6.034 Final Examination
December 18, 2006

Name	Scott Young
EMail	

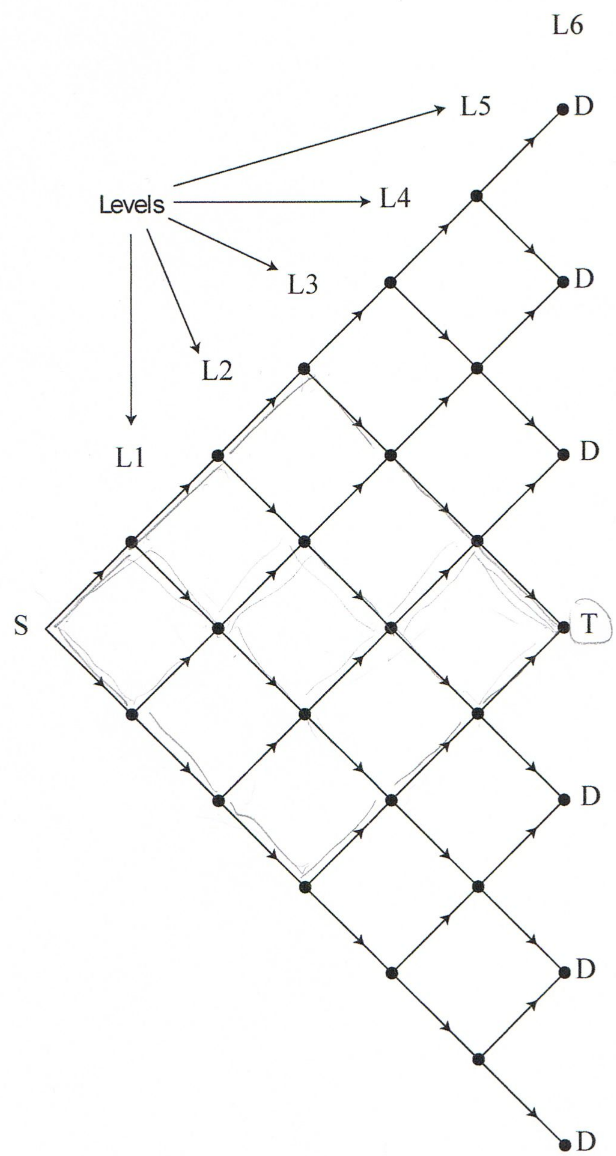
Quiz number	Maximum	Score	Grader
1	100	80	
2	100	85	
3	100	65	
4	100	69	
5	100	69	

= 73.6

There are 30 pages in this final, including this one and a tear-off sheet provided at the end with duplicate drawings and data. You must do Quiz 5. All others are optional.

Quiz 1
Problem 1: Search (75 points)

This problem has a certain resemblance to a problem on Quiz 1, but it is not the same.
Several 6.034 students are stranded in a strange city at the node marked S. Desperate for a meal, they want to get to Hose's food truck located on the map at the node marked T. All the streets are one-way streets. Each link between node pairs is 1 unit long.



Part A (10 points)

Amy calculates how many paths terminating at nodes T or D she would produce if she used the British Museum Algorithm. As expected with the British Museum Algorithm, she does not use an extended list. Her result is:

$\boxed{2^6 = 64}$

Part B (10 points)

David decides to use plain branch and bound, **without an extended list** and with no estimate of remaining distance, to compute the shortest path between S and T. If there is ever a choice in which path to extend, he picks the path closest to the top of the page. Before he can be sure he has the shortest path, he produces n paths starting at S and ending at a node marked T or D and adds those paths to the queue. In this case, n is

$\boxed{2^6 = 64}$

Part C (15 points)

James thinks it would be better to use an estimate of remaining distance with the basic branch and bound procedure. Also, **he does use an extended list**. His estimate for each node is the straight-line distance to T.

Is James's heuristic admissible? $\boxed{\text{Yes}}$

If there is ever a choice in which path to extend, he picks the path closest to the top of the page. Before he can be sure he has the shortest path, he produces n paths starting at S and ending at a node marked T or D and adds those paths to the queue. In this case, n is

$\boxed{2}$

Part D (10 points)

Natalya thinks James is on to something, but in her analysis, Natalya uses branch and bound together with a different estimate of remaining distance:

L1	L2	L3	L4	L5	L6
5	4	3	2	1	0

Is Natalya heuristic admissible? Yes

Natalya uses an extended list. If there is ever a choice in which path to extend, she picks the path closest to the top of the page. Before she can be sure she has the shortest path, she produces n paths starting at S and ending at a node marked T or D and adds those paths to the queue. In this case, n is 2x 12

-5

Part E (20 points)

Olga thinks Natalya is on to something and uses the same method, except that Olga's estimate of remaining distance is Natalya's + 1:

L1	L2	L3	L4	L5	L6
6	5	4	3	2	1

E1: Is Olga's heuristic admissible? No, heuristic overestimates distance.

E2: Does Olga get the same path as Natalya? Yes

-5

Suppose that Natalya changes her mind and uses different estimates (but still ≥ 0). Olga's estimates are still Natalya's +1. Can your answer to E1 change? No ~~Yes if Natalya underestimates distance by at least 1~~

Again, suppose that Natalya has changed her mind and used different estimates (but still ≥ 0). Olga's estimates are still Natalya's +1. Can your answer to E2 change? No, because the shape of the graph, they will always get the same paths

Part F (10 Points)

Raluca thinks it would be a good idea to use branch and bound **with an extended list**, and she also decides not to bother with an estimate of remaining distance. However, she starts at T and looks for S, moving only backwards along the one way streets.

If there is ever a choice in which path to extend, she picks the path closest to the top of the page. Before she can be sure she has the shortest path, she produces n paths starting at T and ending at a node marked S and adds those paths to the queue. In this case, n is

$\boxed{6 \times 2}$

-10

Quiz 1
Problem 2: Rules (25 points)

Susan is developing a chemical synthesis system. Her system has powerful rules that figure out what simpler chemicals are needed to synthesize a given chemical, and the even simpler chemicals needed to synthesize the simpler pieces, and so on for many levels.

She estimates costs for chemicals by making a lot of telephone calls. Generally, the simpler the chemical, the cheaper it is. It is worth noting that the calls take a lot of time and that Susan can never remember the cost of a chemical she checked out previously.

In any case, because she wants to use the cheapest chemicals possible, she decides to have her system to do a lot of rule chaining before she calculates the cost of the chemicals at the bottom of the tree produced by the chaining.

Part A (10 Points)

Suppose that for any given chemical, Susan's system knows how to synthesize it out of b simpler chemicals and that she decides to let her system backward chain through d levels of rules. How many costs of chemicals will she have to determine to estimate the cost of the synthesis?

$\boxed{b^d}$

Part B (15 Points)

Unfortunately, Susan's hyper critical boss, says he cannot always wait for an answer---he wants a just-in-time system that provides a cost estimate whenever he asks for one, and perhaps a cheaper solution if he asks later.

