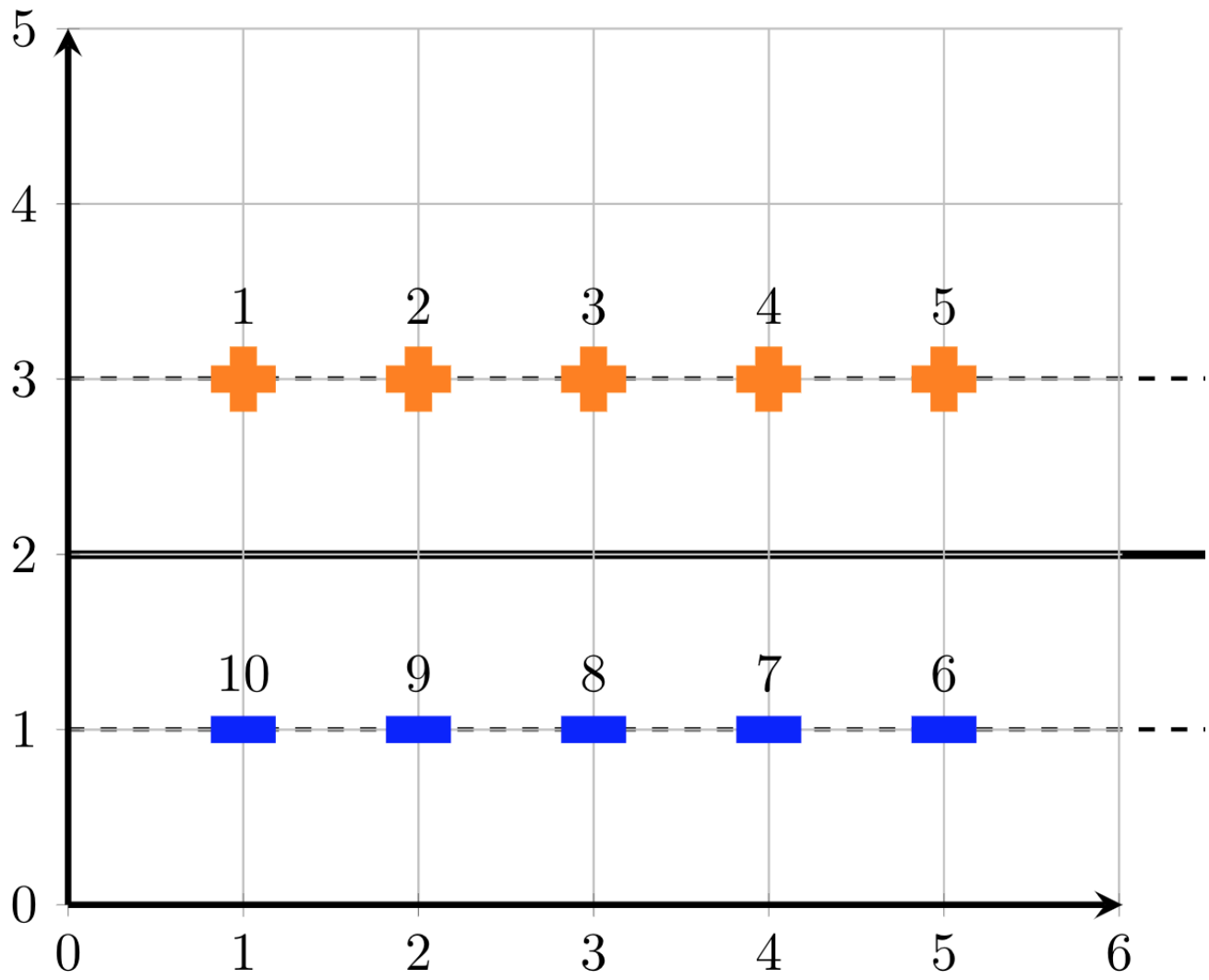


27. (10 points) How would labeling **point C** as a **positive** example, instead of a negative example as shown, change the following values? For each value, circle one answer.

Margin width	Increase	Decrease	Doesn't Change	Can't Tell
Training error	Increase	Decrease	Doesn't Change	Can't Tell
Number of support vectors	Increase	Decrease	Doesn't Change	Can't Tell
α_A	Increase	Decrease	Doesn't Change	Can't Tell
α_C	Increase	Decrease	Doesn't Change	Can't Tell

Part D: Setting up SVMs (13 points)

Finally, Tim the Beaver wants to guarantee that he can trust your judgment to ensure quality confessions, so he gives you the following problem to solve.



