6.034 Quiz 2 Solutions
November 17, 2003

Name	
EMail	

Problem number	Maximum	Score	Grader
1	12		
2	37		
3	31		
4	20		
Total	100		

Question 1: Semantic Transition Trees (12 points)

Use the grammar shown below (in the syntax used for PS3) to answer all parts of this question. Lines of the grammar file are numbered for reference.

```
1: (?sentence ?noun-phrase ?verb-phrase
    %end `(action = ,?verb-phrase thing = ,?noun-phrase))
2: (?noun-phrase ?determiner ?noun %end ?noun)
3: (?verb-phrase ?verb ?direction ?place %end ?verb)
4: (?place ?determiner ?noun %end ?noun)
5: (?direction (onto %end '())
               (into %end '())
               (around %end '()))
6: (?verb (hopped %end 'jump)
          (ran %end 'traverse)
          (leaped %end 'jump)
          (walked %end 'traverse))
7: (?noun (rabbit %end 'rabbit)
          (table %end 'table)
          (bunny %end 'bunny)
          (chair %end 'chair))
8: (?determiner (the %end '())
                (a %end '())
                (an %end '()))
```

Part 1.1: Parsing (6 points)

1.1.1: (3 points)

What will be returned from a parse of the sentence, "The rabbit hopped onto the table"? (If the parse fails, enter "NULL")

(action = jump thing = rabbit)

1.1.2: (3 points)

What will be returned from a parse of the sentence, "The rabbit leaped over to the chair"? (If the parse fails, enter "NULL")

NULL

not parsed

Part 1.2: Augmenting the Grammar (6 points)

Suppose we want to augment the grammar so the system can handle sentences with zero or one adjective. The goal is to change the grammar so a parse of sentences such as, "The cute bunny walked onto the table" will not fail but that all sentences that previously parsed correctly will still continue to do so. To accomplish this, We start by creating the adjective tree shown below. This line is added to the end of our grammar file.

```
9: (?adjective (cute %end '()) (red %end '()) (fuzzy %end '()))
```

1.2.1: (3 points)

One other line must be changed so our parser will make use of this adjective tree and achieve the goal described above. Changing what line will accomplish this task? Note that there is more than one correct answer.

2 (1 or 2 or 7 ok)

1.2.2: (3 points)

Enter your changed line here:

```
(?noun-phrase (?determiner ?noun ?end ?noun)
              (?determiner ?adjective ?noun ?end ?noun))
```

Question 2: Nearest Neighbors and Identification Trees (37 points)

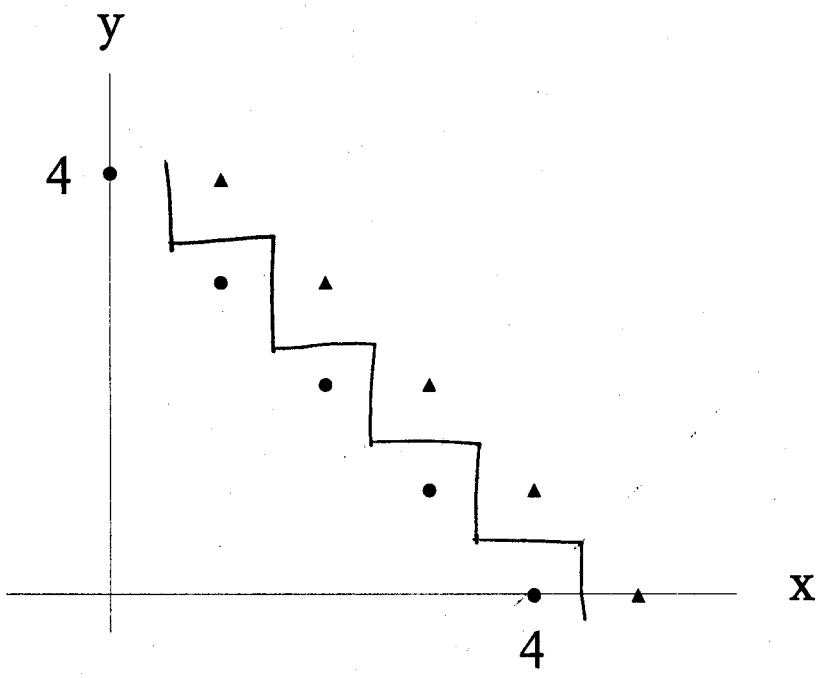
For your convenience in working this problem, we provide a table of logarithms, which you may find useful when checking some of your answers. The table is also on the tear-off sheet at the end of the quiz.