Holly suggests that Susan exploit the progressive deepening idea commonly used in playing games. This means Susan's system will have to compute costs at every level of the tree, not just the bottom level. Susan worries that much time will be wasted. Holly assures her that the time wasted in ensuring a just-in-time answer down to level d-1 will only be a fraction of the time needed to do the estimates at level d. The fraction is approximately

$\boxed{\doteq 1/b}$

$= \sum_{i=1}^{d-1} (1/b^i)$

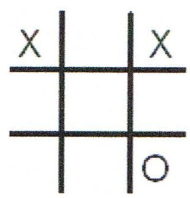
Quiz 2
Problem 1: Search in Games (50 points)

In the game of Tic-Tac-Toe, your goal is to get three O's in a row, either horizontally, vertically, or diagonally. Players alternate placing X's and O's on the board. If the opponent places three X's in a row, you lose.

You decide to use minimax-with progressive deepening to determine which moves you should make. You use the following heuristic to perform static evaluation:

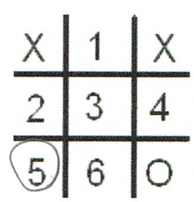
heuristic(state) = -5 if there are three X's in a row
5 if there are three O's in a row
the maximum number of O's in a diagonal, row, or column with no X's.
0 otherwise

Suppose the game board looks like this:



Part A (10 points)

The six possible places that you can place an O are indicated in the diagram, labeled 1, 2, 3, 4, 5, and 6.



Where would you place an O given the static values associated with each choice and no further search? (In case of a tie, enter the lower/lowest number)

5

Part B (20 points)

Because you do not know how deep you will be able to search, you decide to use the progressive deepening idea. Accordingly, you decide to reorder the branches before searching to the next level. Which ordering do you use to evaluate the children of nodes 1-6? (In case of a tie, enter the lower/lowest number)

1 2 3 4 5 6

1st 5
2nd 6
3rd 1
4th 2
5th 3
6th 4

10/10

If you place the next O in cell *c*, which spot on the tic-tac-toe board does the opponent place the next X, given the static values produced by the heuristic. (In case of a tie, enter the lower/lowest number)

If c = 1, the opponent places the next X in cell: ~~2~~ 6
If c = 2, the opponent places the next X in cell: 1
If c = 3, the opponent places the next X in cell: 1
If c = 4, the opponent places the next X in cell: 1
If c = 5, the opponent places the next X in cell: 1
If c = 6, the opponent places the next X in cell: 1

5/10

X	1	X
2	3	4
5	6	O

Part C (10 points)

Now, consider the possible board positions after you have placed a O and your opponent has placed an X. After applying the minimax procedure to the static values of those board positions, which square do you choose to place the first O? (In case of a tie, enter the lower/lowest number)

1

Part D (10 points)

You desire to use progressive deepening to the next level, depth 3. Accordingly, you decide to reorder the six branches of the minimax tree that descend from the board position given in Part A. Which ordering do you use? (In case of a tie, enter the lower/lowest number)

1st 1

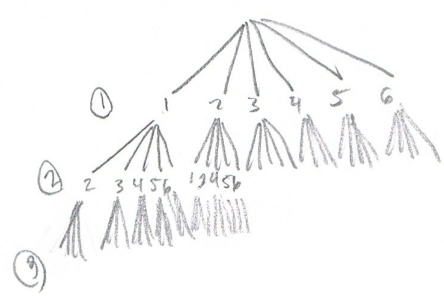
2nd 2

3rd 3

4th 4

5th 5

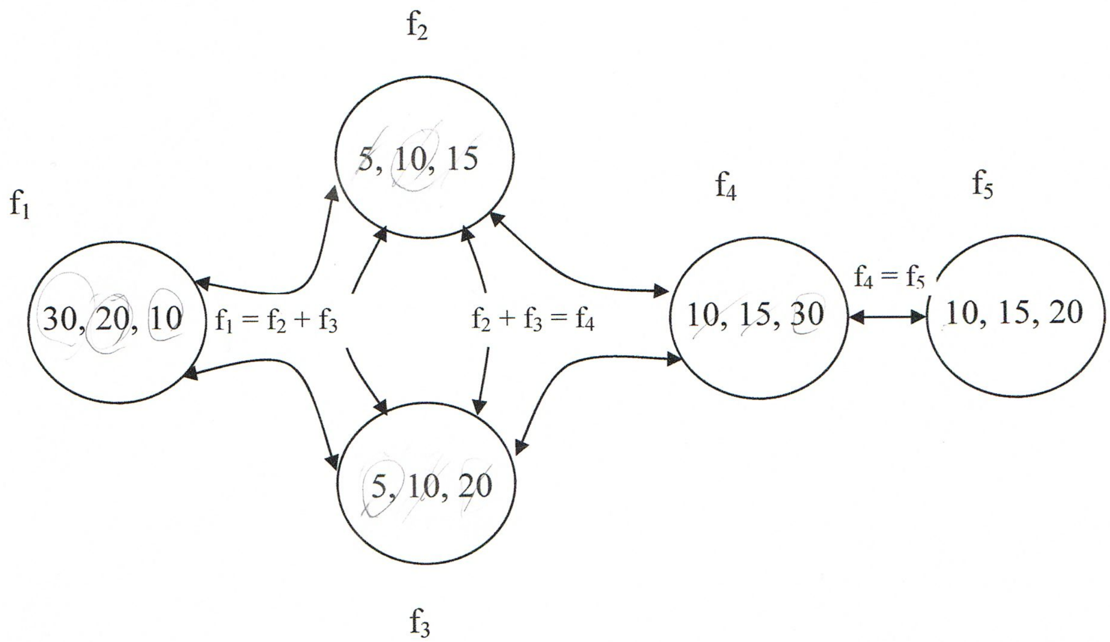
6th 6



Quiz 2
Problem 2: Constraint Propagation (50 points)

Joseph has encountered a network capacity diagram and he is trying to solve for maximum flow in the network.

He creates the following constraint network with the integer domains shown:



Note that there are two constraints among three variables. These are indicated by the curvy lines that come together.

In this constraint network, the variables are the integer values of flow through each node. The constraints between variables are indicated by the curved and straight lines between variables. The domains are the possible values of flow through each node. These are represented by the integer values in each node.

