6.034 Final Examination
December 15, 2008

Name	
EMail	

Quiz number	Maximum	Score	Grader
1	100		
2	100		
3	100		
4	100		
5	100		

There are 33 pages in this final, including this one. Additional pages of tear-off sheets are provided at the end with duplicate drawings and data. As always, open book, open notes, open just about everything.

Quiz 1, Question 1, Rules (50 points)

Before swimming in an unknown river, you want to figure out which animals are dangerous. You have a set of rules and assertions, given below.

Rules:

```
P0:    IF( '(?x) is a mammal',
           THEN( '(?x) is not a crocodile' ) )
P1:    IF( AND( '(?x) is not a crocodile',
               '(?x) lives underwater' ),
           THEN( '(?x) is a manatee'
P2:    IF( AND( '(?x) is a mammal',
               '(?x) lives underwater' ),
           THEN( '(?x) is a hippo' ) )
P3:    IF( OR( '(?x) is a crocodile',
              '(?x) is a hippo' ),
           THEN( '(?x) is dangerous' ) )
P4:    IF( '(?x) is not a crocodile',
           THEN( '(?x) is safe' ) )
```

Assertions:

```
A0:    ('Spike is a mammal')
A1:    ('Fido is a mammal')
A2:    ('Fido lives underwater')
A3:    ('Rover is a crocodile')
```

Part A: Forward Chaining (30 points)

You may make the following assumptions about forward chaining:

- Assume rule-ordering conflict resolution
- New assertions are added to the bottom of the dataset
- If a particular rule matches assertions in the dataset 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 A1 and A2, the match with A1 is considered first.

Run forward chaining on the rules and assertions provided. For the first two iterations, fill out the table below, noting the rules matched, fired, and new assertions added to the data set.

	Matched	Fired	New Assertions Added to Data Set
1			
2			

Which animals (Fido, Spike, Rover) are determined to be dangerous?

Which animals (Fido, Spike, Rover) are determined to be safe?

Would changing the order of the rules affect the final decision of which animals are dangerous or safe?

Part B: Backwards Chaining (20 points)

Make the following assumptions about backwards chaining:

- When working on a hypothesis, the backward chainer tries to find a matching assertion in the dataset. If no matching assertion is found, the backward chainer tries to find a rule with a matching consequent. In case none are found, then the backward chainer assumes the hypothesis is false.
- The backward chainer never alters the dataset, 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.

Evaluate the hypothesis 'Spike is dangerous' using backwards chaining. Draw a goal tree in the space below.

Is Spike dangerous?

Quiz 1, Question 2, Search (50 points)

Part A

Your 6.034 Tas have invented a new game based on baseball called Blurnsball. As many people know, baseball is incredibly boring, so to jazz it up, they included several rules variants, including a variant for running the bases. In Blurnsball, it is legal to run across the pitcher's mound and to the opposite side (thus making it legal to run from 1st base to the pitcher's mound to 3rd base or from 2nd base to the pitcher's mound to home). They have hired you as a consultant to use 6.034 Search techniques in order to analyze the new rules variant. See the graph below for a diagram of the new set-up.

Break all ties in lexicographic order, treating the path from start to finish as a character string. Thus, S-B1-B2-B3 comes before S-B1-P13-B3.
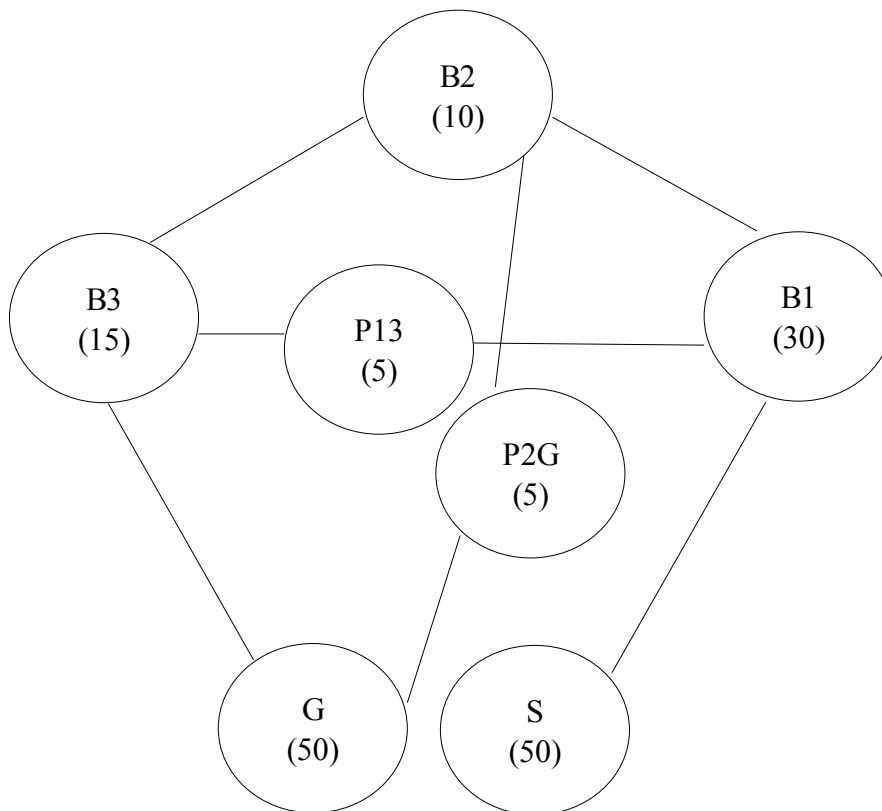
Part A1 (10 points)

The TAs are impatient and demand that you perform a hill-climbing search using the heuristic distances to the goal, provided in parentheses in the diagram.

What path do you find from the starting node S to the goal node G? Do not test a path to see if it reaches the goal until that path reaches the front of the search queue.