Positive and negative points are uniformly arranged with the decision boundary (solid line at y=2) and gutters (dashed lines at y=1, y=3) drawn on the graph above.

28.1 (7 points) Circle all sets of points that could **potentially be a valid set of support vectors** for the given decision boundary.

a. [3, 8]

b. [5, 6]

c. [2, 7]

d. [3, 7, 9]

e. [4, 6, 9]

f. [2, 9, 7]

g. [5, 7, 10]

28.2 (6 points) Now, determine if/when the decision boundary shown above changes to a different line as Tim systematically removes points from the graph. Circle each set of points **whose removal would** alter the decision boundary. If removing a set of points would cause the decision boundary to change, write the equation of the line describing the new decision boundary in the space provided next to the circled set of points.

The decision boundary changes when we remove:	The boundary equation becomes:
a. [1, 6]	_____
b. [1, 2, 7, 6]	_____
c. [1, 2, 3, 6, 7, 8]	$y = -x + 5$
d. [1, 2, 3, 4, 6, 7, 8, 9]	$y = -2x + 8$
e. None of the Above	

Quiz 4, Problem 1: Boosting through Boston (50 points)

Part A: Boston Sports (25 points)

You decide to use Adaboost to convince your skeptical New York friends that Boston sports teams are superior to New York teams. You ask 6 friends which team they think is better in baseball (B), basketball (K), football (F), hockey (H), and soccer (S). Treating each friend as a **weak classifier (h1-h6)**, the table below shows their classifications. The correct classification for best team is shown in the last row.

	Baseball (B)	**Basketball (K)**	**Football (F)**	**Hockey (H)**	**Soccer (S)**
h1	Yankees	Knicks	Giants	Bruins	Red Bulls
h2	Yankees	Knicks	Giants	Rangers	Revolution
h3	Yankees	Celtics	Patriots	Rangers	Red Bulls
h4	Yankees	Celtics	Patriots	Bruins	Revolution
h5	Yankees	Celtics	Patriots	Bruins	Red Bulls
h6	Red Sox	Celtics	Giants	Bruins	Red Bulls
Correct Classification	**Red Sox**	**Celtics**	**Patriots**	**Bruins**	**Revolution**

29.1 (3 points) Fill in the following table to show each classifiers' misclassifications. Assume that a training point is positive (+) if it matches the correct classification and negative (-) otherwise. You may abbreviate the training points according to sports as: B (baseball), K (basketball), F (football), H (hockey), S (soccer).

Weak Classifier	**Points Misclassified**
h_1	B, K, F, S
h_2	B, K, F, H
h_3	B, H, S
h_4	B
h_5	B, S
h_6	F, S

29.2 (12 points) Fill out the table below to complete the first 3 rounds of Adaboost, choosing the classifier with the error rate **farthest from ½**. Break ties according to which classifier comes first in this list: h1, h2, h3, h4, h5.

Blue text is the alternate set of solutions

	Round 1	**Round 2**	**Round 3**
Weight of B (w_B)	1/5	1/8 4/8	1/14 4/12
Weight of K (w_K)	1/5	1/8 1/8	1/14 1/12
Weight of F (w_F)	1/5	1/8 1/8	1/14 3/12
Weight of H (w_H)	1/5	1/2 1/8	4/14 1/12
Weight of S (w_S)	1/5	1/8 1/8	7/14 3/12
Error rate of h1 (ε_{h1})	4/5	4/8 7/8	10/14 11/12
Error rate of h2 (ε_{h2})	4/5	7/8 7/8	7/14 9/12
Error rate of h3 (ε_{h3})	3/5	6/8 6/8	12/14 8/12
Error rate of h4 (ε_{h4})	1/5	1/8 4/8	1/14 4/12
Error rate of h5 (ε_{h5})	2/5	2/8 5/8	8/14 7/12
Error rate of h6 (ε_{h6})	2/5	2/8 2/8	8/14 6/12
Chosen classifier	h1 h4	h2 h6	h4 h4
Error rate of classifier (ε)	4/5 1/5	7/8 2/8	1/14 1/3
Voting power (α)	½ ln(1/4) ½ ln(4)	½ ln(1/7) ½ ln(3)	½ ln(13) ½ ln(2)

For partial credit, show your work in the box below.