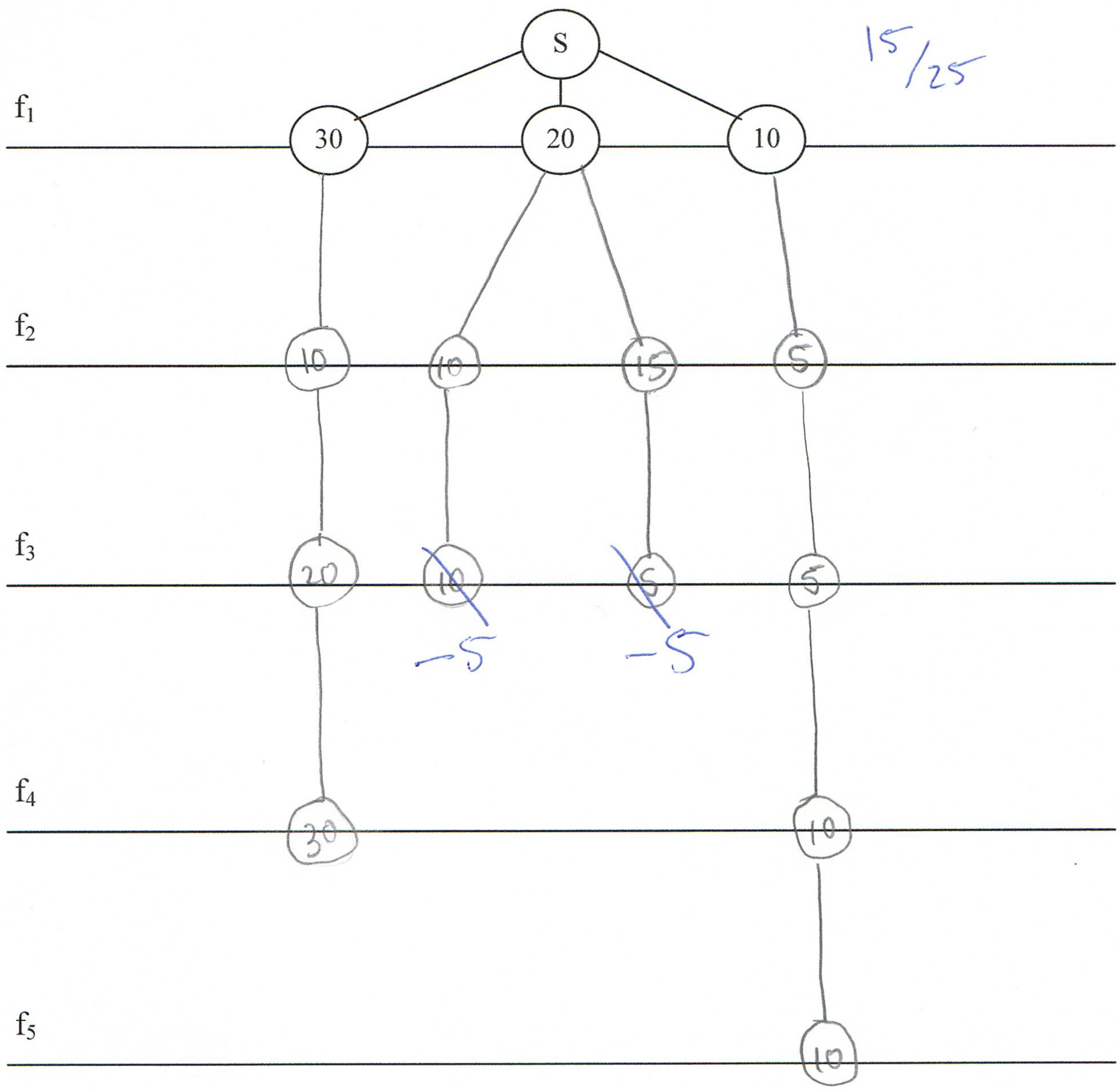
Joseph has a device which measures flow. He obtains an estimate of flow that passes through each node, but is unsure. The possible values are indicated by the domain of each variable.

Part A: (25 points)

Draw a search tree using **backtracking with forward checking** (no propagation beyond neighbors of just-assigned variable).

In addition, for every node in the tree, draw only the valid descendants at that point. Values are to be tried in the same order shown in the nodes on the flow diagram. A portion of the tree is drawn for you.

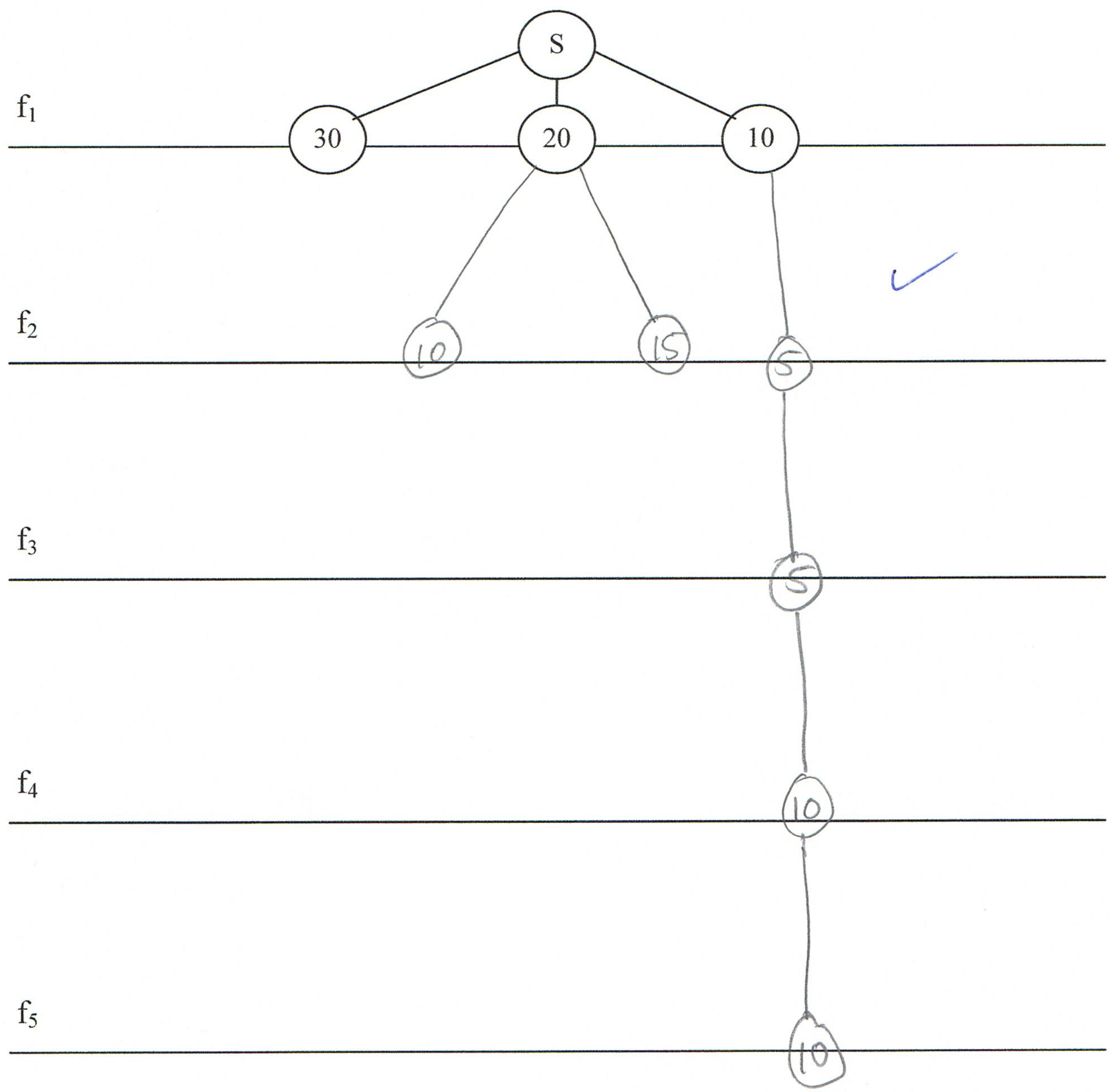
15/25



Part B: (25 points)

Draw the search tree that results from using **backtracking with forward checking AND propagating through domains that are reduced to size 1 [singleton domains].**