How many paths do you extend? Be sure to count the path that contains just S.
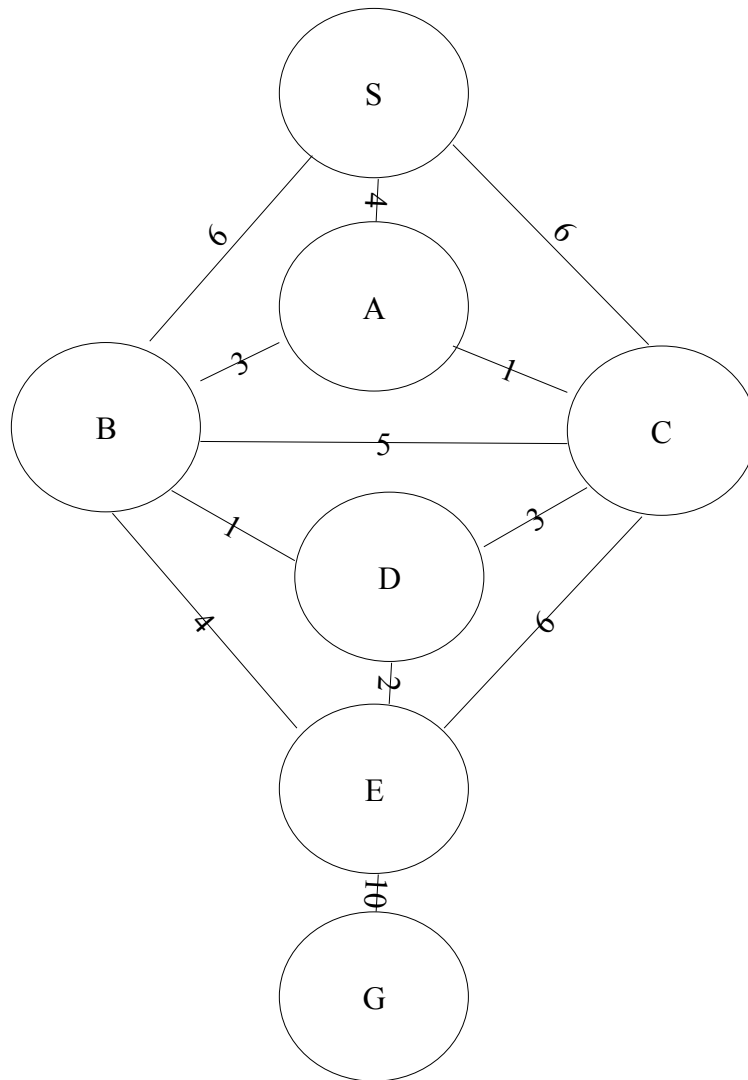
Part A2 (10 points)

Something seems funky, so Sam suggests a depth-first search.

What path do you find? Do not test a path to see if it reaches the goal until that path reaches the front of the search queue.

How many paths do you extend? Be sure to count the path that contains just S.

Part A3 (10 points)

Mark is bored by your answer and decides that you should use Beam Search instead to keep the options from getting too large. Perform beam search with a beam width of 2. Sort after each extension and keep the two paths with the best heuristic distance.

What path do you find this time? Be sure to use the tie breaker on the previous page in case ties need breaking.

How many paths do you extend? Be sure to count the path that contains just S.

Part B

Alex has been chatting online for weeks with his new online girlfriend, Eliza. Stephanie and Maria are skeptical of this "Eliza" and insist that Alex meet with her in person, so when Alex asks Eliza out, Stephanie and Maria decide to come along to make sure the girlfriend is legit. They have to travel along the following streets, from S to G, and Alex wants to make sure he's on time to meet Eliza, so he insists that they take the shortest path.

Part B1 (10 points)

Using Branch and Bound with an Extended List, what is the final path from S to G? Be sure to show your goal tree for partial credit.

What nodes are in your extended list?

Part B2 (10 points)

Stephanie and Maria each suggest a heuristic:

Stephanie's heuristic:
S—25, A—16, B—3, C—15, D—12, E—0, G—0

Maria's heuristic:
S—4, A—16, B—16, C—15, D—14, E—10, G—0

Which of the following searches will find the same path you found using branch and bound with an extended list?

Circle those that work; cross out those that do not.

- Branch and bound with Stephanie's heuristic and **no** extended list.

- Branch and bound with Maria's heuristic and **no** extended list.

- Branch and bound with Stephanie's heuristic and an extended list (aka A*)

- Branch and bound with Maria's heuristic and an extended list (aka A*)

Quiz 2, Question 1, Games (50 points)

You are playing a new Sim game called Obamaquest, the Legend of the Lost International Credibility. In this game, you play a charismatic incoming president who must make a choice on various issues in order to save your country. After each of your turns, the outgoing president will attempt to perform the most meddlesome acts possible to make it less likely that you will succeed. You realize quickly that you can model this game using a simple Game Tree from 6.034, as shown below.

Static values are shown underneath leaf nodes. Ignore the numbers in parentheses for now.
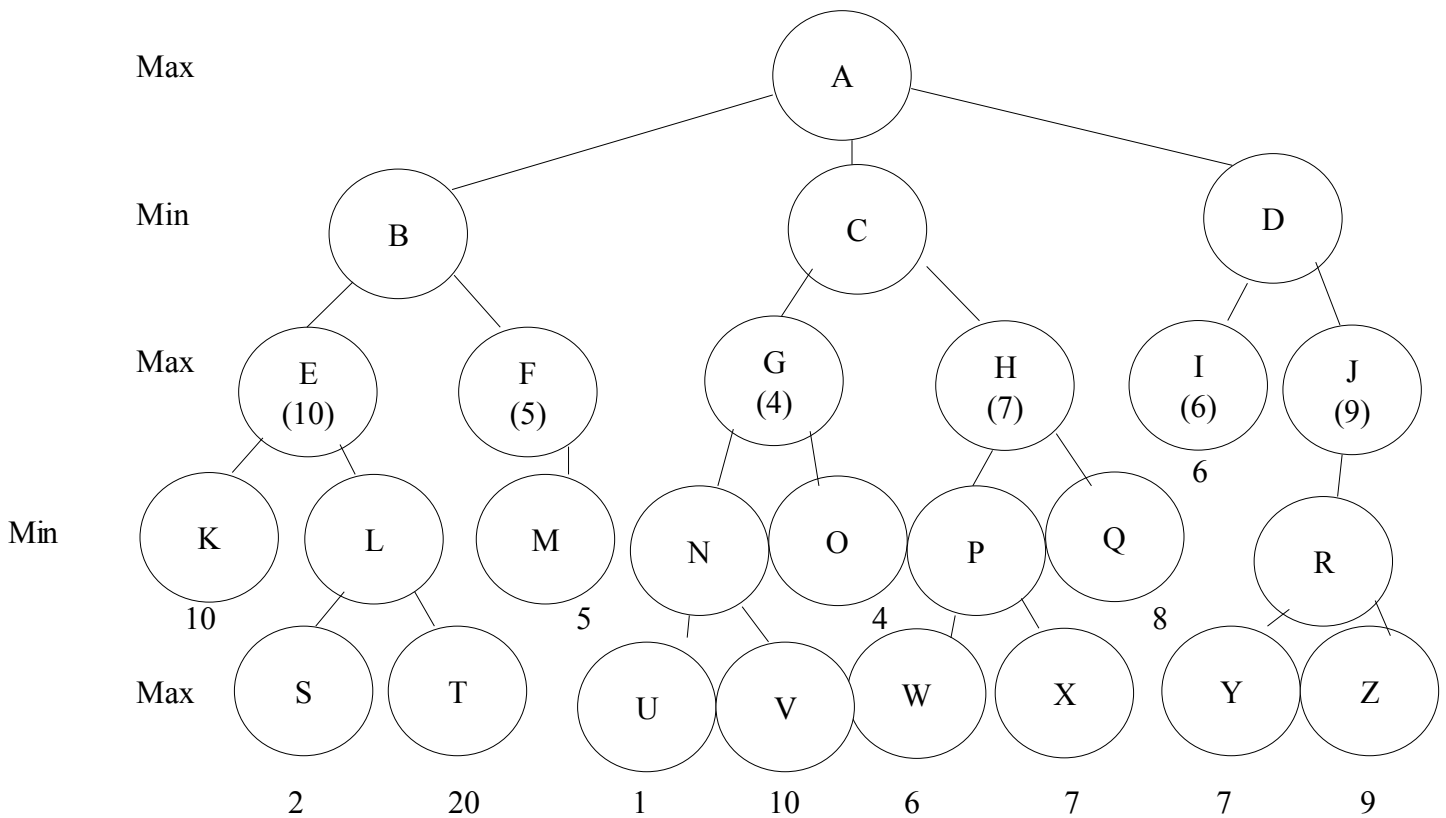
Part A (15 points)

First, you decide to perform a simple minimax algorithm on the tree.