29.3 (4 points) What ensemble classifier **H(x)** would you generate after 3 full rounds of Adaboost?
Hint: Don't forget the SIGN(...).

sign(½ ln(1/4) * h1 + ½ ln(1/7) * h2 + ½ ln(13) * h4)

29.4 (1 point) Is the ensemble classifier H(x) from A3 a "perfect" classifier i.e. with no misclassifications of any training points?

YES **NO** **CAN'T TELL**

29.5 (5 points) Another friend is not convinced that choosing the classifier with the smallest error farthest from ½ will generate a good classifier. After the end of 3 rounds, what will the voting powers look like if you instead choose a weak classifier at each round with the **smallest error less than ½**? *Hint: Some of the alphas can be 0.*

$$H(x) = \text{SIGN}(\alpha_1 h_1(x) + \alpha_2 h_2(x) + \alpha_3 h_3(x) + \alpha_4 h_4(x) + \alpha_5 h_5(x) + \alpha_6 h_6(x))$$

$\alpha_1 = 0$	$\alpha_2 = 0$	$\alpha_3 = 0$	$\alpha_4 = ½$ (ln4 + ln2)	$\alpha_5 = 0$	$\alpha_6 = 01$

For partial credit, show your work in the boxes below.

Part B: Vermonsters (10 points)

In search of more supportive friends, you turn to your local New England friends and decide to take them to a baseball game. You know the weights of their sports team choices (shown in the table below), but not the error rates or the original training points.

	Round 1	**Round 2**	**Round 3**
Weight of B (w_B)	1/5	1/6	1/5
Weight of K (w_K)	1/5	1/6	1/7
Weight of F (w_F)	1/5	1/4	3/10
Weight of H (w_H)	1/5	1/6	1/7
Weight of S (w_S)	1/5	1/4	3/14
Chosen Classifier	Vermont	Maine	NH

30.1 (5 points) Assume that the classifiers chosen at each round were the **smallest error rate less than ½**. Circle all the points below that were misclassified by Vermont in Round 1. *Hint: This problem doesn't require extensive calculations. Remember what we know about the weights in each round.*

B **K** **F** **H** **S**

Show your work in the box below.

30.2 (5 points) Assume that the classifiers chosen at each round were the **smallest error rate less than ½**. Circle all the points below that were misclassified by Maine in Round 2.

B **K** **F** **H** **S**

Show your work in the box below.

Part C: Football Weight (5 points)

31. (5 points) Bill Belichick claims that the weights in the table below are possible weights for the w_F (weight of Football). Assume that Adaboost has been running for many rounds, picking the weak classifier with the **smallest error rate less than ½**. For each w_F value below, **circle whether it's POSSIBLE or NOT POSSIBLE and explain why.**

w_F value	Possible?	Explanation
$-\frac{1}{3}$	NOT POSSIBLE	Negative weight is not possible
$\frac{1}{200}$	POSSIBLE	
$\frac{5}{7}$	NOT POSSIBLE	> ½ weight is not possible
$\frac{11}{10}$	NOT POSSIBLE	> 1 weight is not possible
0	NOT POSSIBLE	Zero weight is not possible

Part D: Sweet Caroline (10 points)

Fenway Park rings often with the tune of Sweet Caroline, but good times never feel "so good" until you understand boosting. Circle **True** or **False** for each statement below, and provide a one sentence explanation or provide a counterexample.

32.1 (2 points) If we have 3 weak classifiers in H(x) that satisfy the triangle inequality and the weak classifiers have non-overlapping misclassified points, H(x) is a perfect classifier.

TRUE **FALSE**

Explain briefly in one sentence, or provide a counterexample.

32.2 (2 points) Compared to using a single weak classifier, boosting typically reduces overfitting.

TRUE **FALSE**

Explain briefly in one sentence, or provide a counterexample.

32.3 (2 points) If *n* is the total number of weak classifiers available, you **cannot** run Adaboost more than *n* times.

TRUE **FALSE**

Explain briefly in one sentence, or provide a counterexample.

32.4 (2 points) One training point can have the same weight in the first and second rounds of Adaboost.

TRUE **FALSE**

Explain briefly in one sentence, or provide a counterexample.

32.5 (2 points) A perfect classifier will never have a weak classifier h that has a negative voting power α.

TRUE **FALSE**

Explain briefly in one sentence, or provide a counterexample.