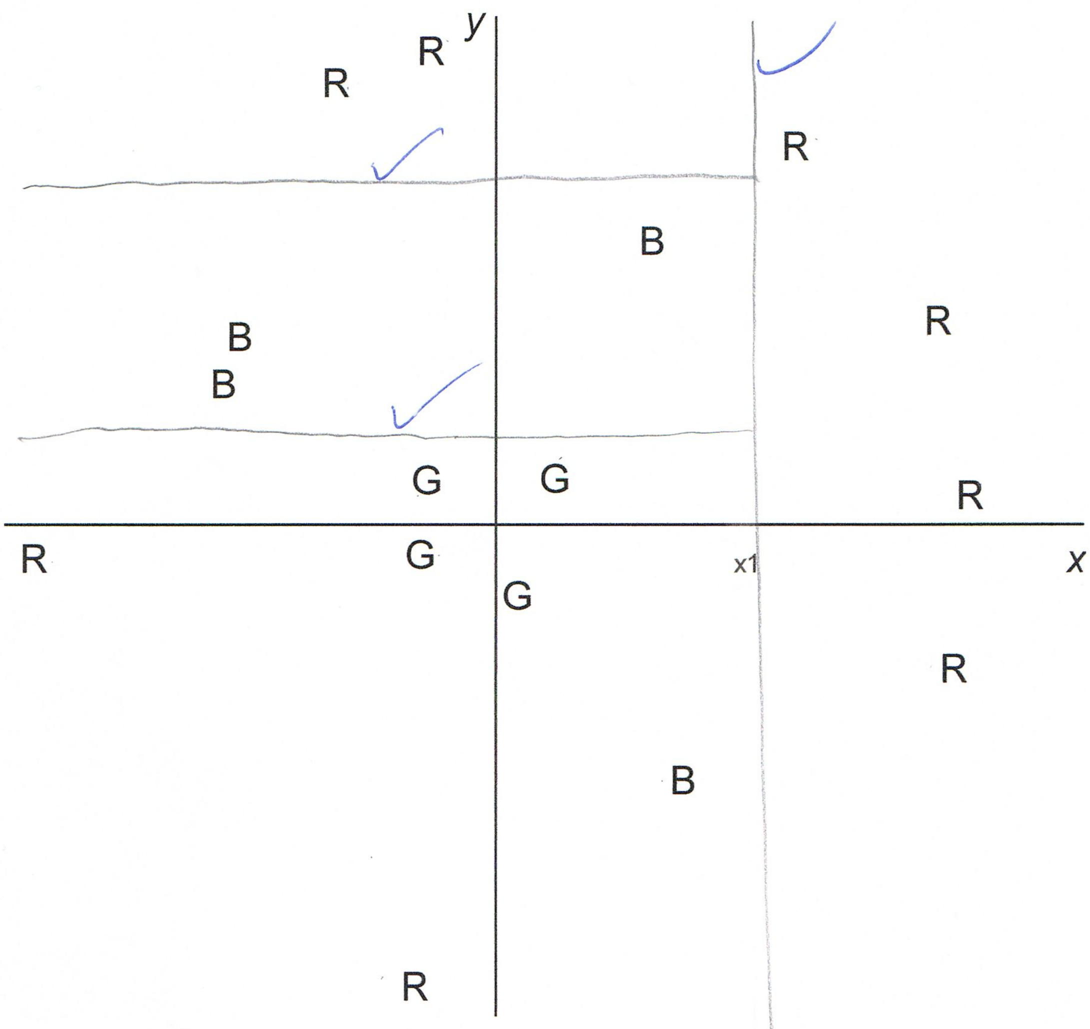
In addition, for every node in the tree, draw only the valid descendants at that point. Values are to be tried in the same order shown in the nodes on the flow diagram.



Quiz 3
Problem 1: Classification (50 points)

Part A: (30 points)

Angelique needs to find a way to identify objects that belong to category R, G, and B using the samples provided in the following diagram.



Part A1 (10 points)

In preparation for placing the first decision boundary, what disorder does Angelique compute for the decision boundary x = x1. You may write your answer in terms of a variable-free expression involving only integers, arithmetic operations, and logarithms.

$(.25 * 0) + (.75 * 1) = .75$ X 0/10

Part A2 (10 points)

Assume that the best decision boundary at any given point in building a classification tree is the boundary that puts the most samples in a homogeneous set. On the diagram on the previous page, draw the first 3 decision boundaries Angelique produces using the identification-tree algorithm. In case two decision boundaries are equally good, use a horizontal rather than a vertical. If there is still a tie, use the smaller threshold.

10/10

Part A3 (10 points)

The **worst possible average disorder** producible by a decision boundary is the same as the disorder of the complete data set handled by the decision boundary. Write the equation of the worst possible decision boundary for this problem.

5/10

$x = 0$ ✓

Write the disorder produced by the **worst possible choice** of a first decision boundary. In this part, we want an exact numerical answer, with no logarithms.

1 X 3/2

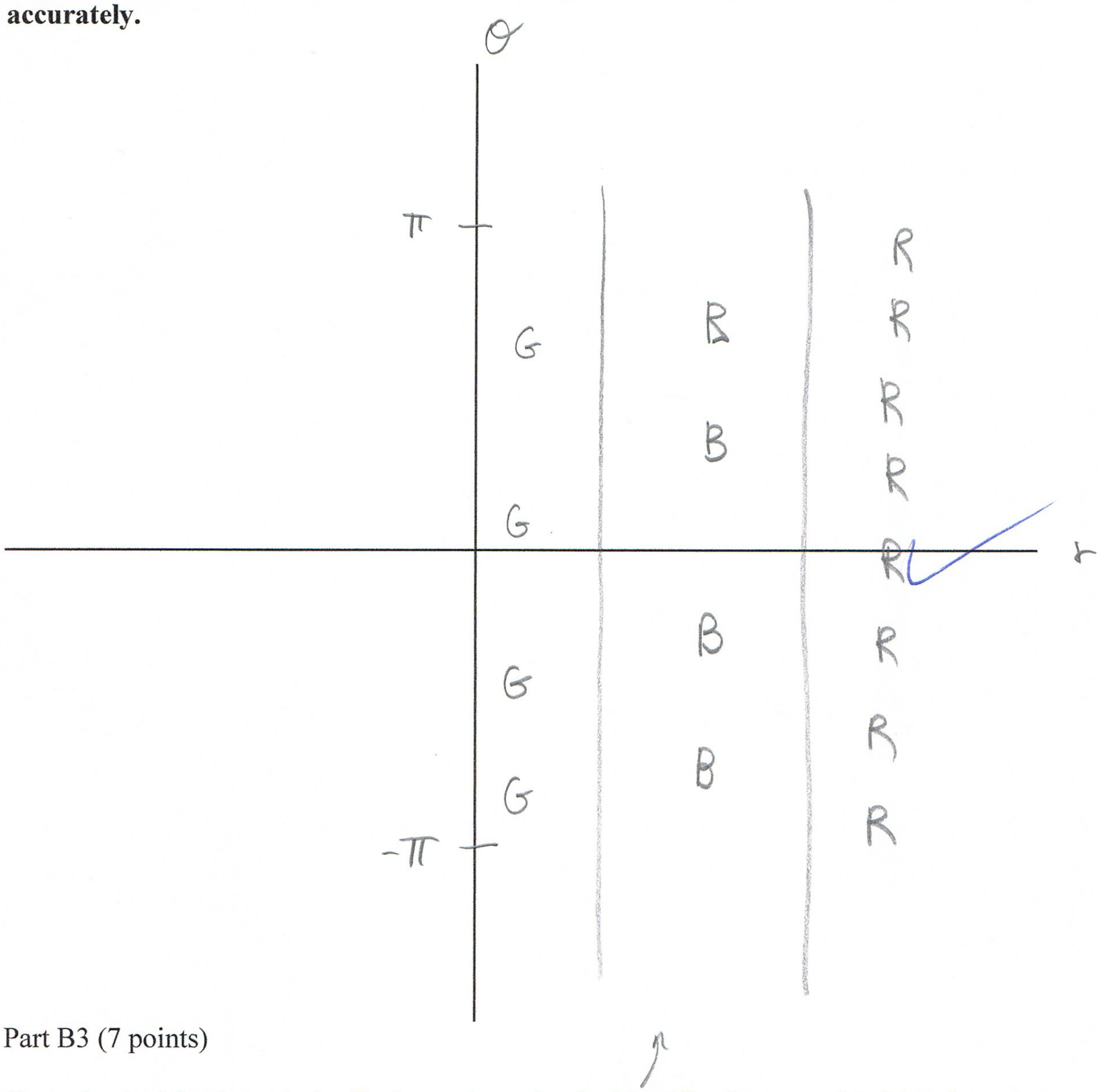
Part B (20 points)