Which direction will the maximizer choose to go at node A?

What is the minimax value of node A?

Which static evaluations did you perform? (write the nodes you statically evaluated, in order).

Part B (25 points)

Minimax was taking too many static evaluations, so you use alpha-beta instead.

This time what direction will the maximizer choose to go at node A?

Which static evaluations did you perform? (write the nodes you statically evaluated, in order).

Part C (10 points)

The end result is still rather depressing in number of required static evaluations, so you decide to perform progressive deepening up to the second level and then reorder the tree to try for a more optimal pruning, using the static values for E,F,G,H,I, and J found in parentheses inside each circle. Draw your new ordering below. **You need not not draw any of the nodes below E, F, G, H, I, and J.**

Quiz 2, Question 2, Constraints (50 points)

Four 6.034 TAs (**Mark, Mike, Rob, Sam**) are trying to write eight questions for a final exam (somewhat like this one):

1. **C**onstraints
2. **O**ptimal Search
3. **G**ames
4. **R**ules
5. **ID**-Trees
6. **N**eural Nets
7. **S**VMs
8. **B**oosting

Some questions are harder to write than others, so they should be distributed about equally. And some problems are so similar to others that for variety, we don't want those written by the same TA. This leads to some constraints:

- TA(Constraints) != TA(Rules)
- TA(Boosting) != TA(ID-Trees)
- TA(Boosting) != TA(SVMs)
- TA(SVMs) != TA(Optimal Search)
- TA(SVMs) != TA(Games)
- TA(SVMs) != TA(Rules)
- TA(SVMs) != TA(Neural Nets)
- TA(Neural Nets) != TA(Games)

Also, there are some demands that have to be honored.
- Mike insists on doing Boosting
- Mike will not do Constraints
- Mark insists on doing Games
- Only Rob and Mike are willing to do SVMs.

Part A (10 points)

Draw lines between variables with a != constraint, and use the TAs' demands to reduce domains by crossing out names. Continue to reduce domains using the constraints while possible.
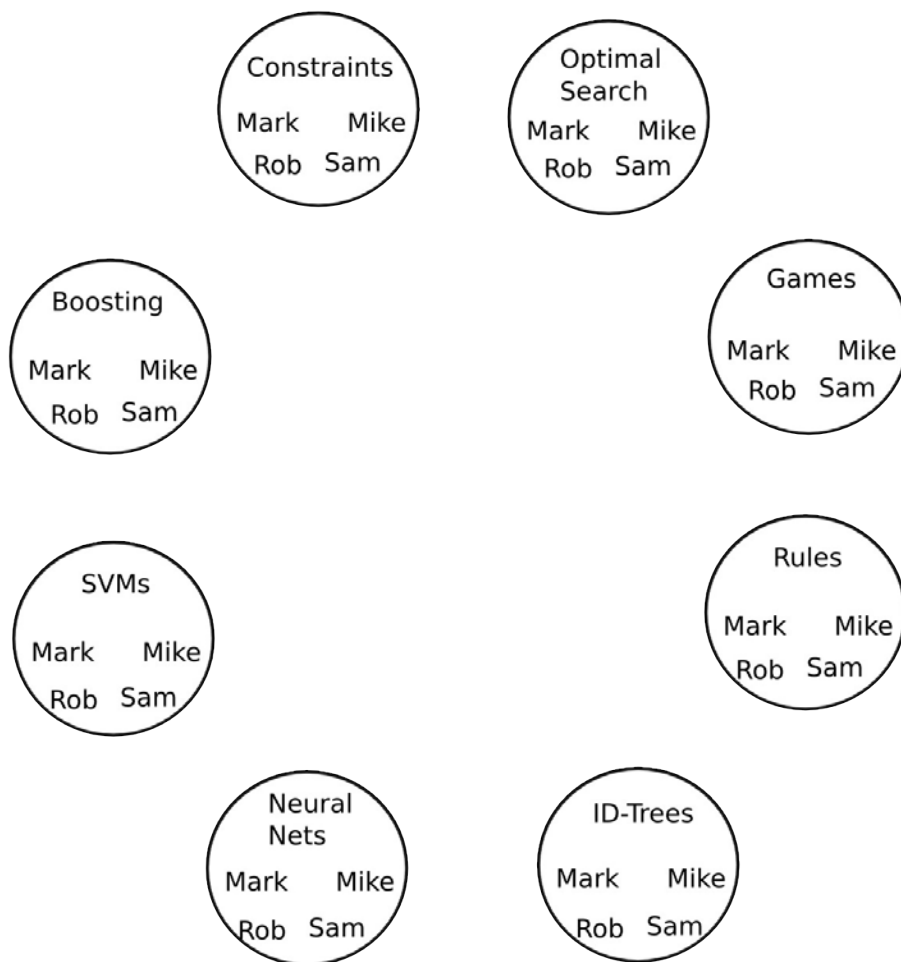
Part B (20 points)

Starting with your domains reduced in Part A, find a solution using depth-first search only, using no constraint propagation , checking constraints at assignments only. Consider TAs in alphabetical order: Mark, Mike, Rob, Sam. Please feel free to abbreviate unambiguously. **Draw your search tree on the next page.**

Constraints	Mark	Rob	Sam
Optimal Search			
Games			
Rules			
ID-Trees			
Neural Nets			
SVMs			
Boosting			

Part C (15 points)

You consider two more advanced constraint propagation algorithms that
- propagate choices to neighbors only
- propagate choices to neighbors and continue through any domains reduced to size one

Do these algorithms find the same result as the DFS? Which domains do these algorithms reduce during the course of their runs? Circle the answers.

Neighbors only

Same result as DFS?			YES		NO			
Domains reduced:	C	O	G	R	I	N	S	B

Neighbors and any domain reduced to size one

Same result as DFS?			YES		NO			
Domains reduced:	C	O	G	R	I	N	S	B

Part D (5 points)

One TA thinks the assignments made in this problem are unfair. Suggest a way to ensure the test questions are more evenly distributed across the four TAs.

