6.034 Final Examination
19 December 2012

Name	MYCROFT HOLMES
email	mycroft@diogenes.mit.edu

Circle your TA so that we can more easily identify with whom you have studied.

Dylan Holmes	**Sarah Lehmann**	**Igor Malioutov**	**Robert McIntyre**
Ami Patel	**Mark Seifter**	**Sila Sayan**	**Stephen Serene**

Question 1. Indicate the approximate percent of the 23 lectures, 7 Right-now talks, 5 mega recitations, and 13 tutorials which you have attended so that we can better gauge the correlation of quiz and final performance with attendance. Your answers have no effect on your grade.

	Lectures	Right-now talks	Mega Recitations	Tutorials
Percent attended				

Question 2. Indicate how often you took notes in the final weeks of the semester during the lectures/talks you attended.

Regular lectures:	Never	Rarely	About half the time	Often	Always
Right-now talks:	Never	Rarely	About half the time	Often	Always

Question 3. Indicate what you think of the Right Now talks.

Good idea as is. Bad idea. Would be a good idea if...

Comments:

Quiz	Score	Grader	SRN	Score	Grader
Q1			SRN 2		
Q2			SRN 3		
Q3			SRN 4		
Q4			USRN		

There are 40 pages in this final examination, including this one, but not including blank pages and tear-off sheets. Tear-off sheets with duplicate drawings and data are located after the final page of the exam. As always, open book, open notes, open just about everything, including a calculator, but no computers.

Quiz 1, Problem 1: Rule-Based Systems (50 pts)

After discovering the magic of friendship, Robert and Dylan decide to apply their knowledge of rule-based systems to study something more complex: romantic relationships. After much argument, they decide on the following rules, which they will test on their fellow TAs.

Rules:

P0

```
IF( AND( '(?a) threw a penny into a well',
         '(?b) is in (?a)'s way')
    THEN( '(?a) notices (?b)' ) )
```

P1

```
IF( AND( '(?a) is popular',
         '(?b) is crazy')
    THEN( '(?b) stares at (?a)' ) )
```

P2

```
IF( AND( '(?a) just met (?b)',
         OR( '(?a) is crazy',
             '(?a) is attracted to (?b)' ) ),
    THEN( '(?a) gives his/her number to (?b)' ) )
```

P3

```
IF( OR( AND( '(?a) is wearing ripped jeans',
             '(?b) notices (?a)' ),
        '(?b) stares at (?a)' ),
    THEN( '(?b) is attracted to (?a)' ) )
```

P4

```
IF( AND( '(?b) is attracted to (?a)',
         '(?a) gives his/her number to (?b)')
    THEN( '(?b) will call (?a)' ) )
```

Assertions:

A0: Sila threw a penny into a well
A1: Igor is in Sila's way
A2: Igor is crazy
A3: Sila is popular
A4: Sila is crazy
A5: Igor is wearing ripped jeans
A6: Sila just met Igor

(For your convenience, this table of rules is also provided on a tear-off sheet located at the end of the final examination.)

Part A: Backward Chaining (25 pts)

Make the following assumptions about backward chaining:

- When working on a hypothesis, the backward chainer tries to find a matching assertion in the list of assertions. If no matching assertion is found, the backward chainer tries to find a rule with a matching consequent. If no matching consequent is found, then the backward chainer assumes the hypothesis is false.
- The backward chainer never alters the list of assertions, so it can derive the same result multiple times.
- Rules are tried in the order they appear.
- Antecedents are tried in the order they appear.
- Lazy evaluation/short circuiting is in effect (e.g., if the first part of an AND clause is false, the rest does not need to be evaluated to determine that the whole clause is false).

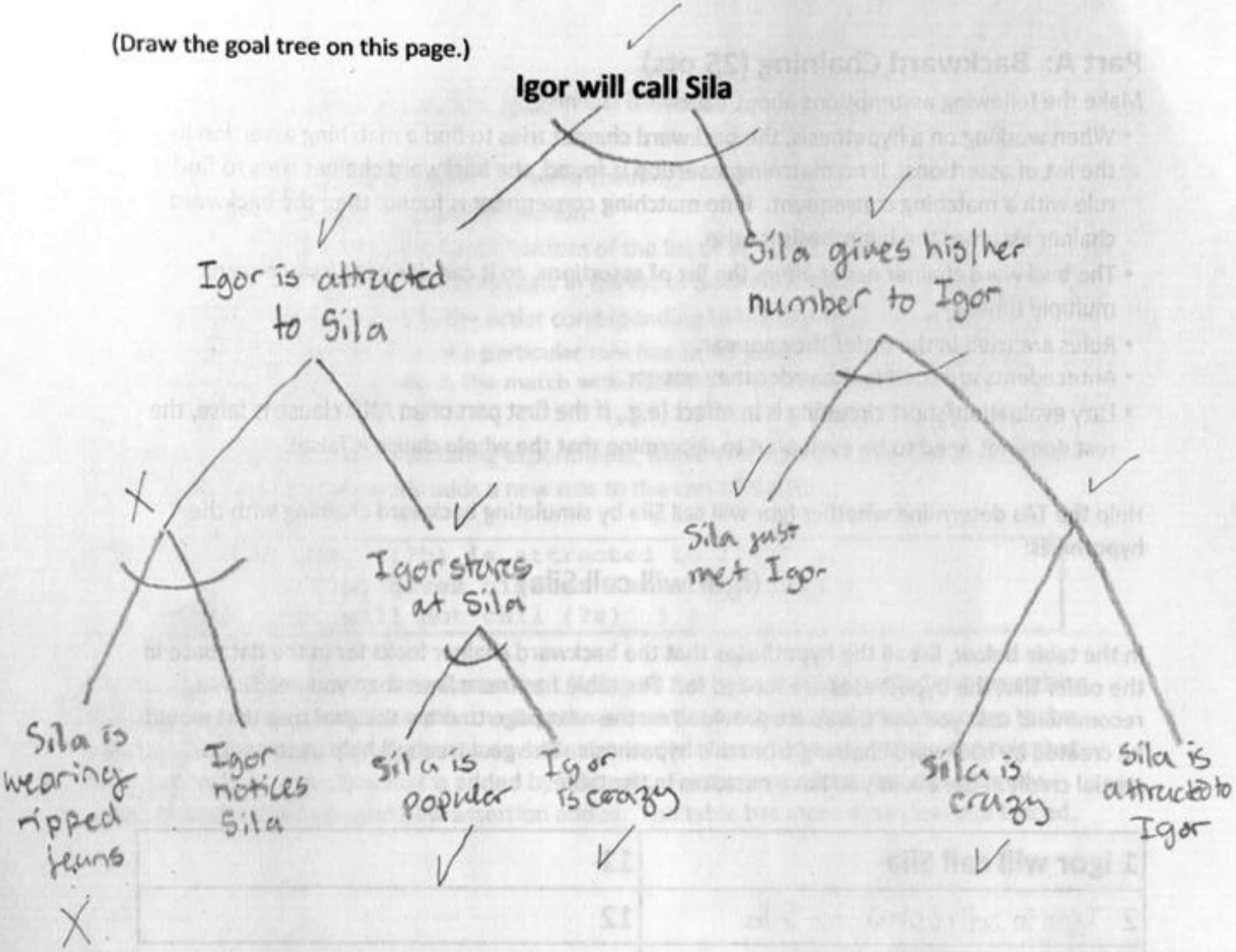
Help the TAs determine whether Igor will call Sila by simulating backward chaining with the hypothesis:

(Igor will call Sila)

In the table below, list all the hypotheses that the backward chainer looks for in the database in the order that the hypotheses are looked for. The table has more lines than you need. We recommend that you use the space provided on the next page to draw the goal tree that would be created by backward chaining from this hypothesis. The goal tree will help us to assign partial credit in the event you have mistakes in the table.

1 **Igor will call Sila**	11
2 Igor is attracted to Sila	12
3 Sila is wearing ripped jeans	13
4 Igor stares at Sila	14
5 Sila is popular	15
6 Igor is crazy	16
7 Sila gives his/her number to Igor	17
8 Sila just met Igor	18
9 Sila is Crazy	19
10	20

(Draw the goal tree on this page.)



Part B: Forward Chaining (25 pts)

Make the following assumptions about forward chaining:

- Assume rule-ordering conflict resolution.
- New assertions are added to the bottom of the list of assertions.
- If a particular rule matches assertions in the list of assertions in more than one way, the matches are considered in the order corresponding to the top-to-bottom order of the matched assertions. Thus, if a particular rule has an antecedent that matches both Assertion 1 and Assertion 2, the match with Assertion 1 is considered first.

B1 (20 points) After some frustrating experiments, Robert realizes that the TAs are missing a very important case, and quickly adds a new rule to the end of the list:

P5

```
IF( AND( NOT( '(?b) is attracted to (?a)' ),
           '(?a) gives his/her number to (?b)')
THEN( '(?b) will not call (?a)' ) )
```

Run forward chaining on the list of assertions (**using all the rules** `P0-P5`) to find out everything that happens with Sila and Igor. For the first three iterations, fill out the first three rows in the table on the next page, noting the rules whose antecedents match the assertions, the rule that fires, and the new assertion that is added by that rule. For the rest of the iterations, simply write down the fired rule and new assertion added. The table has more lines than you'll need.

Iteration	Rules Matched	Rule Fired	New Assertion Added
1	P_0 P_1 P_2	P_0	Sila notices Igor
2	P_0 P_1 P_2 P_3	P_1	Igor stares at Sila
3	P_0 P_1 P_2 P_3	P_1	Sila stares at Sila
4	[redacted]	P_2	Sila gives his/her number to Igor
5	[redacted]	P_3	Sila is attracted to Igor
6	[redacted]	P_3	Igor is attracted to Sila
7	[redacted]	P_3	Sila is attracted to Sila
8	[redacted]	P_4	Igor will call Sila
9	[redacted]		
10	[redacted]		
11	[redacted]		
12	[redacted]		
13	[redacted]		
14	[redacted]		

B2 (5 points)

If you moved the new rule, P5, to the top of the list of the rules, would the results of forward chaining be affected? (Circle the best answer.)