Peak says he thinks the data come from planets of three types in a distant star system. He suggests that it would be interesting to try a transform to ease the problem of placing decision boundaries.

Part B1 (6 points) What is Peak's transform? to polar coordinates

Part B2 (7 points)

Label the axes and **roughly** sketch the data on the following coordinate system. **The way you organize the points must show you have the right idea, but you need not place each point accurately.**

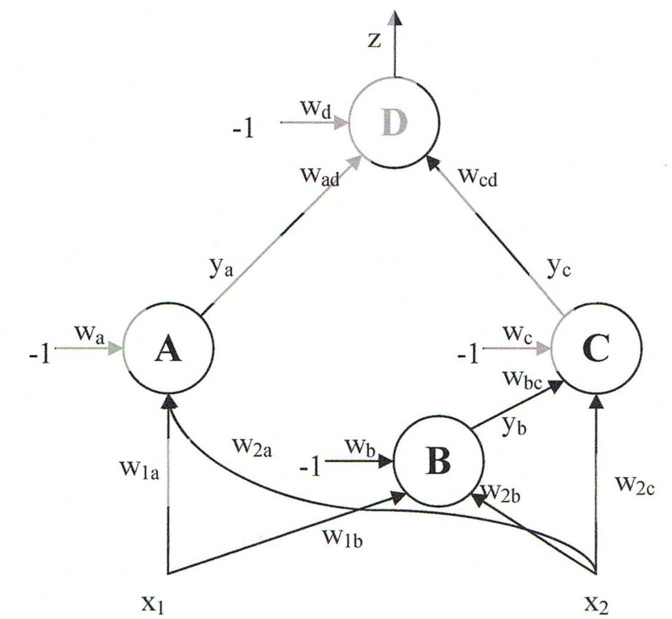


Part B3 (7 points)

Show the decision boundaries Peak produces by the identification-tree algorithm from these samples.

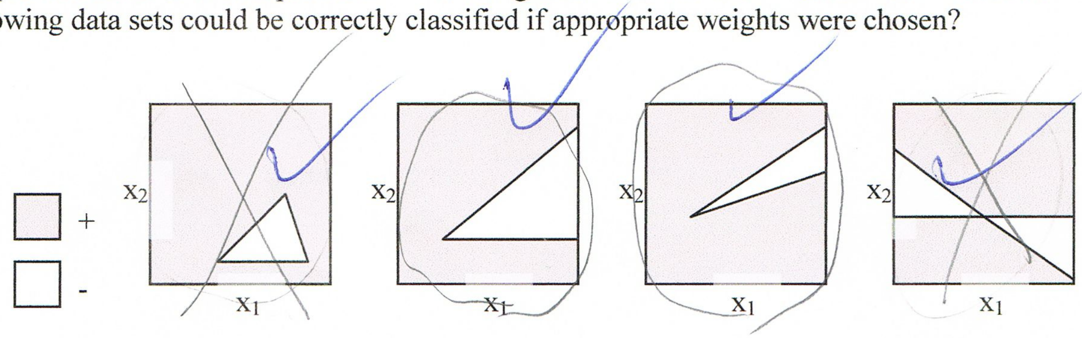
Quiz 3
Problem 2: Neural Nets (50 points)

In the neural network shown in the figure, the outputs of neurons A, C and D are computed by the sigmoid function. The output of neuron B is computed by a linear function, that is, by multiplying the weighted sum of B's inputs by a constant, c. The standard performance function, $P = -\frac{1}{2}(d-z)^2$, is used to evaluate the output on training data.



Part A (12 points)

The output of neuron D is interpreted as '+' if it is greater than 0.5 and '-' otherwise. Which of the following data sets could be correctly classified if appropriate weights were chosen?



Part B (20 points)

After running the first example in your training data through this net, you note that the output, z differs from the desired value d. Write an expression for Δw_{cd}, the amount by which w_{cd} is adjusted during backpropagation of the result. The learning rate is r. Write your expression in terms of r, d, z, c, y_a, y_b, y_c, x_1, x_2, and any weights shown on the diagram. Recall that the sigmoid function $o = s(i)$ is $1/(1+e^{-i})$ and its derivative is $o(1-o)$.

$$\Delta w_{cd} = -r(z-d)\left(\frac{ds}{dz}\right)\left(\frac{\partial z}{\partial w_{cd}}\right)$$

$$= -r(z-d)\underbrace{\left(\frac{1}{1+e^{-z}}\right)\left(1-\frac{1}{1+e^{-z}}\right)}_{z(1-z)} y_c$$

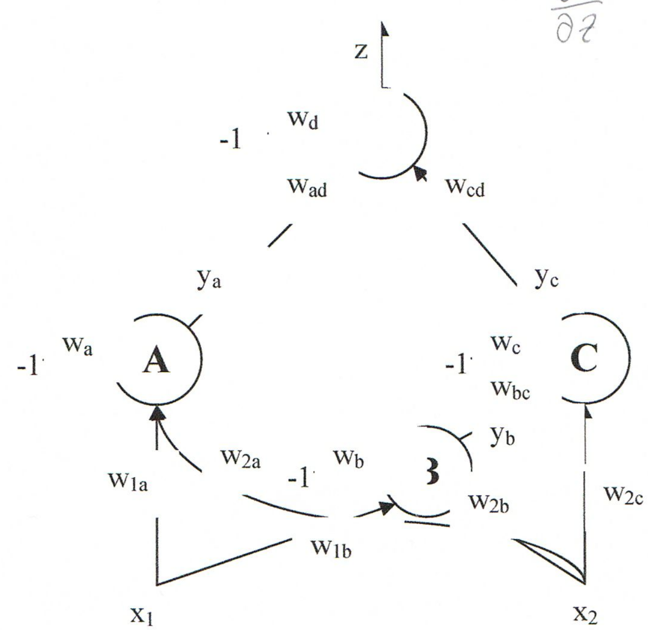
0/10

Write an expression for Δw_{1b}, the amount by which w_{1b} is adjusted during backpropagation of the result. Remember that the output of neuron B is computed by a linear function, that is, by multiplying the weighted sum of B's inputs by a constant, c. That is, no sigmoid function is involved. Write your expression in terms of r, d, z, c, y_a, y_b, y_c, x_1, x_2, and any weights shown on the diagram.