Quiz 3, Question 1, NNand ID trees (50 points)

Part A: Nearest Neighbors

On the following graph, draw the decision boundaries produced by 1-Nearest Neighbor. Ignore the letters A and B.
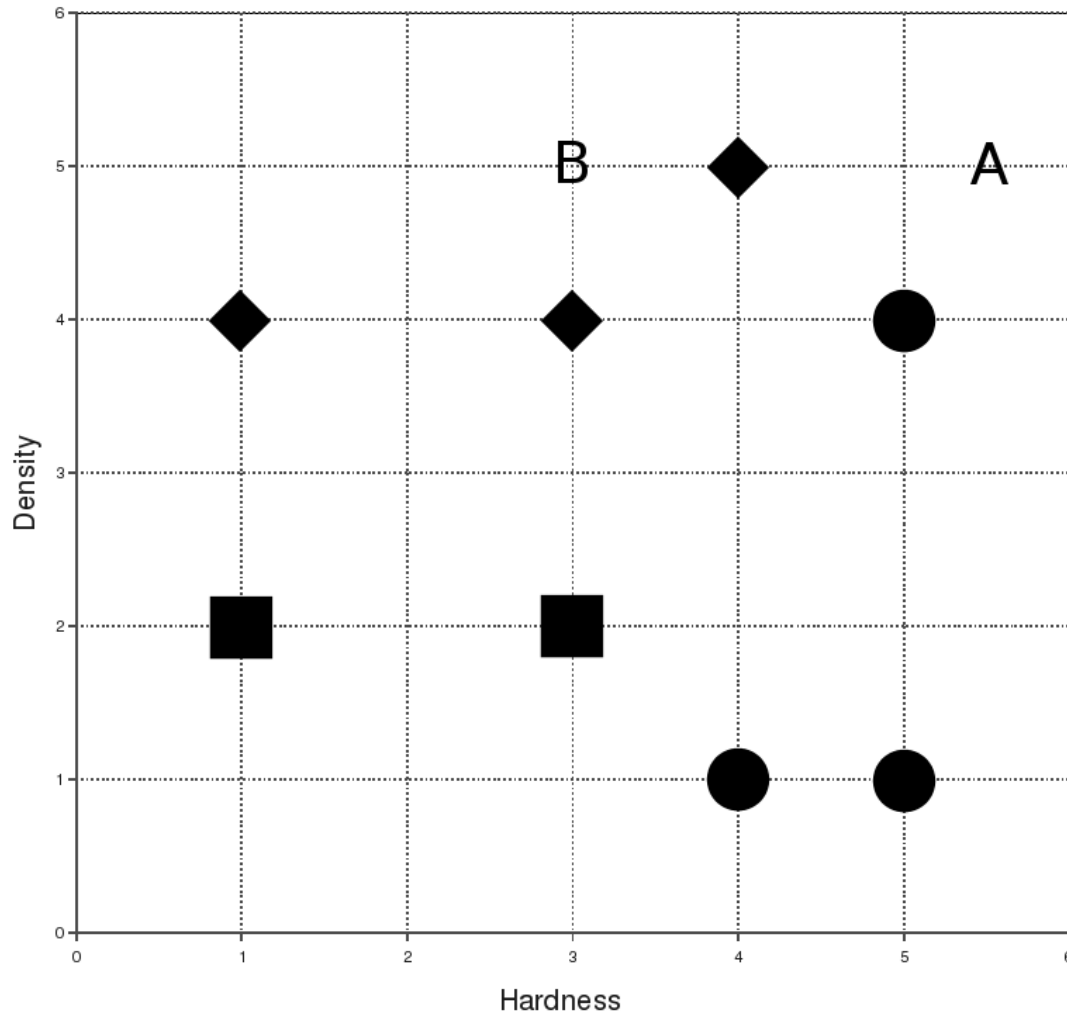
◆ Sedimentary
■ Metamorphic
● Igneous

How is Sample A classified by 1-NN?

By 3-NN?

How is Sample B classified by 1-NN?
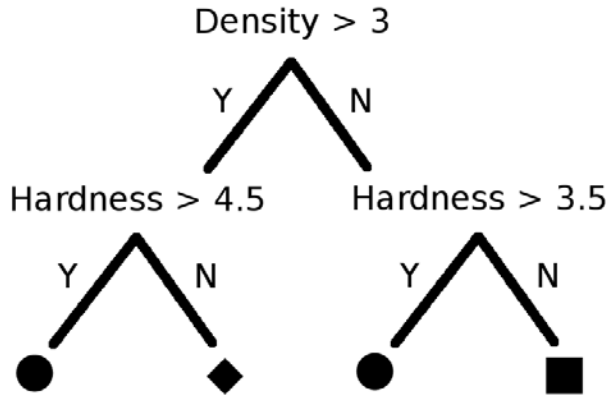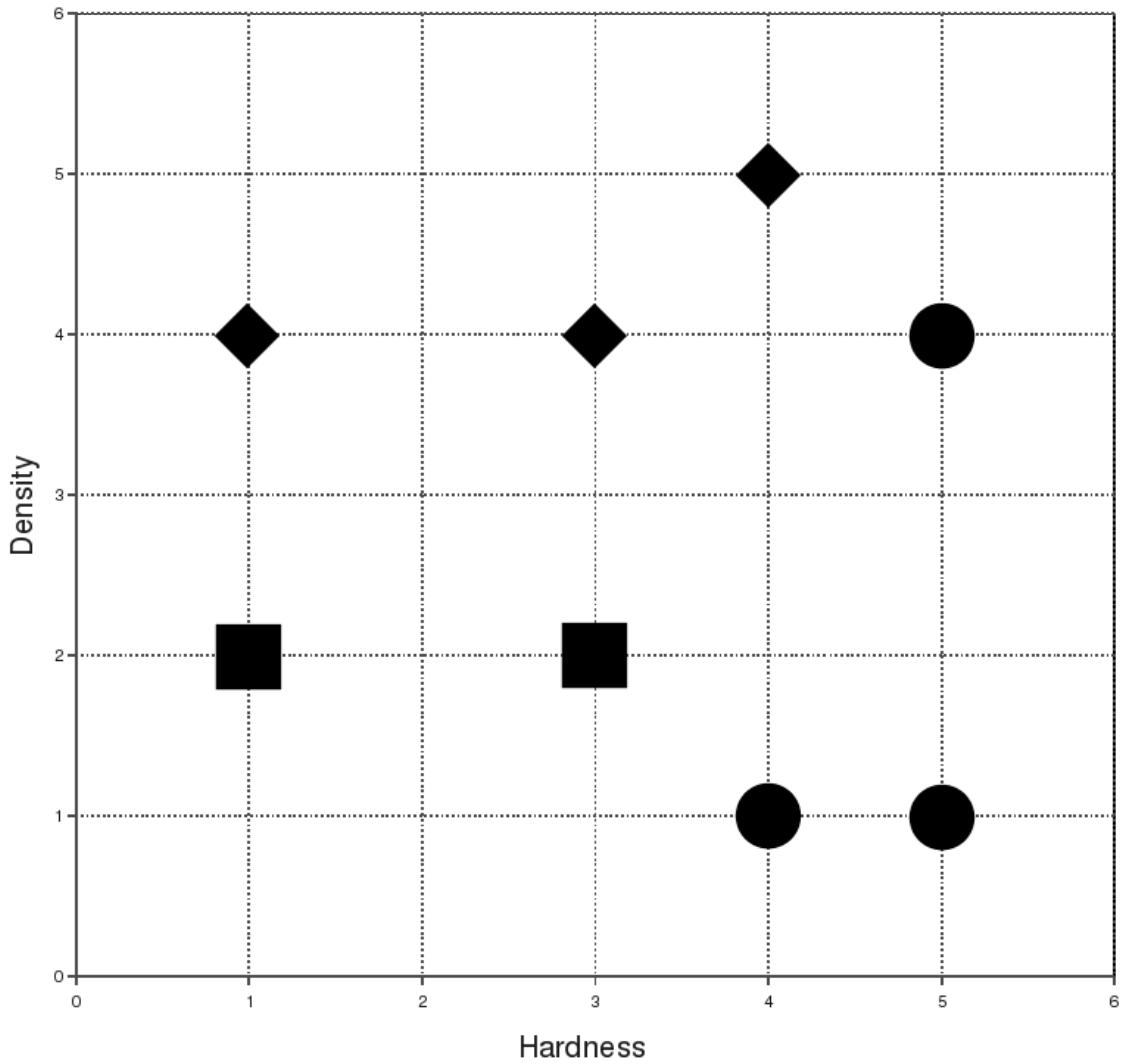
By 3-NN?