[✓ YES] NO

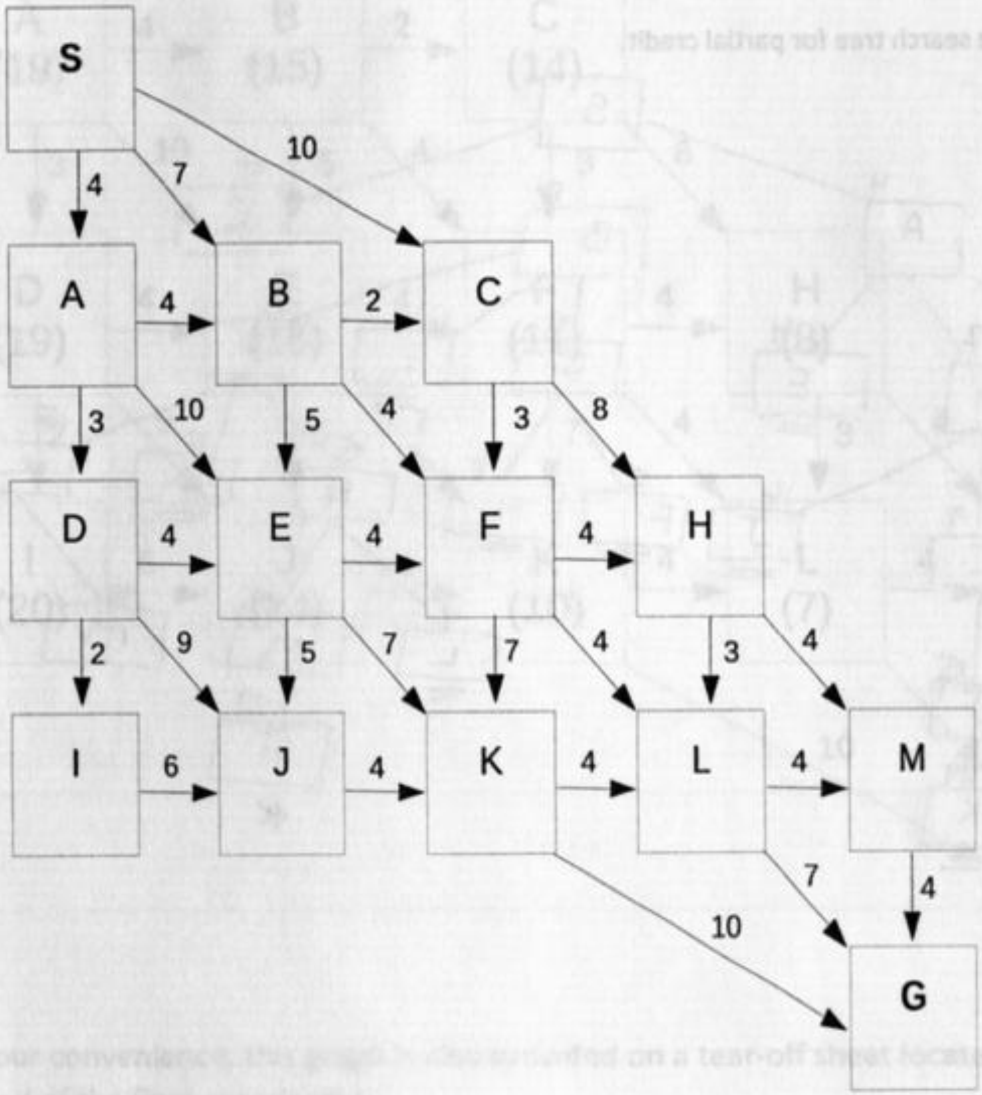
Concisely explain.

When P_5 is at the end of the list, it will match from the end of Iteration 4 until ~~the~~ end of Iteration 6 — but it will not fire, because earlier rules take precedence. When P_5 is at the beginning of the list, it will both match and fire, adding the new assertion "Igor will not call Sila".
(Later, P_4 will also add "Igor will call Sila" as usual.)

Quiz 1, Problem 2: Search & Games (50 pts)

Fed up with academic politics, Igor decides to change careers and pursue his first love, Russian folktales. Alas, as soon as he starts writing his own story, he is hit by a severe case of writer's block and can't decide what should happen next. For inspiration, he borrows Vladimir Propp's breakdown of traditional Russian folktales, which he uses to construct **a graph of possible events that could happen in his story**. He is dismayed to find that Propp's analysis is so full of unoriginal clichés, though, so he also adds **edge weights that correspond to the clichédness** of every development.

Having accomplished this, Igor attempts a search for a sequence of events that leads from the start of his story (S) to the end of his story (G), with the smallest amount of clichédness possible.



For your convenience, this graph is also provided on a tear-off sheet located at the end of the final examination.

Part A: Branch and Bound + Extended List (15 pts)

As a first attempt, Igor decides to use **branch-and-bound search with an extended list,** breaking ties in lexicographic order. In the following sections, **list the nodes** that Igor will add to his extended list, in the order they are extended; **write down** the final path that Igor produces; and **draw the search tree** for partial credit.

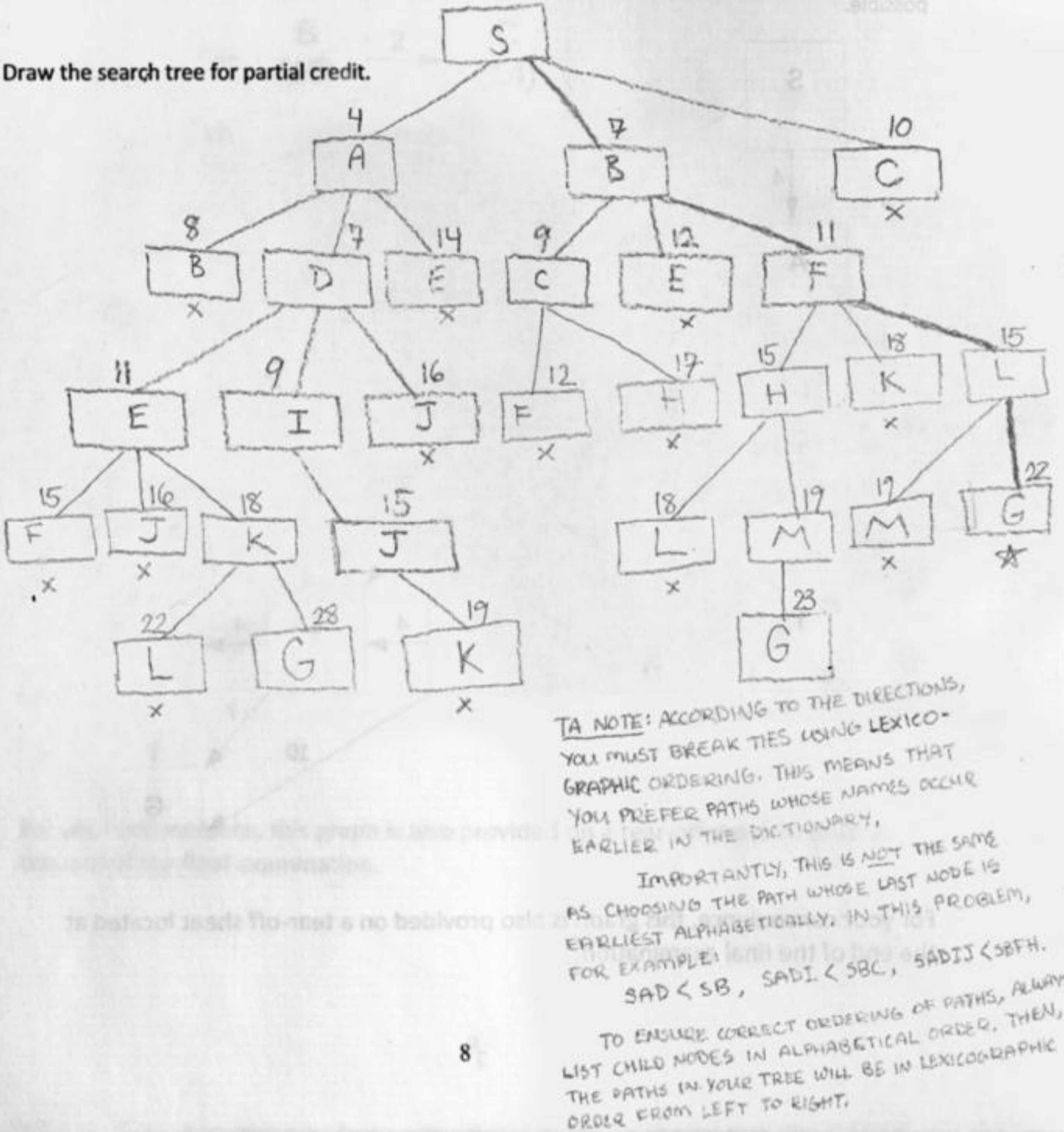
List the nodes that Igor will add to his extended list, in the order they are extended:

S A D B I C E F J H L K M G

Write down the final path that Igor produces:

S → B → F → L → G

Draw the search tree for partial credit.