$$\Delta w_{1b} = -r\delta\frac{\partial z}{\partial w_{1b}} = -r(z-d)\underbrace{\left(\frac{1}{1+e^{-z}}\right)\left(1-\frac{1}{1+e^{-z}}\right)}_{\frac{\partial s}{\partial z}} \underbrace{c}_{\frac{\partial s}{\partial z}} \underbrace{(x_1+x_2)}_{\text{input}}$$

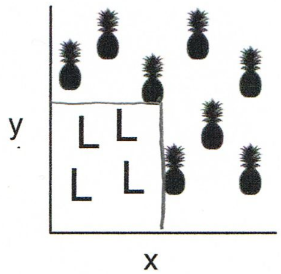
X

0/10

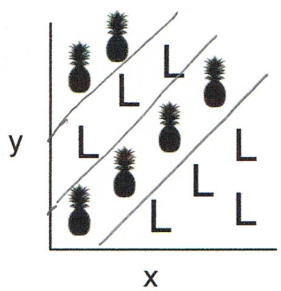


Part C (18 points)

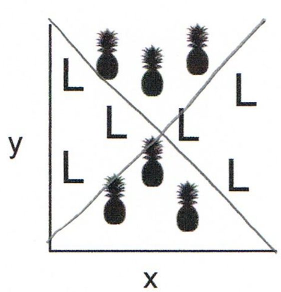
Shown are three possible mappings of training data onto a 2D feature space. For each mapping, draw the minimum number of decision boundaries required to completely classify the training data. In the box next to each mapping, write the letter of the neural network **with the fewest nodes** that could separate the classes. **You may use the same network more than once.**



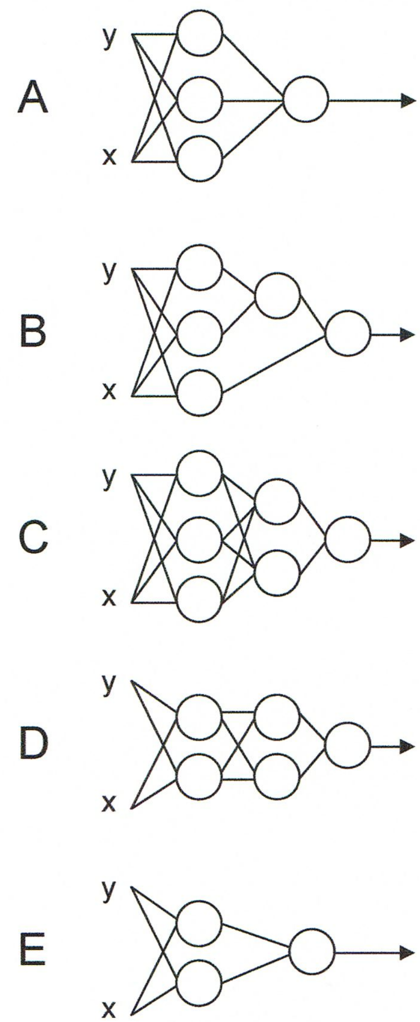
$\boxed{E}$



$\boxed{A}$



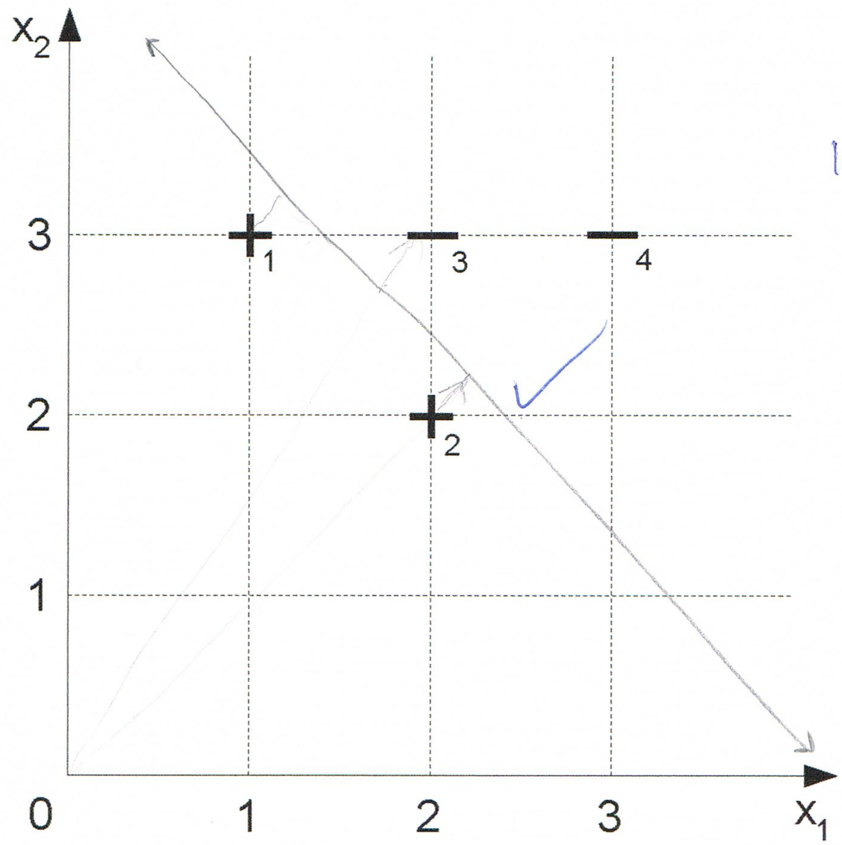
$\boxed{D}$



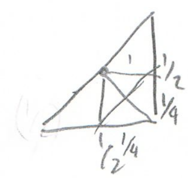
Quiz 4, Problem 1 (50 Points)

Part A (10 points)

Use a linear SVM, with kernel $K(\mathbf{u}, \mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$, to separate the training data points in the sample space provided below. Sketch the decision boundary in the diagram.