n	$-n \log_2 n$	$-n \log_2 n - (1-n) \log_2 (1-n)$	n	$-n \log_2 n$	$-n \log_2 n - (1-n) \log_2 (1-n)$
0.00	0.00	0.00	0.50	0.50	1.00
0.05	0.22	0.29	0.55	0.47	0.99
0.10	0.33	0.47	0.60	0.44	0.97
0.15	0.41	0.61	0.65	0.40	0.93
0.20	0.46	0.72	0.70	0.36	0.88
0.25	0.50	0.81	0.75	0.31	0.81
0.30	0.52	0.88	0.80	0.26	0.72
0.35	0.53	0.93	0.85	0.20	0.61
0.40	0.53	0.97	0.90	0.14	0.47
0.45	0.52	0.99	0.95	0.07	0.29
0.50	0.50	1.00	1.00	0.00	0.00

Part 2.1: Decision boundaries, warm up (9 points)

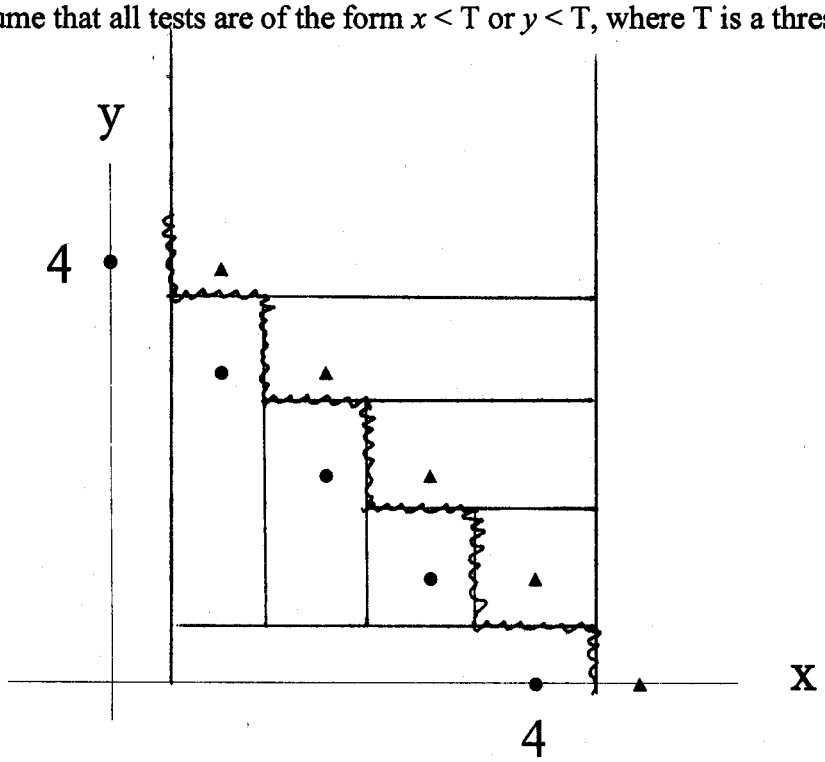
2.1.1: Nearest neighbors (3 points)

Draw the decision boundaries produced by a 1-nearest neighbor classifier on the following samples of circles and triangles.



2.1.2: Standard identification tree (3 points)

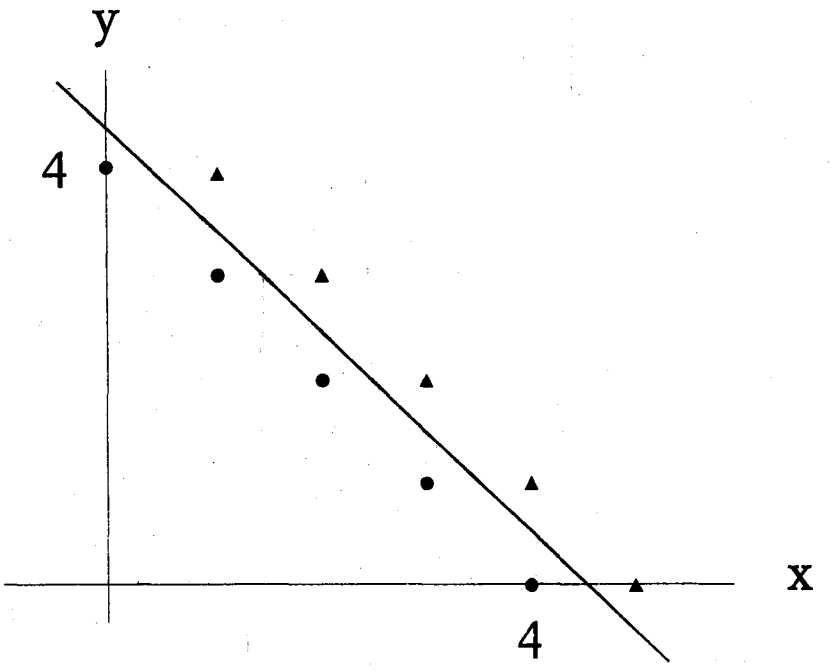
Repeat, but this time, draw the decision boundaries found by an identification tree builder. Assume that all tests are of the form $x < T$ or $y < T$, where T is a threshold value.