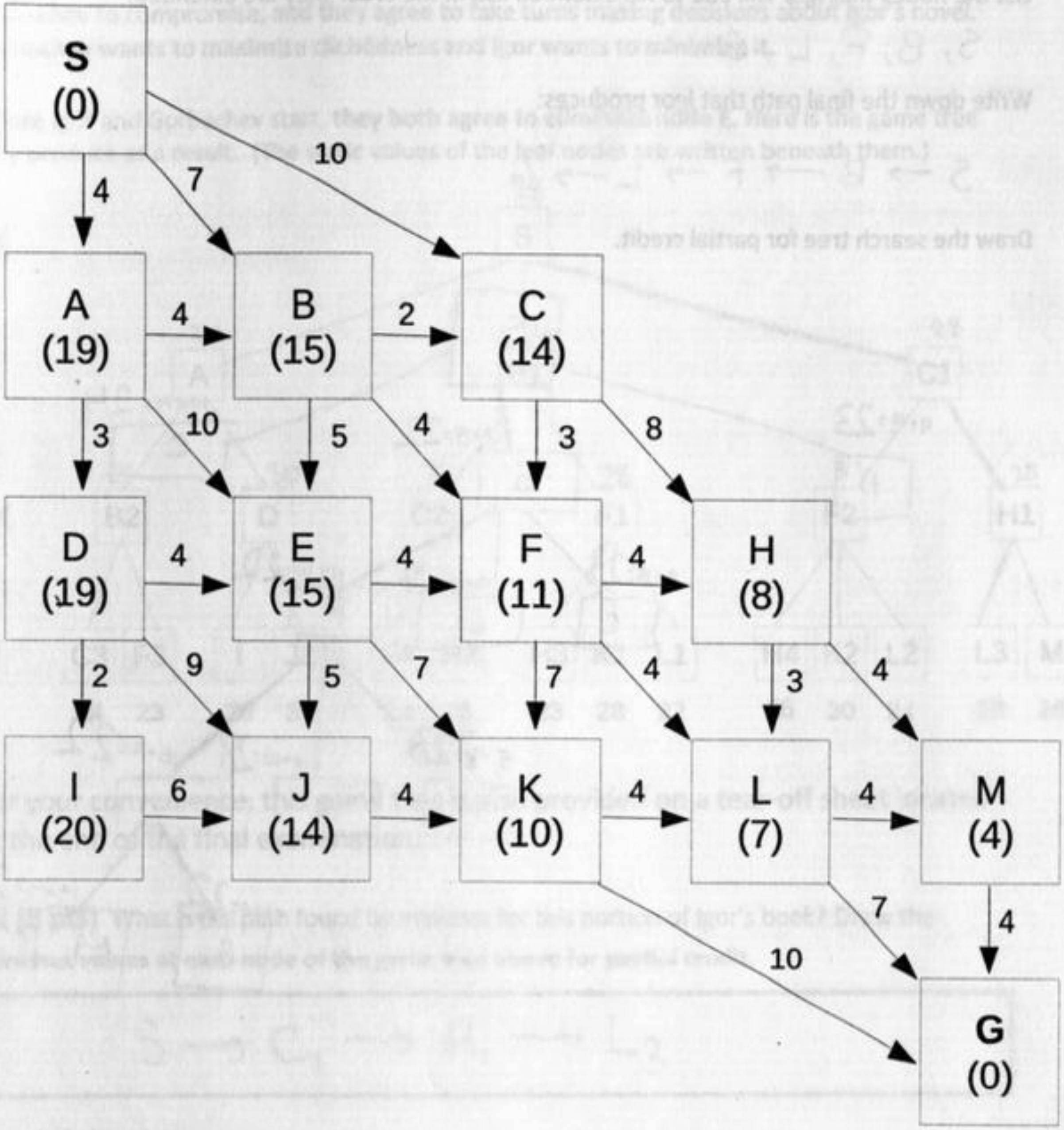
TA NOTE: ACCORDING TO THE DIRECTIONS, YOU MUST BREAK TIES USING LEXICOGRAPHIC ORDERING. THIS MEANS THAT YOU PREFER PATHS WHOSE NAMES OCCUR EARLIER IN THE DICTIONARY.

IMPORTANTLY, THIS IS NOT THE SAME AS CHOOSING THE PATH WHOSE LAST NODE IS EARLIEST ALPHABETICALLY. IN THIS PROBLEM, FOR EXAMPLE: SAD < SB, SADI < SBC, SADIJ < SBFH.

TO ENSURE CORRECT ORDERING OF PATHS, ALWAYS LIST CHILD NODES IN ALPHABETICAL ORDER. THEN, THE PATHS IN YOUR TREE WILL BE IN LEXICOGRAPHIC ORDER FROM LEFT TO RIGHT.

Part B: A* search (15 pts)

Frustrated with the speed of his algorithm, Igor decides add heuristic guesses at each node (written in parentheses) which represent the remaining amount of clichédness until the end of the story.



For your convenience, this graph is also provided on a tear-off sheet located at the end of the final examination.

With his new graph in hand, Igor uses **A* search,** breaking ties in lexicographic order. In the following sections, **list the nodes** that Igor will add to his extended list, <u>in the order they are extended</u>; **write down** the final path that Igor produces; and **draw the search tree** for partial credit.

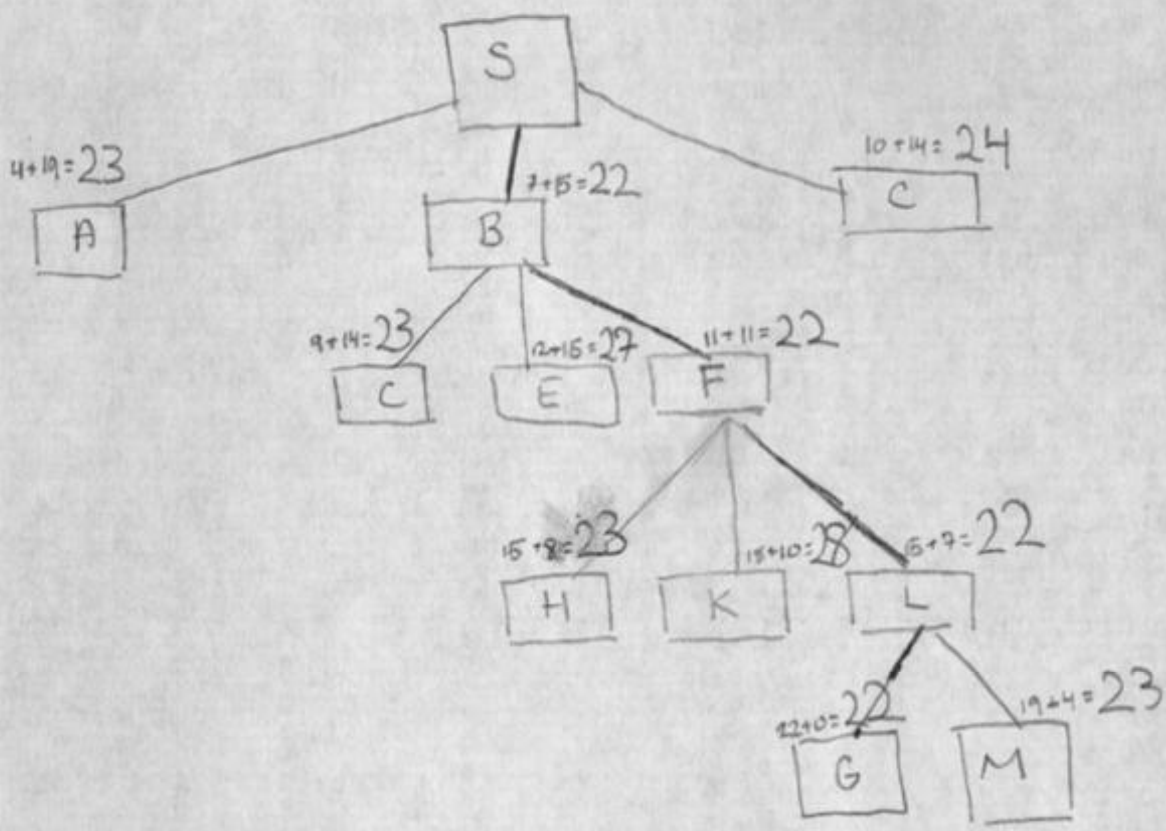
List the nodes that Igor will add to his extended list, in the order they are extended:

S, B, F, L, G

Write down the final path that Igor produces:

S → B → F → L → G

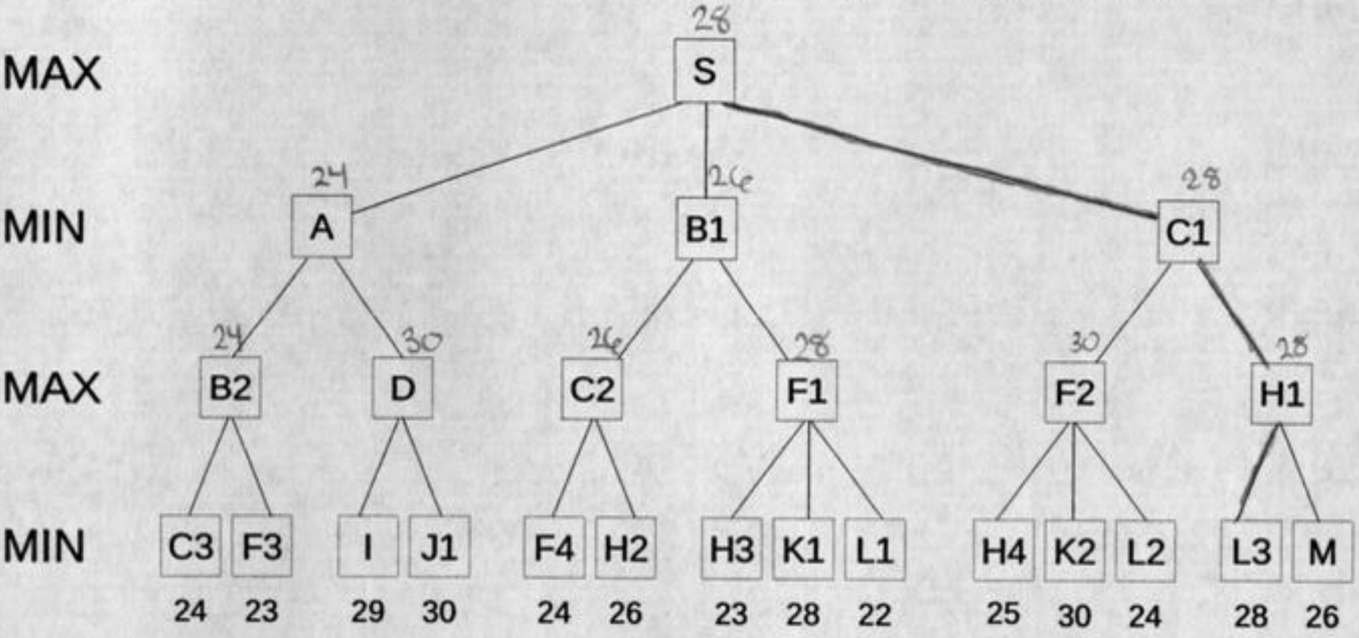
Draw the search tree for partial credit.