10/10



$(1/4)^2 + (1/4)^2 = \sqrt{1/8}$

$\frac{1}{2\sqrt{2}}$

$[1, 3] + [2, 3] + [2, 2] = (1, 1)$

5 8

Part B (20 points)

What are the parameters of your linear SVM?

w = [1, 1]

α1 = $1/2\sqrt{2}$

α2 = $1/2\sqrt{2}$

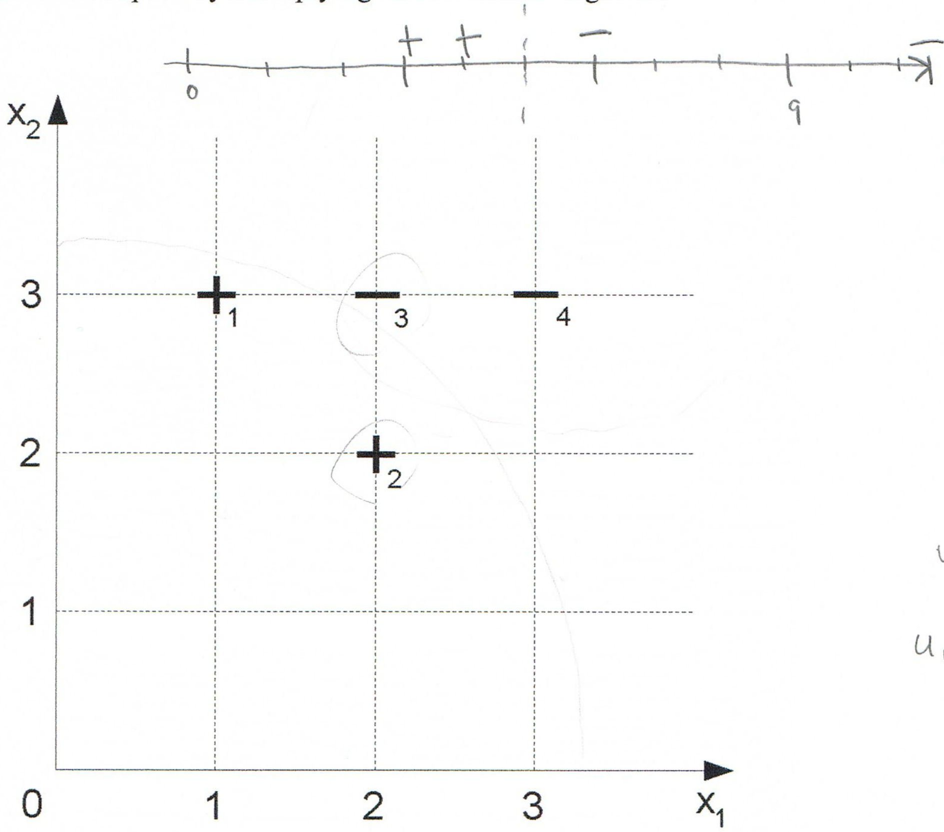
α3 = $1/2\sqrt{2}$

α4 = 0

4/20

Part C (20 points)

Now you are to classify the same data by projecting it into a new space with $\phi(\mathbf{u}) = \langle u_1 \bullet u_2 \rangle$. That is, you project the two-dimensional vector **u** into a one-dimensional vector in a one-dimensional space by multiplying its coordinates together.



$||u|| = \sqrt{u_1^2 + u_2^2}$

$||u|| \, ||v|| =$

$u_1 u_2 v_1 v_2$

$(u_1 v_1 + u_2 v_2)^2$

$u_1^2 v_1^2 + 2 u_1 v_1 u_2 v_2 + u_2^2 v_2^2$

$u_1 v_1 + u_2 v_2 = a$

$u_1 u_2 + v_1 v_2$

What is the kernel function for this transformation?

K(u, v) = $u_1 u_2 v_1 v_2$ ✓

$K(u \circ v)$

What are the parameters of the SVM produced in the new space?

9/20

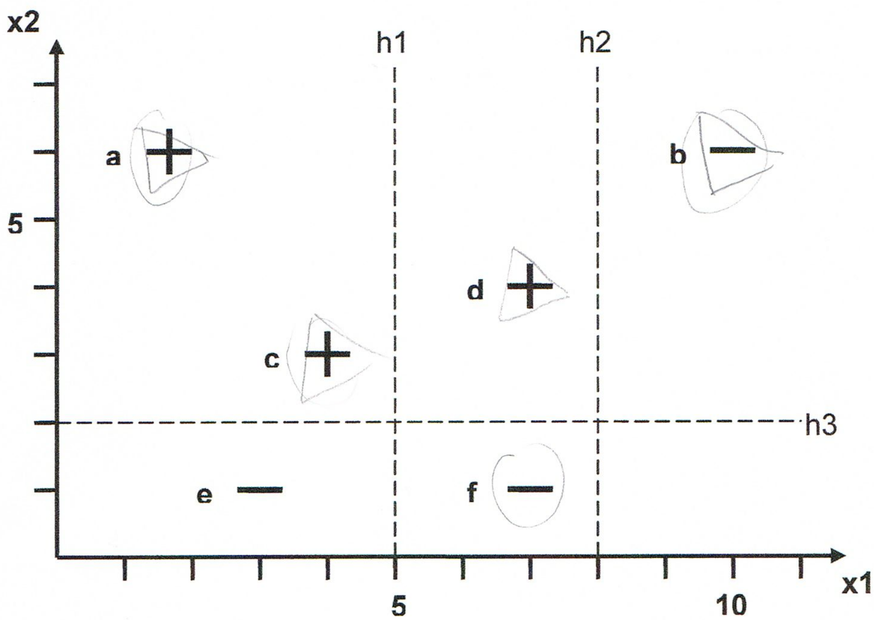
w = 5 ✗

α1 = 0 ✓ α2 = 1 ✗

α3 = 1 ✗ α4 = 0 ✓

Quiz 4
Problem 2: Boosting (50 points)

You would like to build a classifier on the training data below, but the linear classifiers you've tried up to this point have been performing poorly. Instead of starting over from scratch with a different type of classifier, you decide to feed your previous attempts into the AdaBoost algorithm and see what happens.



You are using a simple weak learning algorithm that considers three possible classifiers h1, h2, and h3, as shown in the table below. Each classifier outputs + if its test condition is true, and – if the test condition is false.

Classifier	Test Condition
h1	x1 < 5
h2	x1 < 8
h3	x2 > 2

R3:
5/16
1/2
5/16

h1: 4/2
h2: 4/2
h3: 5/1

Part A (42 Points)

Simulate the AdaBoost algorithm for 3 rounds by filling in the table below. **If at any step two classifiers have the same error rate, the algorithm breaks ties by choosing the one appearing first in the table (h1, then h2, then h3).**

Round	**1**	**2**	**3**
D(a)	1 / 6	$\frac{1}{12} \cdot \frac{6}{5} = \frac{1}{10}$	$\frac{1}{20} \cdot \frac{5}{4} = \frac{1}{16}$
D(b)	1 / 6	$\frac{1}{12} \cdot 6 = \frac{1}{2}$	$\frac{1}{4} \cdot \frac{5}{4} = \frac{5}{16}$
D(c)	1 / 6	$\frac{1}{10}$	$\frac{1}{16}$
D(d)	1 / 6	$\frac{1}{10}$	$\frac{1}{4}$
D(e)	1 / 6	$\frac{1}{10}$	$\frac{1}{20} \cdot 5 = \frac{1}{4}$
D(f)	1 / 6	$\frac{1}{10}$	$\frac{1}{16}$
Weak classifier h_t	h_3: $x_2 > 2$	h_1: $x_1 < 5$	$\cancel{h_3}$: $x_2 > 2$
Error rate ε_t	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{5}{16}$
Weight a'_t	$\ln 5$	$\ln 5$	$\cancel{\ln(16/5)}$ $\ln 11/5$

38
42

incorrect correct

$$\frac{D_{t-1}}{2} \cdot \frac{1}{\varepsilon} \qquad \frac{D_t}{2} \cdot \frac{1}{1-\varepsilon}$$