Part B: ID Trees

Using the same data as in Part A, calculate the disorder of the following ID Tree tests. Your answers may contain logarithms.

Density > 3 Y N	**Disorder:**
Hardness > 3.5 Y N	**Disorder:**

Suppose you created the following ID Tree.

Density > 3

Y / \ N

Hardness > 4.5   Hardness > 3.5

Y / \ N   Y / \ N

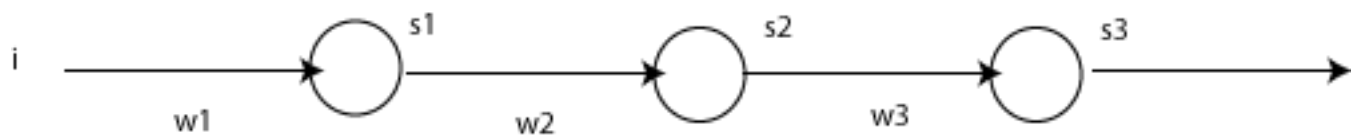● ◆   ● ■

Draw the decision boundaries produced by that ID Tree on the graph below.

Quiz 3, Question 2, Neural Nets (50 points)

Part A (21 points)

Consider the following neural net. Note that all the neurons have sigmoid units, $s(z)=\dfrac{1}{1+e^{-z}}$ and

the performance function is $P=-\dfrac{1}{2}(s_3-d)^2$

Note that the input to the net is i and the outputs of the sigmoid units are $s_1, s_2,$ and $s_3.$

In terms of $i, w_1, w_2, w_3, i, s_1, s_2, s_3,$ and d calculate the following partial derivatives:

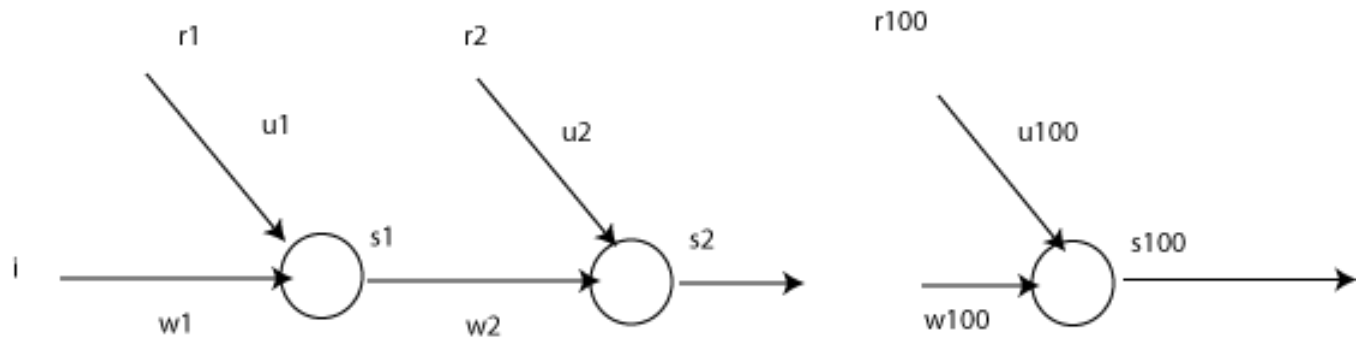$$\frac{\partial P}{\partial w_3}=$$

$$\frac{\partial P}{\partial w_2}=$$

$$\frac{\partial P}{\partial w_1}=$$

Part B (24 points)

Consider the following neural net. There are 100 neurons. All neurons have sigmoid units, $s(z)=\dfrac{1}{1+e^{-z}}$ and the performance function is $P=-\dfrac{1}{2}(s_{100}-d)^2$

The inputs to the network are i and r_i. The outputs of the sigmoid units in the network are s_i

At a certain time, t, $i=y$ and

$$r_1=r_2=\cdots=r_{100}=\rho$$

$$s_1=s_2=\cdots=s_{100}=\sigma$$

$$u_1=u_2=\cdots=u_{100}=1$$

$$w_1=w_2=\cdots=w_{100}=\omega$$

Calculate $\dfrac{\partial P}{\partial w_1}$ in terms of y ρ σ ω and d .

$$\frac{\partial P}{\partial w_1}=$$

Part C (5 points)

Finally, exhibit an equation that relates σ, ρ and ω.
**Do not attempt to solve the equation.**

$$\sigma =$$

Quiz 4, Question 1, Support Vectors (50 points)

Miriam has a lot of finals. She wants to know how best to use her time, so naturally she collects some data to train a support vector machine.

She finds that nine of her friends have already taken the class that her next final is in, so she asks them how long they spent reviewing for the class's final exam, how many hours of sleep they got the night before the final, and what grade they got (A, B, or C). The results appear on this graph:
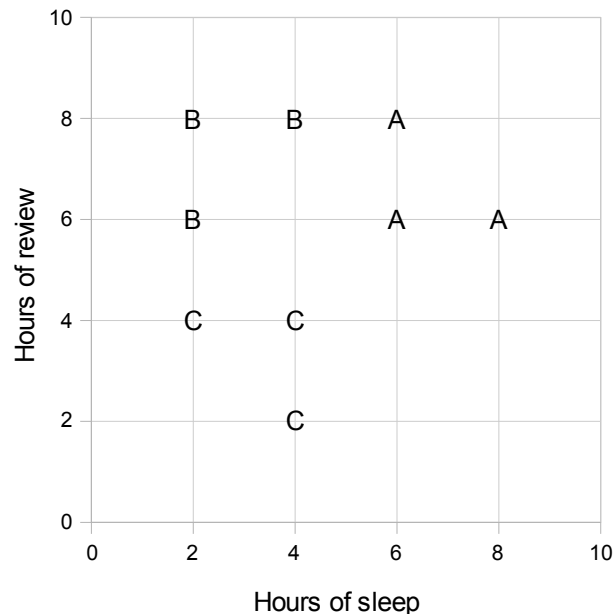

The problem she runs into is that SVMs can typically only distinguish two classes, and her data contains three classes of grades. However, she finds a way to make a three-class SVM:

- Divide the problem among three ordinary, two-class SVM classifiers, called $h_A(\mathbf{x})$, $h_B(\mathbf{x})$, and $h_C(\mathbf{x})$.