Part C: Games (20 pts)

Igor has now created the perfect story and is excited to release it to the world. Unfortunately, his publisher, retired Soviet premier Mikhail Gorbachev, seeing the success of poorly-written clichéd novels like *Twlight* and *Fifty Shades of Grey* claims that clichés (and graphic scenes) are required for a bestseller in this economy, and wants Igor to rewrite his story. Igor implores Gorbachev to compromise, and they agree to take turns making decisions about Igor's novel. **Gorbachev wants to maximize clichédness and Igor wants to minimize it.**

Before Igor and Gorbachev start, **they both agree to eliminate node E.** Here is the game tree they produce as a result. (The static values of the leaf nodes are written beneath them.)



For your convenience, this game tree is also provided on a tear-off sheet located at the end of the final examination.

C1 (8 pts) What is the path found by minimax for this portion of Igor's book? **Draw the minimax values at each node of the game tree above for partial credit.**

$S \to C_1 \to H_1 \to L_3$

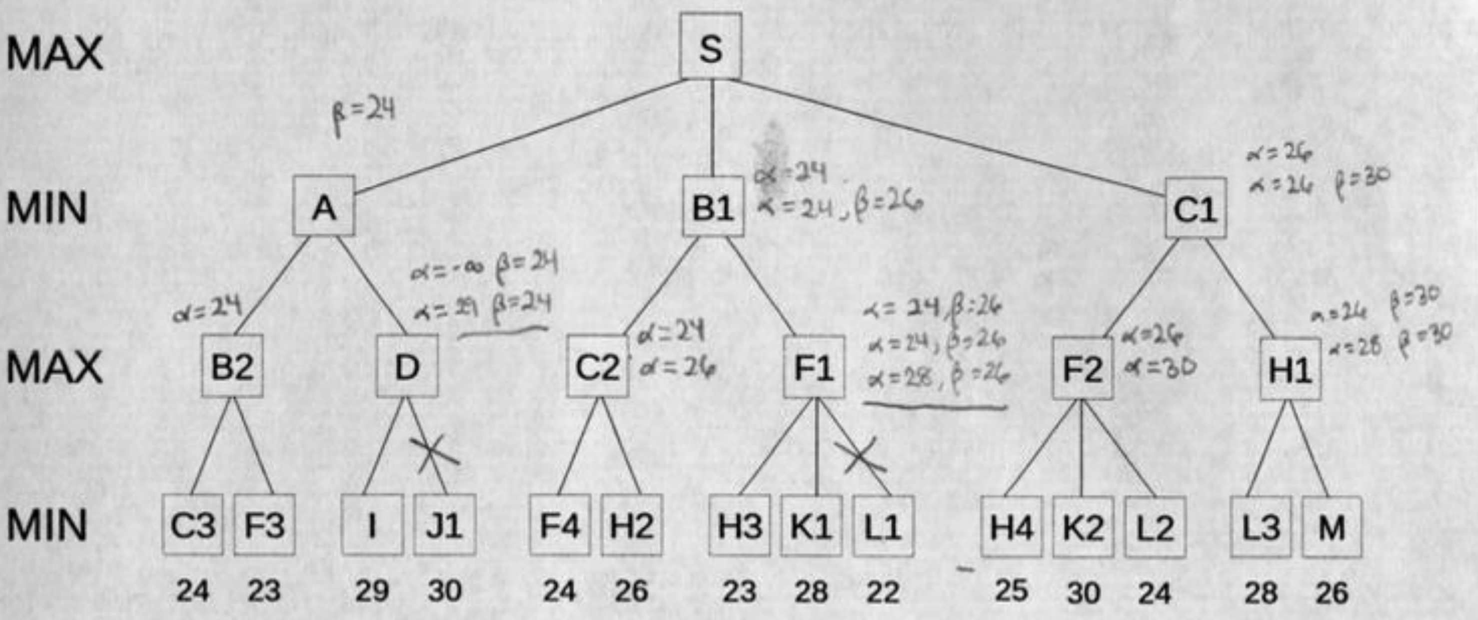
C2 (12 pts) If Igor and Gorbachev use alpha-beta pruning, which nodes would be statically evaluated? (List all the nodes that would be statically evaluated. If none would, write NONE instead.)

C_3 F_3 I F_4 H_2 H_3 K_1 H_4 K_2 L_2 L_3 M

Show your work for partial credit.

$\alpha = 24$
$\alpha = 26$

(all the leaves except J_1 and L_1)



Quiz 2, Problem 1: Constraint Propagation (50 pts)

There are five friends (nicknamed P, A, R, D, F) living in the same town. Each friend has a unique **favorite color** and owns a unique **pet animal.** Your job, as a newcomer to the town, is to figure out the favorite color and pet animal that belongs to each friend.

Pet animals:	dog, cat, turtle, rabbit, gator.
Favorite colors:	pink, white, blue, orange, yellow

! Also, out of the five pets, the **dog, cat,** and **rabbit** are **Mammals,** while the **turtle** and the **gator** are **Reptiles.**

Part A (5 pts): In how many different ways can you assign favorite colors and pet animals to the five friends? (Remember, no two friends can have the same favorite color or the same pet animal.)

$5! \cdot 5!$

Constrained Search

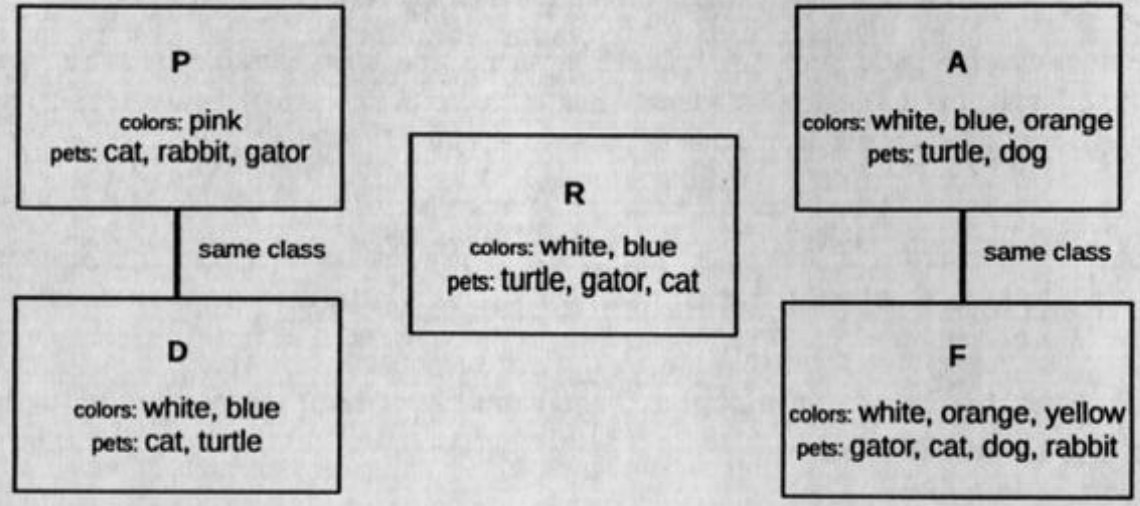
After some investigation, you manage to eliminate some of the possible colors and pets from each friend. **Then, you arrange the remaining possibilities with the most likely ones first.** Here are the results:

Friend	**P**	**A**	**R**	**D**	**F**
Colors	pink	white, blue, orange	white, blue	white, blue	white, orange, yellow
Pets	cat, rabbit, gator	turtle, dog	turtle, gator, cat	cat, turtle	gator, cat, dog, rabbit

You have also learned that:

1. **D** and **P** both own pets with the **same class.** (This means that their pets are both Mammals, or both Reptiles.)
2. **A** and **F** also own pets with the **same class.**
3. One of the five friends has blue as a favorite color and also owns a pet turtle.

You decide to use your skills with constraint propagation to solve this problem. Here is the constraint graph you draw:



Not drawn: Every friend must have a different favorite color and pet
Not drawn: One friend has blue as a favorite color and also owns a pet turtle

- The **variables** of this problem are the five friends (P, A, R, D, F).
- Each **value** in this problem is a combination of a favorite color <u>and</u> a pet animal, for example [white, turtle], or [orange, rabbit].
- The size of a domain is, therefore, the **number of different color/pet <u>pairs</u>** remaining.
- The "same class" constraint between two friends means that their pets must either be both Mammals, or both Reptiles.
- The constraints "Every friend must have a different favorite color and pet" and "One friend has blue as a favorite color and also owns a pet turtle" are not drawn—but of course, they still apply.