Part B (8 Points)

Cross out the training examples that the new AdaBoost classifier (after 3 rounds of training) gets **wrong**. **Circle** the examples the classifier gets **right**. You may find the provided table of logarithms useful if you don't have a calculator.

a ~~b~~ c d ~~e~~ f

CARRY FWD ERROR:

8/8

Table of natural logarithms:

ln(1)	0
ln(2)	0.69
ln(3)	1.10
ln(4)	1.39
ln(5)	1.61
ln(6)	1.79
ln(7)	1.95
ln(8)	2.08
ln(9)	2.20
ln(10)	2.30
ln(11)	2.40
ln(12)	2.48
ln(a / b)	ln(a) - ln(b)

Quiz 5, Problem 1 (44 points)

Rick thinks he can use the heuristics that emerged from research on arches (require link, forbid link—climb tree—extend set—drop link) to create a program that learns to recognize the amino acid sequences in the reverse transcriptase protein of various strains of AIDS (true story, Rick got his PhD some years ago and he is now a senior professor at UC Irvine). Suppose you are transported back in time to when Rick was a TA in 6.034, just starting his research. To clarify his thinking, he creates a simplified kind of biology, and produces problems you are to solve.

1. There are five amino acids: a, b, c, d, and e.

2. Some sequences of amino acids form beta strands, denoted as B.

3. Amino acids have a size property. They can be small, medium, or large. Amino acids a and b are small, c is medium, d and e are large.

4. Amino acids are classified as hydrophilic, neutral, or hydrophobic. Amino acids b and c are hydrophilic. Amino acids a and d are neutral. Amino acid e is hydrophobic.

5. When there are two way to use a positive or negative example, your system prefers model modifications that add *must* relations to modifications that add *must-not* relations.

A seed sequence found in a protein known to be an AIDS reverse transcriptase is as follows:

B—a—c—B—d—b—e

The available training samples are:

44/44

1. B—a—c—B—c—B—a
2. a—a—c—B—d—b—e
3. B—a—c—B—c—b—e
4. B—a—c—B—b—b—e
5. B—a—c—B—a—b—e
6. B—a—c—B—b—b—b
7. a—a—c—B—e—b—e
8. B—a—b—B—d—b—e
9. B—a—b—B—b—b—e
10. B—a—c—B—a—b—B

Part A

Circle the protein sequence you would you use to teach your program that a beta strand must be in the first position.

1 2 3 4 5 6 7 8 9 10

What arch learning heuristic is involved? require-link

Is the sample you selected an example or a near miss? near miss

Part B

Suppose you start over, ignoring Part A. Circle the protein sequence you would use to teach your program that the amino acid in the third position may be any hydrophilic amino acid.

1 2 3 4 5 6 7 8 9 10

What arch learning heuristic is involved? climb-tree

Is the sample you selected an example or a near miss? example

Part C

Suppose you start over, ignoring Part A and Part B. Circle the protein sequence you would use to teach your program that the amino acid in the fifth position may be either small or large.

1 2 3 4 5 6 7 8 9 10

What arch learning heuristic is involved? extend set

Is the sample you selected an example or a near miss? example

Part D

Circle the protein sequence you would use in combination with the sequence you selected in Part C to teach your program that the amino acid in the fifth position may be any size.

1 2 3 4 5 6 7 8 9 10

What arch learning heuristic is involved? drop link

Is the sample you selected an example or a near miss? example

Quiz 5, Problem 2 (20 points)

Circle the best answer for each of the following question. There is no penalty for wrong answers, so it pays to guess in the absence of knowledge.

Yip and Sussman set out to develop a theory of how:

1. Syntactic rules can be learned from examples.
2. Phonological rules can be learned from examples.
3. Distinctive features govern muscle relaxation and contractions.
4. Structures such as arches can be learned from examples.
5. None of the above.

The Yip and Sussman learning algorithm

1. Generalizes starting with a seed pattern.
2. Specializes starting with a seed pattern.
3. Generalizes and specializes starting with a seed pattern.
4. Does not use a seed pattern.
5. None of the above.

Yip and Sussman believed their algorithm worked because

1. The problem involved high-dimensional sparsely populated spaces
2. The problem involved low-dimensional densely populated spaces
3. The problem involved uniformly populated spaces
4. The problem involved binary features
5. None of the above

9/20

Yip and Sussman could not have succeeded without

1. Beam search
2. A* search
3. Bidirectional search
4. Left to right search
5. None of the above

Yip and Sussman's engineering model of language included

1. Multiway constraint boxes
2. Fourier transforms
3. Circuit models
4. Hash coding
5. None of the above

1/4

Quiz 5, Problem 3 (20 points)

Disorientation experiments with rats, children, and adults demonstrate

1. Rats cannot distinguish the short side of a rectangular room from the long side
2. Children become confused if each wall of a room has a different color
3. Rats cannot combine information about geometry with information about color
4. Children and adults cannot reorient themselves as well if they are chewing gum
5. None of the above

Chomsky believes the distinguishing characteristic of human intelligence is

1. A doubling of short term memory relative to other primates
2. An ability to combine concepts into new concepts without limit
3. An ability to name concepts
4. An ability to form visual images of concepts delivered through language
5. None of the above

Brooks subsumption architecture is best described as

1. A way of building robots using constraint propagation
2. An approach to natural language understanding based on statistics
3. An extension of the 6.001 abstraction idea to a higher level
4. An approach to problem solving based on choosing operators that reduce differences
5 None of the above

Winston believes the best way to develop an account of human intelligence is

1. To focus on collections of rule-based systems interacting through short-term memory
2. To focus on creating robots that interact with real, rather than simulated worlds
3. To understand tightly coupled loops connecting language with imagined visual scenes
4. To collect commonsense knowledge from volunteers on the world-wide web
5. To study differences in the chimp and human genomes

Representations are important because

1. They can provide a framework for recording real and surrogate experience
2. They can provide a substrate for building models
3. They can make constraints explicit
4. All of the above
5. None of the above

12/20

Quiz 5, Problem 4 (16 points)

Coen's method seems capable of

1. Partitioning words into clusters corresponding to phonemes.
2. Partitioning sensory input into clusters corresponding to vowels.
3. Partitioning multiple modes of labeled sensory input into clusters
4. Partitioning multiple modes of unlabeled sensory input into clusters
5. None of the above

The powerful idea in Coen's work is

1. Observation that constraints on muscle movement limit the number of vowels produced
2. Observation that constraints on cochlear geometry limit the number of vowels recognizable
3. Observation that proximity in one space can suggest clustering in another space
4. Observation that simultaneous proximity in two spaces is needed for clustering
5. None of the above

Coen believes in cross-modal cooperation of sensory systems

1. As late as possible, after each system has done all it can independently
2. As early as possible, close to the physical sensors
3. From top to bottom, so as to support contextual processing
4. From bottom to top, so as to prevent hallucination
5. None of the above

4/16

Coen's algorithm

1. Starts with large clusters, which it splits
2. Starts with small clusters, which it merges
3. Starts with large clusters in one space and small clusters in the other
4. Starts with either large clusters or small clusters depending on the problem
5. None of the above