- Each SVM treats one of the classes (A, B, and C respectively) as +, and treats the other two classes as −.

- The final classifier outputs A, B, or C based on which of the three sub-classifiers outputs the highest value. For example, if h_A outputs -5, h_B outputs -3, and h_C outputs -2, the overall result should be C. If h_A outputs 2, h_B outputs -1, and h_C outputs -3, the overall result should be A.

Part A (30 points)

The three SVM classifiers are linear, so their output will be defined by the equation $h = \mathbf{w} \cdot \mathbf{x} + b$. On the next three graphs:
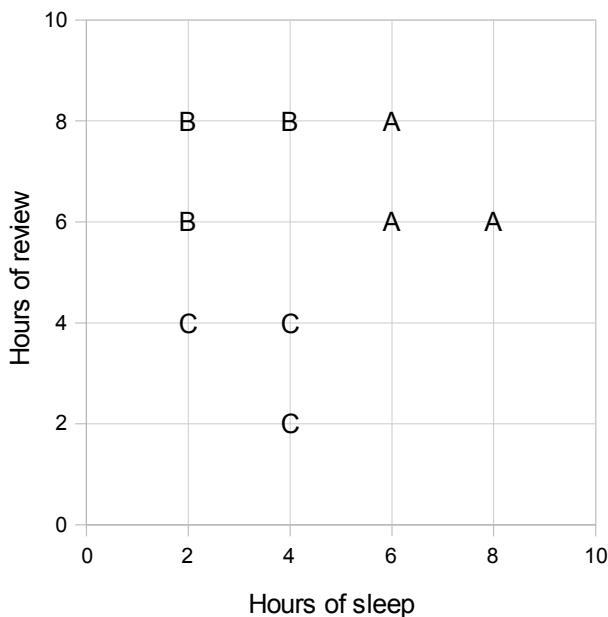
•Draw the "street" that separates the data by drawing a dotted line at the $h = 0$ boundary, and solid lines at $h = 1$ and $h = -1$.

•Write the values of $\mathbf{w}$ and b for that classifier.
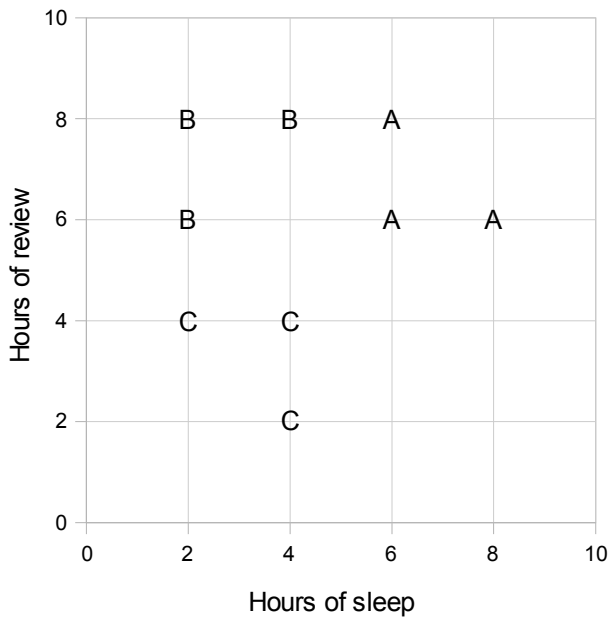
Classifier A: $h_A = \mathbf{w}_A \cdot \mathbf{x} + b_A$
(This classifier separates A's from other grades.)
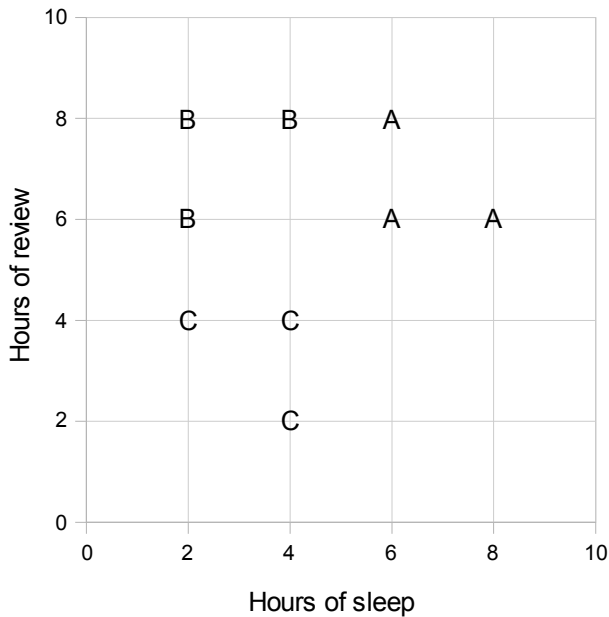
$\mathbf{w}_A =$

$b_A =$

Classifier B: $h_B = \mathbf{w}_B \cdot \mathbf{x} + b_B$
(This classifier separates B's from other grades.)

$\mathbf{w}_B =$

$b_B =$

Classifier C: $h_C = \mathbf{w}_C \cdot \mathbf{x} + b_C$
(This classifier separates C's from other grades.)

$\mathbf{w}_C =$

$b_C =$

Part B (6 points)

After running the three SVMs, Miriam's computer needs to determine which class wins overall. Remember that this is determined by which classifier outputs the *highest* value.

To determine whether classifier A beats classifier B, one can subtract their equations, giving a new equation in terms of the vector **x**. Calculate these equations for all pairs of classifiers – this will help to analyze which classifier wins at each point.

$h_A(\mathbf{x}) - h_B(\mathbf{x}) =$

$h_B(\mathbf{x}) - h_C(\mathbf{x}) =$

$h_A(\mathbf{x}) - h_C(\mathbf{x}) =$

Part C (8 points)

Draw the boundary lines that separate classes A, B, and C in the final classifier on this graph.