Part B (7 pts) Because of the constraint "One friend has blue as a favorite color and also owns a pet turtle", you can immediately eliminate several possible color/pet pairs from each domain. For example, what are the three color/pet pairs remaining in the domain of R?

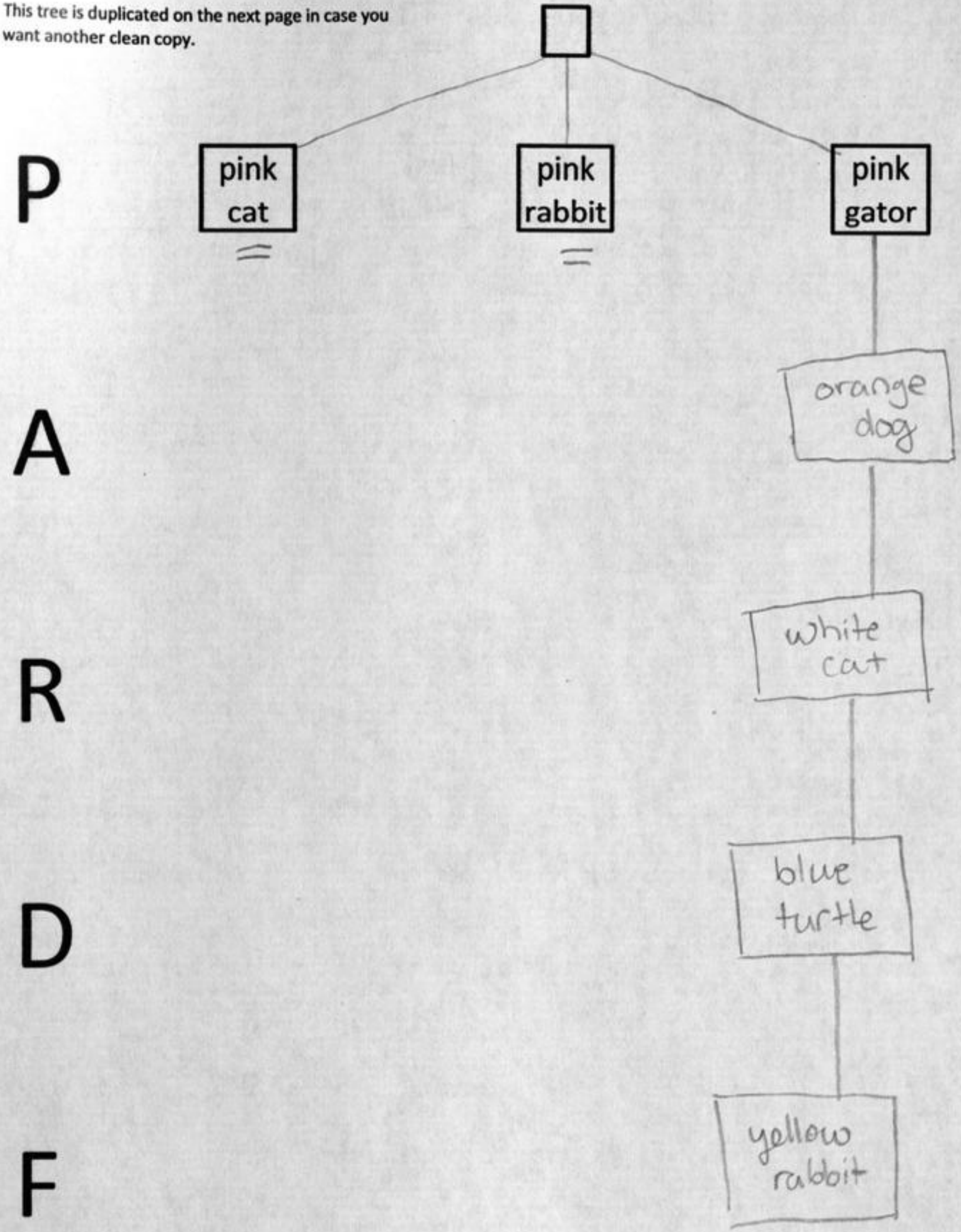
blue turtle, white gator, white cat.

Part C (38 pts) Now use depth first search <u>with propagation through domains reduced to size one</u> to find a consistent assignment of values to variables. Draw the search tree on the following page.

! You must search through the values in the order specified in the constraint graph above. For each friend, start by assigning the first color and first pet, then the first color and second pet, and so on. As noted above, you should <u>completely ignore</u> pairs that are impossible given the constraint "One friend has blue as a favorite color and also owns a pet turtle".

For your convenience, a tear-off sheet with additional copies of the constraint graph are provided after the last page of the final exam.

This tree is duplicated on the next page in case you want another clean copy.



This is a duplicate page. Only use it if you want. ☐

P

pink cat	pink rabbit	pink gator

A

R

D

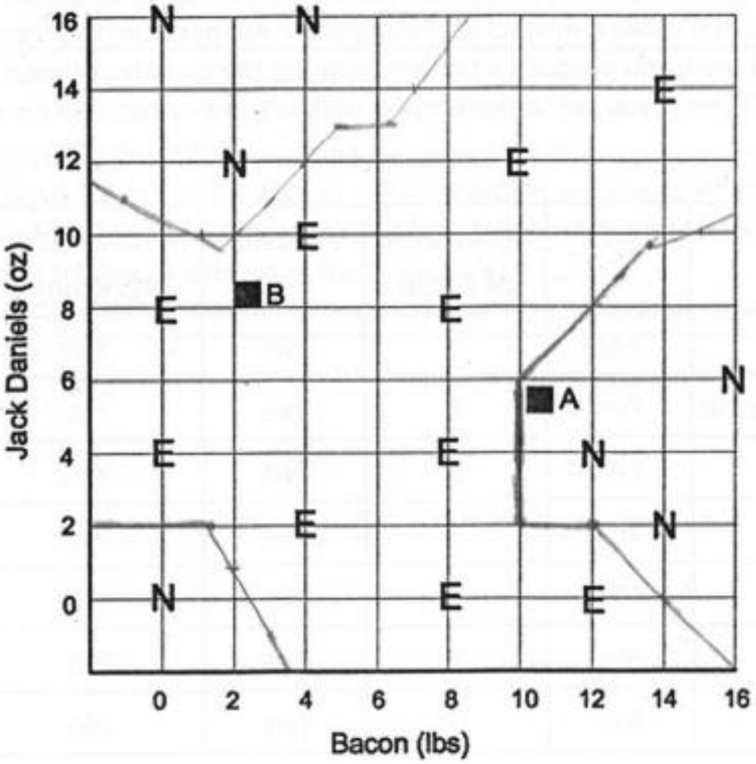
F

Quiz 2, Problem 2: Nearest Neighbors & ID Trees (50 pts)

Lately, your dinners have been sort of bland. After turning to the Internet, you find a set of meals that are absolutely EPIC. You are now inspired to figure out exactly what makes a meal EPIC.

Part A: Nearest Neighbors (20 pts)

The online guides suggest that the most EPIC meals include prodigious amounts of Jack Daniels and Bacon strips (and Bacon strips, and Bacon strips.) Sensing a pattern, you try out a bunch of meals with varying amounts of Jack Daniels and Bacon, then classify them as either EPIC (E) or Not EPIC (N). Your results are plotted in the graph below:



A1 (15 pts)

On the graph above, draw the decision boundaries produced by 1-Nearest-Neighbors. (Ignore the square points A and B for now).

A2 (5 pts):

Your friends invent two new recipes, A and B (the square points plotted on the graph above), which contain varying amounts of Jack Daniels and Bacon. You don't want to waste any more time on terrible food, so you want to predict whether they will be EPIC ahead of time.

How will the points A and B be classified using 1- and 3-Nearest-Neighbors?

	Classification using 1 Nearest Neighbor	Classification using 3 Nearest Neighbors
Point A	N	E
Point B	E	E

Part B: ID Trees (30 points)

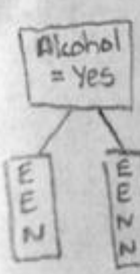
Soon, you realize that EPIC meals are more than just Jack Daniels and Bacon Strips—there are more subtle qualities that make a meal EPIC. Taking some recipes from the Internet and some failed recipes of your own, you produce a table comparing the qualities of each meal. Using your knowledge of ID Trees, you set to work trying to figure out—once and for all—what makes a meal truly EPIC.

Meal	EPIC?	# of Kinds of Meat	Alcohol	Fast Food Ingredients	Deep Fried
TurBacon Epic	Yes	8	No	Yes	No
Cinnaburger Supreme	Yes	1	Yes	No	No
Breakfast Lasagna	Yes	3	No	Yes	No
Meat Car	Yes	3	Yes	No	Yes
Meatloaf	No	3	No	No	No
Fish and Chips	No	1	Yes	No	Yes
Grilled Cheese	No	0	No	No	No

For your convenience, this table is also printed on a tear-off sheet located at the end of the final examination.

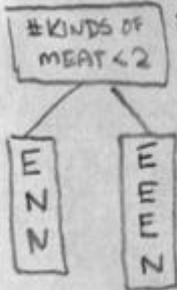
Part B1 (10 pts):

What is the disorder of the test, "Alcohol = Yes"? Express your answer in terms of fractions, real numbers, and logarithms.



$$\frac{3}{7}\left[-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right]+\frac{4}{7}\underbrace{\left[-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}\right]}_{=0}$$

What is the disorder of the test, "# of Kinds of Meat < 2"? Express your answer in terms of fractions, real numbers, and logarithms.