The intent here is ***not*** for you to have to solve the equations down to the decimal point. We're looking for a graph that graphically shows the correct overall behavior. In fact, if you're trying to find the point where all three classifiers output the same value, we'll save you the effort by telling you that it should be at 3.8 hours of sleep and 6.2 hours of review.
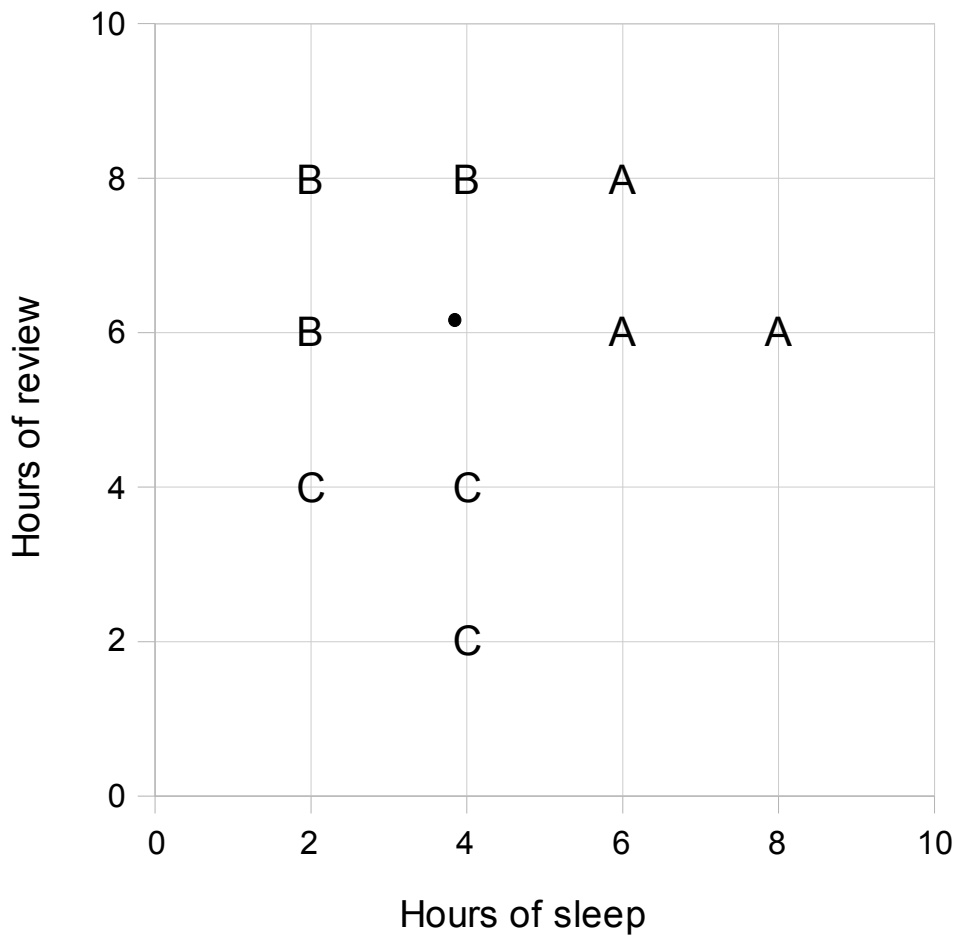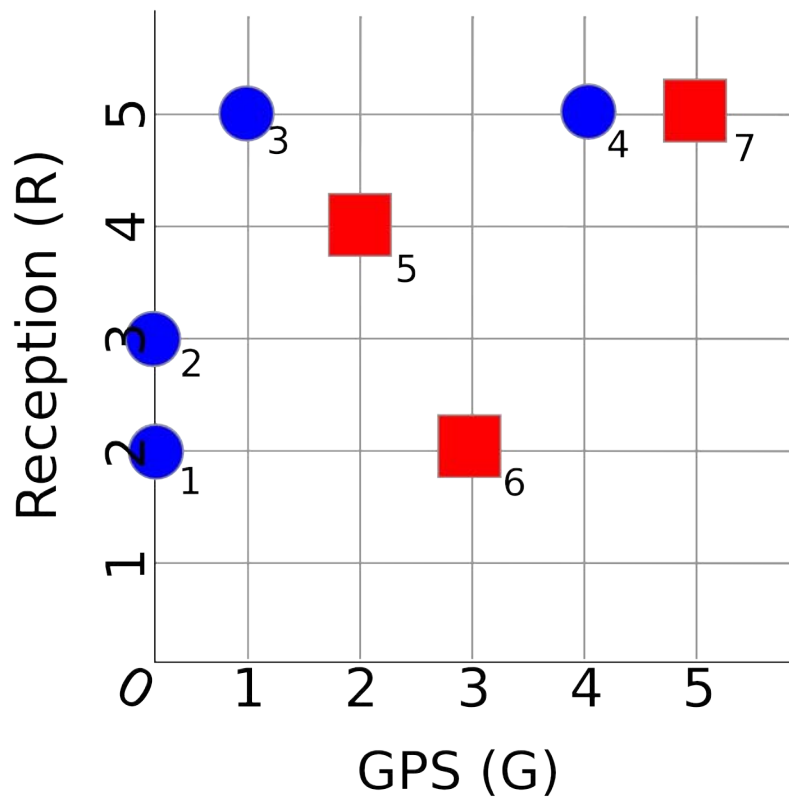

Part D (6 points)

Describe a way in which the behavior of your classifier is counterintuitive, given the meaning of the input data?

Quiz 4, Question 2, Boosting(50 points)

You want to use your cell phone to detect whether you are indoors or outdoors. You use the number of bars of reception (R) and your GPS signal quality (G) as features. Here's the data you gather:

Part A (10 points)

You want to choose a learning algorithm for this data. Circle all the algorithms that can learn this data with zero training error:

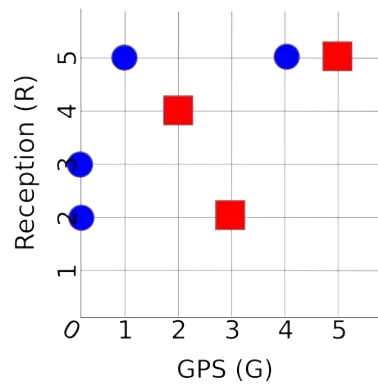   AdaBoost   ID-Tree   Linear SVM   RBF SVM

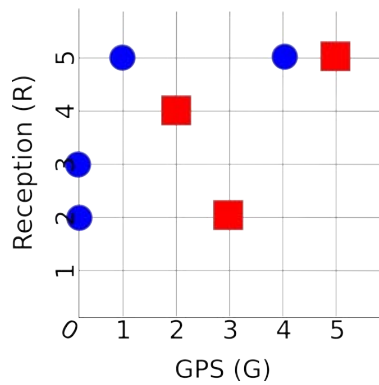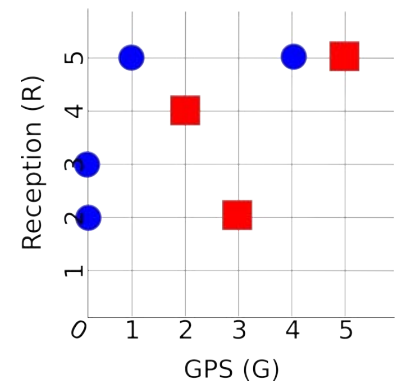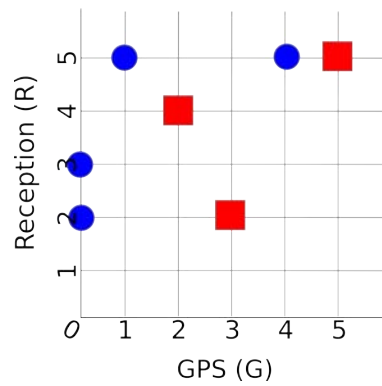RBF = Radial basis function, the exponential kernel; you can choose the RBF's σ.