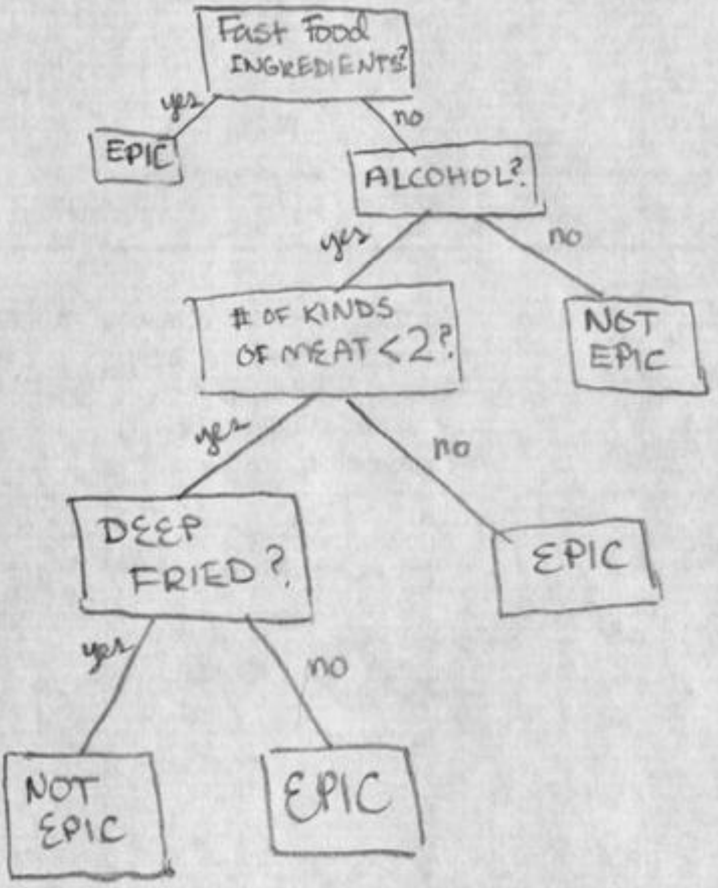


$$\frac{3}{7}\left[-\frac{1}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{2}{3}\right]+\frac{4}{7}\left[-\frac{3}{4}\log_2\frac{3}{4}-\frac{1}{4}\log_2\frac{1}{4}\right]$$

Part B2 (20 pts):

Now, using the features in the table, construct the complete disorder-minimizing identification tree for correctly classifying the given data as "EPIC" (EPIC=Y) or "Not EPIC" (EPIC=N). Note that you can use the "# of Kinds of Meat" test with any threshold. In case of a tie, choose the classifier that appears further to the left in the table.

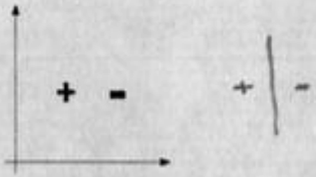
Hint: Neither "Alcohol=Y" nor "# of Kinds of Meat < 2" is the first test.)



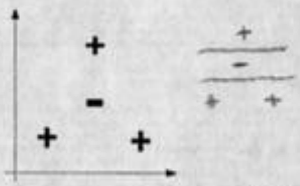
Quiz 3, Problem 1: Neural Nets (50 pts)

Each of the following exhibits specifies the behavior of a function of one or two inputs. Some behaviors are specified with verbal descriptions. Other behaviors are specified with graphs of desired output plotted as a function of input: points marked with "+" should produce an output of 1, while points marked with "−"should produce an output of 0.

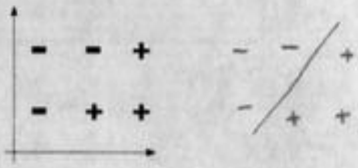
Your task is to choose which of the neural nets on the following page could produce each desired behavior. For each desired behavior, list **all** possible nets that could produce it . If none could, write NONE instead. All neural net thresholds use the unit step function, and threshold inputs are not drawn. If a neural net has only one input, you may choose whether that input receives x-values or y-values. You may choose the same neural net more than once, and some may not be used at all.



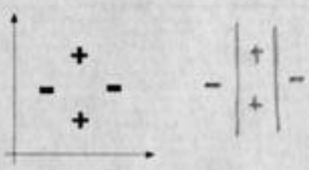
ABCDEF



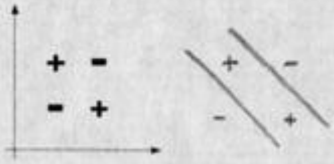
CEF



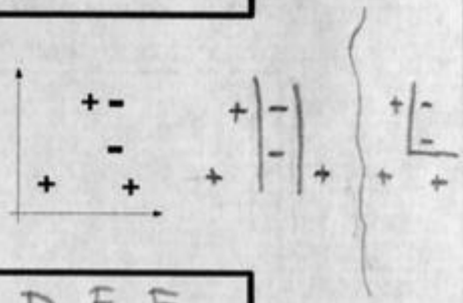
BEF



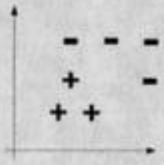
CEF



EF



CDEF



BDEF

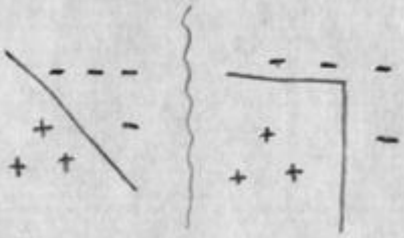
A function which takes as input a temperature in degrees Celsius; it returns one if its input is in a certain (fixed) range of degrees Fahrenheit, otherwise zero.

C (E F)

You can ignore the second input.

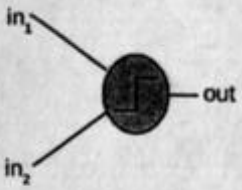
A function which returns one if its inputs are within a certain (fixed) distance of each other, otherwise zero.

E F

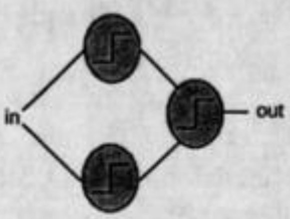



in
out

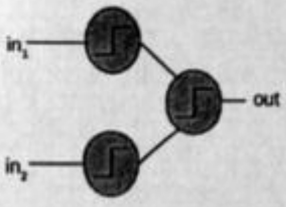
A


in_1
out
in_2

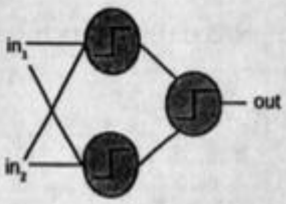
B


in
out

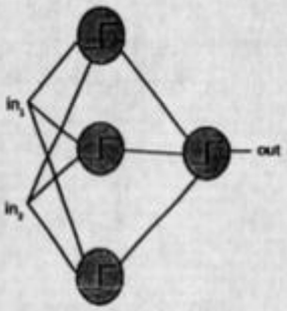
C


in_1
out
in_2

D


in_1
out
in_2

E


in_1
out
in_2

F

Quiz 3, Problem 2: Support Vector Machines (50 pts)

Part A (40 pts)

You have a training set consisting of four training points: A, B, C, D, which are labelled as either positive or negative, as follows:

Positive Points (+)	**A, B**
Negative Points (-)	**C, D**

You created an SVM classifier based on this data. Unfortunately, you've forgotten the decision rule and coordinates of the points. You have, however, jotted down the output of the kernel function for each pair of training points:

K(u,v)	**A**	**B**	**C**	**D**
A	26	4	-4	-6
B	4	6	-15	4
C	-4	-15	2	0
D	-6	4	0	18

You also remember that points B and D are the only support vectors when using this kernel.

A1 (25 pts) Using the above information, compute b and the value of alpha for each point in the training set. Hint: Using the kernel function, you can rewrite the expression $\vec{w}\cdot\vec{x}+b$ as $b+\sum_i \alpha_i \cdot y_i \cdot K(\vec{x}_i,\vec{x})$.

$\alpha_A =$ 0

$\alpha_C =$ 0

$b =$ 3/4

$\alpha_B =$ 1/8

$\alpha_D =$ 1/8

Show your work for partial credit.

Because B and D are the only support vectors, $\alpha_A = \alpha_C = 0$, and B and D must lie on the positive and negative gutters, respectively.

In other words,

(1) $\vec{w}\cdot x_B + b = +1$

(2) $\vec{w}\cdot x_D + b = -1$.

USING THE HINT:

(1) $b + \sum_{i=A,B,C,D} \alpha_i\, y_i\, K(x_i, x_B) = +1$

$b + \alpha_B \cdot (+1) \cdot K(x_B, x_B) + \alpha_D (-1) K(x_D, x_B) = +1$

$\boxed{b + 6\alpha_B - 4\alpha_D = 1}$

(2) $b + \sum_{i=A,B,C,D} \alpha_i\, y_i\, K(x_i, x_D) = -1$

$\boxed{b + 4\alpha_B - 18\alpha_D = -1}$

(3) Finally, because $\sum \alpha_i y_i = 0$ in general, here $\boxed{\alpha_B = \alpha_D}$.

COMBINING EQNS (1), (2), AND (3), WE FIND THAT

$\alpha_B = \alpha_D = 1/8$

$b = 3/4$

A2 (15 pts) Now suppose you also have two test points, X and Y, with unknown classification. The output of your original kernel function with these two test points and each of the training points is:

K(u,v)	**A**	**B**	**C**	**D**
X	2	6	-4	1
Y	0	0	0	0

How would the points X and Y be classified by your SVM? **(Circle the best answer for each):**

Point X: [✓ Positive (+)] Negative (-) Can't be determined

Point Y: [✓ Positive (+)] Negative (-) Can't be determined

Show your work for credit. If you did not solve for the values in Part A1, leave them as variables and write the decision rules for the two test points X and Y, incorporating the information you do know.