Part B (10 points)

You decide to train AdaBoost using these stumps:
- $h_1(G,R) = \text{sign}(1.5 - G)$
- $h_2(G,R) = \text{sign}(G - 3.5)$
- $h_3(G,R) = \text{sign}(4.5 - G)$
- $h_4(G,R) = \text{sign}(R - 2.5)$
- $h_5(G,R) = \text{sign}(R - 4.5)$

For each stump, draw the line and circle the misclassified data points.
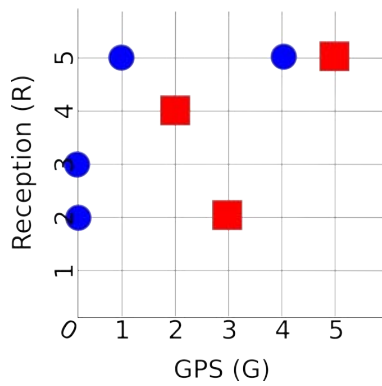
Part C (30 points)

Starting with all weights set to 1/7, run AdaBoost for three rounds. For each round, give the classifier you chose and its error. Also give the final AdaBoost classifier and its error rate. (You don't need to give weights or alphas, but they'll help us give you partial credit.) Break ties by preferring low-numbered stumps.

$h_a(G,R) =$

$\varepsilon_a =$

$h_b(G,R) =$

$\varepsilon_b =$

$h_c(G,R) =$

$\varepsilon_c =$

H(G,R)=

$\varepsilon =$

Quiz 5, Problem 1, (25 points)

Tired of searching for a new president every 10-15 years, the MIT corporation decides to elect a permanent, hereditary monarch. Times are tough, so they are overwhelmed with applicants and decide to hire you to create a kind of filter to reduce the candidate set to a manageable number.

Your first task is to pick some descriptors. You decide to look at intelligence, field, group memberships (if any) , and gender. Men an women form a set. Fields are grouped as at MIT: EE and MechE are part of engineering, Physics and Math are part of science, and XV is part of management. Engineering, science, management, architecture, and humanities are "MIT-type fields."

Using Arch learning, indicate in the tabel what is learned from each example and identify the heuristic involved by name, if known. If nothing is learn, put an x in the corresponding 2 columns.

Part A (18 points)

Candidate	Intelligent	Field	Member of	Gender	Heuristic	What is learned
Good	Yes	EE		Woman		
Bad	No	EE		Woman		
Bad	Yes	EE	Anarchists	Woman		
Good	Yes	MechE		Woman		
Good	Yes	MechE		Man		
Good	Yes	Physics		Man		
Good	Yes	Math		Man		
Bad	No	XV		Man		
Bad	Yes	XV	Anarchists	Woman		

Part B (7 points)

Exhibit an example, which if used as the second example, would teach two characteristics at once.

Candidate	Intelligent	Field	Member of	Gender	What is learned

Quiz 5, Problem 2, (25 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.

Experiments with the newly sighted children indicate that they have difficulty with

1. Naming colors
2. Recognizing objects that are moving
3. Recognizing objects while the child is moving
4. Recognizing overlapping objects
5. None of the above

A newly sighted child, presented with a picture of a ball and a picture of a cube will

1. Readily identify which is which
2. Think they are both the same
3. Think they are both balls
4. Not know which is which
5. None of the above

A newly sighted child will have trouble recognizing a triangle if

1. The triangle is seen against a background of short black lines
2. It is a filled with black, not just three black lines
3. Only three black lines are shown, with no fill
4. The triangle is seen between a circle and a square
5. None of the above

Experiments with newly sighted children suggest the key to visual bootstrapping is

1. Cooperation between vision and sound systems
2. Use of color in figure-ground separation
3. Restricted motion of Infants during the first few months of life
4. Clues provided by the visual motion detection system
5. None of the above

Human children...

1. Like cats, cannot have vision usefully restored if sightless beyond a critical period
2. Like cats, have trouble tracking moving objects if sightless beyond a criical period
3. Prefer to remain blind if sightless beyond a critical period
4. Cannot identfy geometric figures on a computer screen
5. None of the above

Quiz 5, Problem 3, (25 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.

Disorientation experiments with rats, children, and adults demonstrate

1. Rats will not move in the presence of loud rock music
2. Children that behave like rats do not spontaneously use words like left and right
3. Rats and infants do not have full-spectrum color vision
4. Neither small children nor adults cannot reorient themselves if severely sleep derprived
5. None of the above

Chomsky believes the great differentiator between modern humans and humans 100,000 years ago is

1. Improved mitochondria
2. Improved memory
3. Improved hand-eye coordination
4. Ability to plan
5. None of the above

Brooks subsumption architecture is best described as

1. A demonstration of the importance of reasoning and planning in robot activity
2. An approach demonstrated by a robot that collected beer cans on cape-cod beaches
3. An approach to building robots focused on tightly coupled loops between vision and language
4. An idea reminiscent of the notion of abstraction layers in programming methodology
5. None of the above

Winston believes that to develop an account of human intelligence we must

1. Find new ways to exploit the SOAR architectural paradigm
2. Build systems that acquire permanent knowledge from visual experiences
3. Collect commonsense knowledge using volunteers contributing via the web
4. Focus for the time being on what insects can do
5. None of the above

The General Problem Solver Architecture

1. Is a general purpose architecture that unifies reasoning and perception
2. Is a special purpose architecture limited to proving theorems in first order logic
3. Uses a means-ends approach to operator selection
4. Gets stuck if a selected operator cannot be applied in the current state
5. None of the above

Quiz 5, Problem 4, (25 points)

Coen's method

1. Uses boosting to make clusters
2. Uses a neural net to make clusters
3. Uses subsumption to make clusters
4. Uses cross modal coupling to make clusters
5. None of the above

Coen's method

1. Explains why Zebra Finches use a mating song
2. Enables a program to learn to sing a Zebra Finch's mating song
3. Enables a program to learn to to recognize a Zebra Finch's mating song
4. Enables a program to recognize the mating song of particular Zebra Finches
5. None of the above

In Coen's method

- Two clusters are close if they are close in a Euclidian sense
- Two clusters are not close if they are not close in a Euclidian sense
- Two clusters are close if they project proportionately to clusters in another space
- Two clusters are not close if they have different shapes
- None of the above

Coen's method

1. Works well with any distance metric
2. Works well only with the Wasserstein distance metric
3. Works well only when the number of clusters is known in advance
4. Works well only when there are two or three clusters
5. None of the above

Coen's method

1. Works well only if corresponding points in two spaces are provided
2. Works well only if all the points in at lest one cluster are close together
3. Works well only if there is no noise
4. Works well only with labeled data
5. None of the above

This page left almost blank intensionally

MIT OpenCourseWare
http://ocw.mit.edu

6.034 Artificial Intelligence
Fall 2010

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.