A POINT $\vec{u}$ WILL BE CLASSIFIED AS POSITIVE (+) IF $\vec{w}\cdot\vec{u}+b>0$, AND CLASSIFIED AS NEGATIVE (−) IF $\vec{w}\cdot\hat{u}+b<0$. IF $\vec{w}\cdot\vec{u}+b=0$, A CLASSIFICATION CAN'T BE DETERMINED.

USING THE HINT FROM A_1, WE CAN REWRITE THESE CONDITIONS AS:

$$b+\sum_{i=A,B,C,D}\alpha_i y_i K(\vec{x}_i,\vec{u}) \begin{array}{l} >0 \Rightarrow \text{POSITIVE} \\ <0 \Rightarrow \text{NEGATIVE} \\ =0 \Rightarrow \text{CAN'T BE DETERMINED.} \end{array}$$

POINT X:

$$b+\sum_{i=A,B,C,D}\alpha_i y_i K(\vec{x}_i,X)$$

$$=b+\alpha_B(+1)K(B,X)+\alpha_D(-1)K(D,X)$$

$$=\frac{3}{4}+\frac{1}{8}(6)-\frac{1}{8}(1)$$

$>0 \Rightarrow$ POSITIVE

POINT Y:

$$b+\sum_{i=A,B,C,D}\alpha_i y_i K(\vec{x}_i,Y)$$

$$=b+\alpha_B(+1)K(B,Y)+\alpha_D(-1)K(D,Y)$$

$$=\frac{3}{4}+\frac{1}{8}(0)-\frac{1}{8}(0)$$

$>0 \Rightarrow$ POSITIVE

Part B (5 pts) You're now concerned about the amount of time it takes to classify an unknown test point, where time is measured in terms of the number of standard arithmetic operations (additions, multiplications, etc.) that need to be performed.

The amount of time it takes to decide the classification an unknown test point using a table of kernel values and an SVM decision rule (such as the one in Part A) depends on: **(Circle the one best answer)**

3

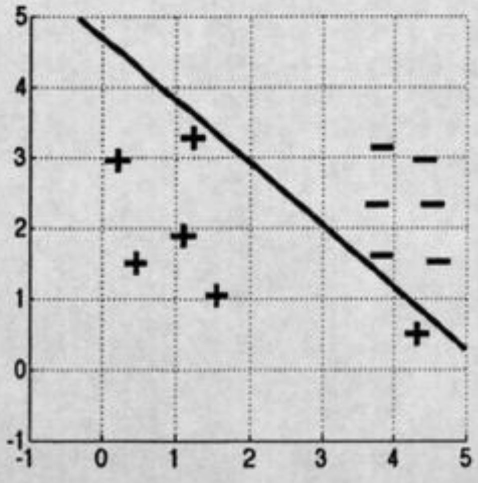
1. The number of training examples
2. The width of the margin
[✓ 3.] The number of support vectors
4. The number of possible classifications
5. Whether the test point lies inside the gutter or not.
6. Whether the test point is an outlier or not.
7. The distance of the test point from the decision boundary.
8. The distance of the test point from the closest support vector

Part C (5 pts)

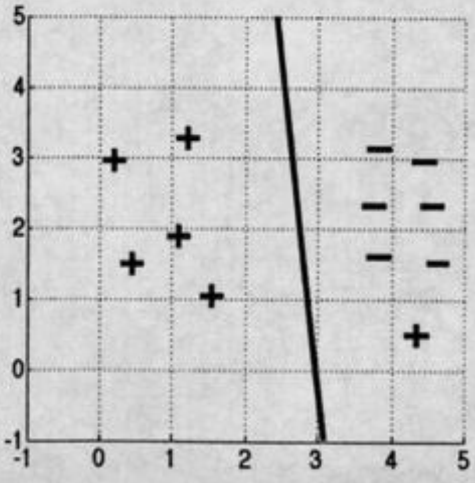
The type of SVM we've been learning about in 6.034 is called the hard-margin SVM: whenever, given the kernel, it is possible to perfectly separate the training data, the hard-margin SVM will do so.

After the hard-margin SVM was published, a modified version known as the soft-margin SVM was described. A soft-margin SVM can allow some of the training data to be misclassified, even when the data could be perfectly separated.

Compare the decision boundaries created by the hard-margin SVM and a soft-margin SVM using a linear kernel on the same training dataset.



Hard-margin SVM



Soft-margin SVM

What problem does the soft-margin SVM avoid in this case? A one-word answer is all you need.

OVERFITTING

Quiz 4, Problem 1: Adaboost (50 pts)

The Science of Deduction

You have just acquired an aging problem set about Adaboost from the year 1995. Because many of the entries are no longer legible, you decide to employ your knowledge of Adaboost to determine some of the missing information.

Below, you can see part of table that was used in the problem set. There are columns for the Round # and for the weights of the six training points (A, B, C, D, E, and F) at the start of each round. Some of the entries, marked with "?", are impossible for you to read.

<u>Note:</u> Questions about this table are located on the next few pages. It's depicted here just to give you a glimpse of the whole table.

	Round #	weight$_A$	weight$_B$	weight$_C$	weight$_D$	weight$_E$	weight$_F$
	1	?	?	1/6	?	?	?
❶	2	?	?	?	?	?	?
	⋮						
	219	?	?	?	?	?	?
❷	220	1/14	1/14	7/14	1/14	2/14	2/14
❸	221	1/8	1/8	7/20	1/20	1/4	2/20
	⋮						
❹	3017	1/2	1/4	1/8	1/16	1/16	0
	⋮						
❺	6034	1/8	3/8	1/8	2/8	3/8	1/8
	⋮						
❻	8888	?	1/10	3/10	?	?	?
	8889	3/20	1/10	?	?	?	3/20
	⋮						
❼	9999	?	?	?	?	?	?
	10000	2/30	2/30	?	9/30	10/30	5/30

In the following problems, you may assume that non-consecutive rows are independent of each other, and that a classifier with error < ½ was chosen at each step.

❶ **(8 pts)** According to a scribbled note in the margin, the weak classifier chosen in **Round 1** correctly classified training points A, B, C, and E but misclassified training points D and F. What should the updated weights have been in the following round, **Round 2**?

weight$_A$	**weight$_B$**	**weight$_C$**	**weight$_D$**	**weight$_E$**	**weight$_F$**
1/8	1/8	1/8	1/4	1/8	1/4

Show your work for partial credit.

❷ **(7 pts)** During **Round 219,** which of the training points (A, B, C, D, E, F) must have been misclassified, in order to produce the updated weights shown at the start of **Round 220**?

Round #	**weight$_A$**	**weight$_B$**	**weight$_C$**	**weight$_D$**	**weight$_E$**	**weight$_F$**
219	?	?	?	?	?	?
220	1/14	1/14	7/14	1/14	2/14	2/14
221	1/8	1/8	7/20	1/20	1/4	2/20

List all the points that were misclassified. If none were misclassified, write NONE instead.

C

(THE SUM OF THE WEIGHTS OF THE MISCLASSIFIED POINTS MUST EQUAL 1/2.)

❸ **(7 pts)** During **Round 220,** which of the training points (A, B, C, D, E, F) must have been misclassified in order to produce the updated weights shown at the start of **Round 221**?

List all the points that were misclassified. If none were misclassified, write NONE instead.

A, B, E

❹ **(7 pts)** Your friend John observes that the weights in **Round 3017** cannot possibly be right. Why not?

Round #	**weight$_A$**	**weight$_B$**	**weight$_C$**	**weight$_D$**	**weight$_E$**	**weight$_F$**
3017	1/2	1/4	1/8	1/16	1/16	0

WEIGHTS CAN NEVER BECOME ZERO.

❺ **(7 pts)** Your friend Sherlock observes that the weights in **Round 6034** cannot possibly be right. Why not?

Round #	weight$_A$	weight$_B$	weight$_C$	weight$_D$	weight$_E$	weight$_F$
6034	1/8	3/8	1/8	2/8	3/8	1/8

THE SUM OF THE WEIGHTS MUST EQUAL ONE.

❻ **(7 pts)** You observe that there must also have been a mistake in the weights of **Round 8888** or **Round 8889**. Why?

Round #	weight$_A$	weight$_B$	weight$_C$	weight$_D$	weight$_E$	weight$_F$
8888	?	1/10	3/10	?	?	?
8889	3/20	1/10	?	?	?	3/20

WEIGHTS MUST CHANGE EVERY ROUND. (ERROR $\neq \frac{1}{2}$)

❼ **(7 pts)** Finally, suppose you know that at least half of the training points were classified correctly during **Round 9999**. Knowing this, which of the training points (A, B, C, D, E, F) were given a lower weight in the next round, **Round 10000**?

Round #	weight$_A$	weight$_B$	weight$_C$	weight$_D$	weight$_E$	weight$_F$
9999	?	?	?	?	?	?
10000	2/30	2/30	?	9/30	10/30	5/30

List all the points that were given a lower weight in Round 10000. If none of the points were given a lower weight, write NONE instead.

A, B, C, D

EXPLANATION: (1) SINCE THE SUM OF THE WEIGHTS MUST EQUAL ONE, weight$_C = 2/30$ IN ROUND 9999.

(2) SINCE THE SUM OF THE WEIGHTS OF THE MISCLASSIFIED POINTS MUST EQUAL $\frac{1}{2}$, AND THE SUM OF THE WEIGHTS OF THE CORRECTLY CLASSIFIED POINTS MUST EQUAL $\frac{1}{2}$, THEN EITHER A B C D WERE CORRECTLY CLASSIFIED AND EF WEREN'T, OR VICE-VERSA.

(3) SINCE WE ARE TOLD TO ASSUME THAT ERROR $< \frac{1}{2}$ EACH ROUND, THE CORRECTLY CLASSIFIED POINTS WILL HAVE GOTTEN A LOWER WEIGHT.

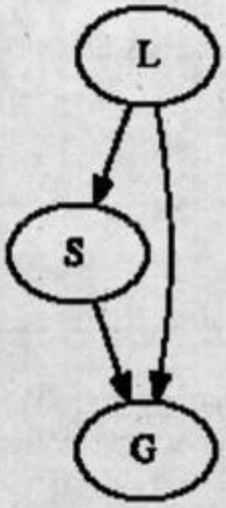
(4) SINCE AT LEAST HALF THE POINTS WERE CLASSIFIED CORRECTLY, IT MUST BE THAT A B C D RATHER THAN EF WERE CLASSIFIED CORRECTLY. HENCE, A B C D GOT LOWER WEIGHTS THE NEXT ROUND.

Quiz 4, Problem 2: Bayes Nets (50 pts)

For his senior thesis, Louis Reasoner decides to model the interaction between Cantabrigians' views on global warming (adherent or skeptic), their political affiliation (liberal or conservative), and their background in science (scientist or Non-scientist).

Part A (10 pts)

Louis starts out with the following Bayes Net to model the belief dynamics . Here, G denotes "Subject is an adherent of Global warming", L denotes "Subject has a Liberal political affiliation", and S indicates "Subject is a Scientist".



A1 (5 pts) **Write out an expression for computing the joint probability distribution—that is, $P(S \wedge L \wedge G)$ —for this model in terms of the marginal/conditional probabilities which are its parameters.**

$P(S \wedge L \wedge G) = P(L) \cdot P(S|L) \cdot P(G|S,L)$

A2 (5 pts) **Is there a joint probability distribution that wouldn't satisfy the constraints imposed by this model? (Circle the best answer.)**

YES [✓ NO]

Concisely explain why or why not.

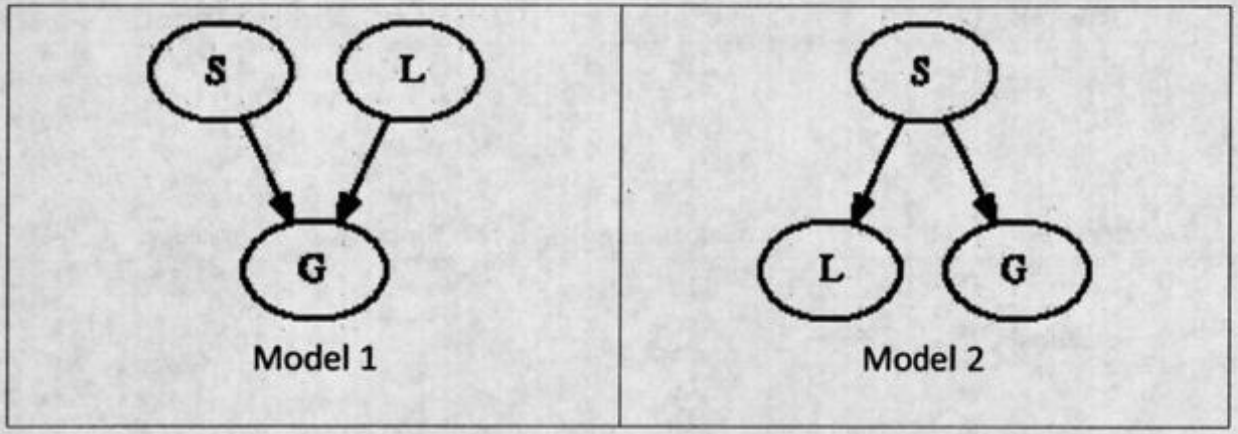
WITH NO INDEPENDENCE ASSUMPTIONS, A MODEL WITH 3 BINARY VARIABLES HAS $2^3 - 1 = 7$ PARAMETERS.

LOUIS' BAYES NET HAS $1+2+4=7$ PARAMETERS, AS WELL — SO, IN FACT, HIS BAYES NET IMPOSES NO CONSTRAINT; EVERY JOINT PROBABILITY DISTRIBUTION WILL AGREE WITH THIS MODEL.

L S G	P(L∧S∧G)
f f f	①
f f T	②
f T f	③
f T T	④
⋮	⋮
T T T	⑧

Part B (30 pts)

After examining his assumptions further, Louis decides that the right model is one of the following two candidates instead:



B1 (10 pts) **In the table below, list the models (Model 1, Model 2) for which the following conditional/marginal independence statements are true. If the statement applies to none of the models, write NONE instead.**

Independence statement	**Models in which the statement is true**
S is independent of L	MODEL 1
G is conditionally independent of S given L	NONE
L is independent of G	NONE
S is conditionally independent of L given G	NONE
S is independent of G	NONE

B2 (10 pts) **Write an expression for the joint probability distribution—that is, $P(S \wedge L \wedge G)$ —for each of the models in terms of the marginal/conditional probabilities which are its parameters:**

Model 1: $P(S \wedge L \wedge G) = P(S) \cdot P(L) \cdot P(G | S, L)$

Model 2: $P(S \wedge L \wedge G) = P(S) \cdot P(L | S) \cdot P(G | S)$

B3 (10 pts) In order to assign probabilities to his three models, Louis has collected the following data from a survey of people in and around MIT campus.

Liberals (L)	Global Warming adherent (G)	Global Warming skeptic $(\overline{G})$
Scientist (S)	20	1
Non-scientist $(\overline{S})$	80	20

Conservatives $(\overline{L})$	Global Warming adherent (G)	Global Warming skeptic $(\overline{G})$
Scientist (S)	10	2
Non-scientist $(\overline{S})$	50	100

Each of the Models has a certain number of parameters, namely the entries in their conditional/marginal probability tables. Use the table of data above to **construct the conditional/marginal probability tables** required for each of the [redacted] models.

Hint #1: Given a table of frequencies like the one above, you can estimate a marginal probability $P(X=x)$ as the ratio of the number of samples in which $X=x$ was observed, to the total number of samples.

Hint #2: Given a table of frequencies like the one above, you can estimate a conditional probabiliy $P(X=x|Z_1=z_{1,}...,Z_n=z_n)$ as the ratio of the number of samples in which $X=x,Z_1=z_{1,}...,Z_n=z_n$ was observed, to the number of samples in which $Z_1=z_{1,}...,Z_n=z_n$ was observed.

Show your work below.

Probability expression	**Which models use this parameter?**	**Numerical value**	**Show your work for partial credit.**
P(S)	1, 2	33/283	[redacted]
P(G|S=T)	2	10/11	
P(G|S=F)	2	13/25	
P(L|S=T)	2	21/33	
P(L|S=F)	2	2/5	
P(G|S,L)	1	20/21	
P(G|S=F,L)	1	4/5	
P(G|S=F, L=F)	1	1/3	
P(G|S, L=F)	1	5/6	
P(L)	1	121/283	

Part C (10 pts) Now that Louis estimated the parameters for the possible campus belief networks, he would like to evaluate each of the models on new evidence. He independently polls two new subjects from the Continental US:

Subject 1 is a non-scientist Liberal who is a skeptic of Global Warming. (not S, L, not G)
Subject 2 is a non-scientist Conservative who is an adherent of Global Warming. (not S, not L, G)

Determine which of the models is better (the model with the highest posterior probability), using these two subjects as evidence and **assuming that the models have equal prior probability** $P(M_1)=P(M_2)$.

Which model is more probable, based on this evidence? (Circle one.)

Model 1 [✓ Model 2]

You must show work for credit.

$$P(M_1|E) = P(E|M_1)\cancel{\frac{P(M_1)}{P(E)}} \propto P(\text{Subj}_1 \,\&\, \text{Subj}_2 | M_1) \overset{\text{indep. poll}}{=} P(\text{Subj}_1|M_1)\,P(\text{Subj}_2|M_1)$$

$$= P(\bar{S} \,\&\, L \,\&\, \bar{G}|M_1) \cdot P(\bar{S} \,\&\, \bar{L} \,\&\, G | M_1)$$

$$= \left[P(\bar{S})P(L)P(\bar{G}|L,\bar{S})\right] \cdot \left[P(\bar{S})P(\bar{L})P(G|\bar{S},\bar{L})\right]$$

$$\boxed{= \left(1-\frac{33}{283}\right)\left(\frac{121}{283}\right)\left(1-\frac{4}{5}\right) \cdot \left(1-\frac{33}{283}\right)\left(1-\frac{121}{283}\right)\left(\frac{1}{3}\right)}$$

The models have equal prior probability

$$P(M_2|E) = P(E|M_2)\cancel{\frac{P(M_2)}{P(E)}} \propto P(\text{Subj}_1 \,\&\, \text{Subj}_2 | M_2) \overset{\text{indep. poll}}{=} P(\text{Subj}_1|M_2)\,P(\text{Subj}_2|M_2)$$

The evidence occurs in the expression for both models, so we can ignore it.

$$= P(\bar{S} \,\&\, L \,\&\, \bar{G}|M_2) \cdot P(\bar{S} \,\&\, \bar{L} \,\&\, G | M_2)$$

$$= \left[P(\bar{S})P(L|\bar{S})P(\bar{G}|\bar{S})\right] \cdot \left[P(\bar{S})P(\bar{L}|\bar{S})P(G|\bar{S})\right]$$

$$\boxed{= \left(1-\frac{33}{283}\right)\left(\frac{2}{5}\right)\left(1-\frac{13}{25}\right) \cdot \left(1-\frac{33}{283}\right)\left(1-\frac{2}{5}\right)\left(\frac{13}{25}\right)}$$

BY CALCULATION/INSPECTION, WE FIND THAT $P(M_2|E) > P(M_1|E)$, SO MODEL $\underline{2}$ IS MORE LIKELY, GIVEN THE AVAILABLE EVIDENCE.