(or other solutions that slice one point off at a time)

2.1.3: Augmented identification tree (3 points)

Again, find the decision boundaries found by an identification tree builder, but this time suppose that your test options include not only $x < T$ and $y < T$ but also $x - y < T$ and $x + y < T$, where T is a threshold value.



Part 2.2: Decision boundaries, movies (28 points)

Excited about the performance of classifiers you used for learning viewer preferences and recommending movies, you decide to apply your expertise to a new domain. This time, you want to recommend music to people, based on what they have rated highly before. However before you make any recommendations, you need to recognize each listener's category, based on each listener's music tastes, determined by the listener's ratings for two songs.

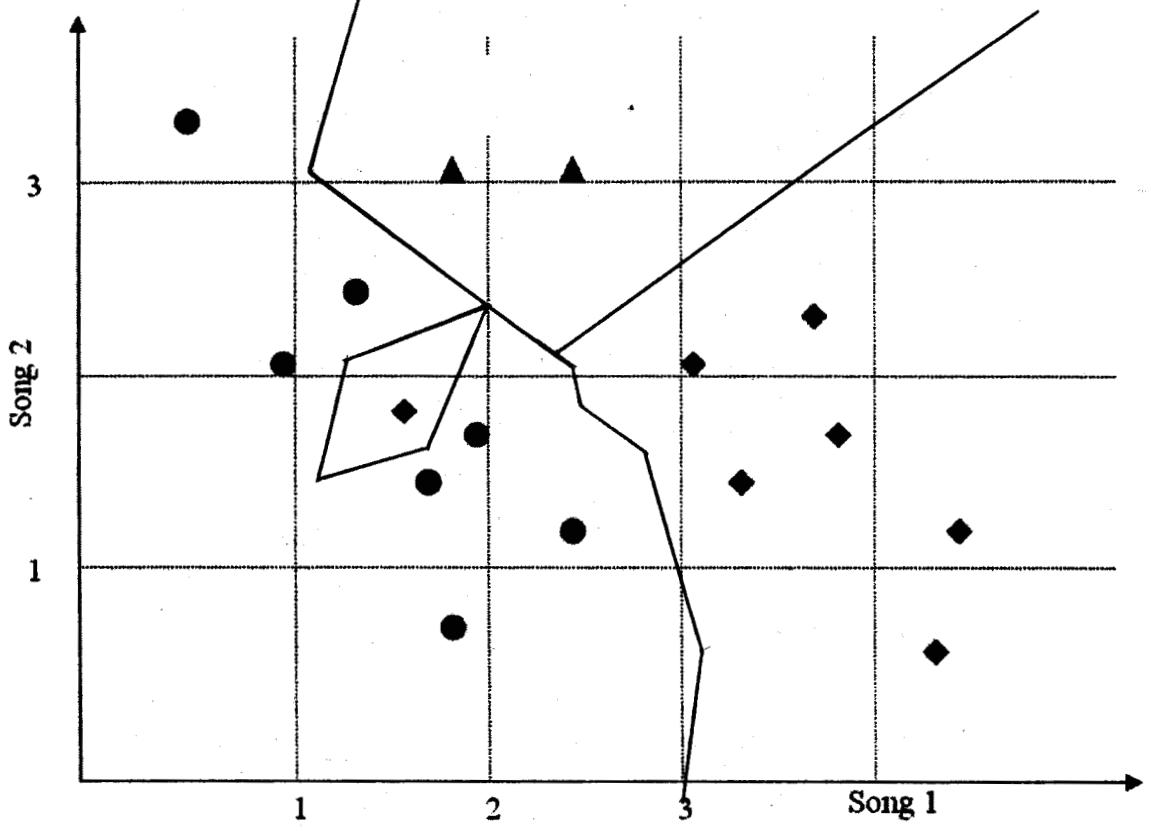
Below, you are given a chart that shows how various people, belonging to 3 different listener groups, rated two songs. To make a recommendation to a new person, you first want to identify to which listener group they belong, and then to make a recommendation based on what others in the listener group rated high.

The circles, triangles and diamonds represent three groups of people with a consistent music taste.

2.2.1: Nearest neighbors (8 points)

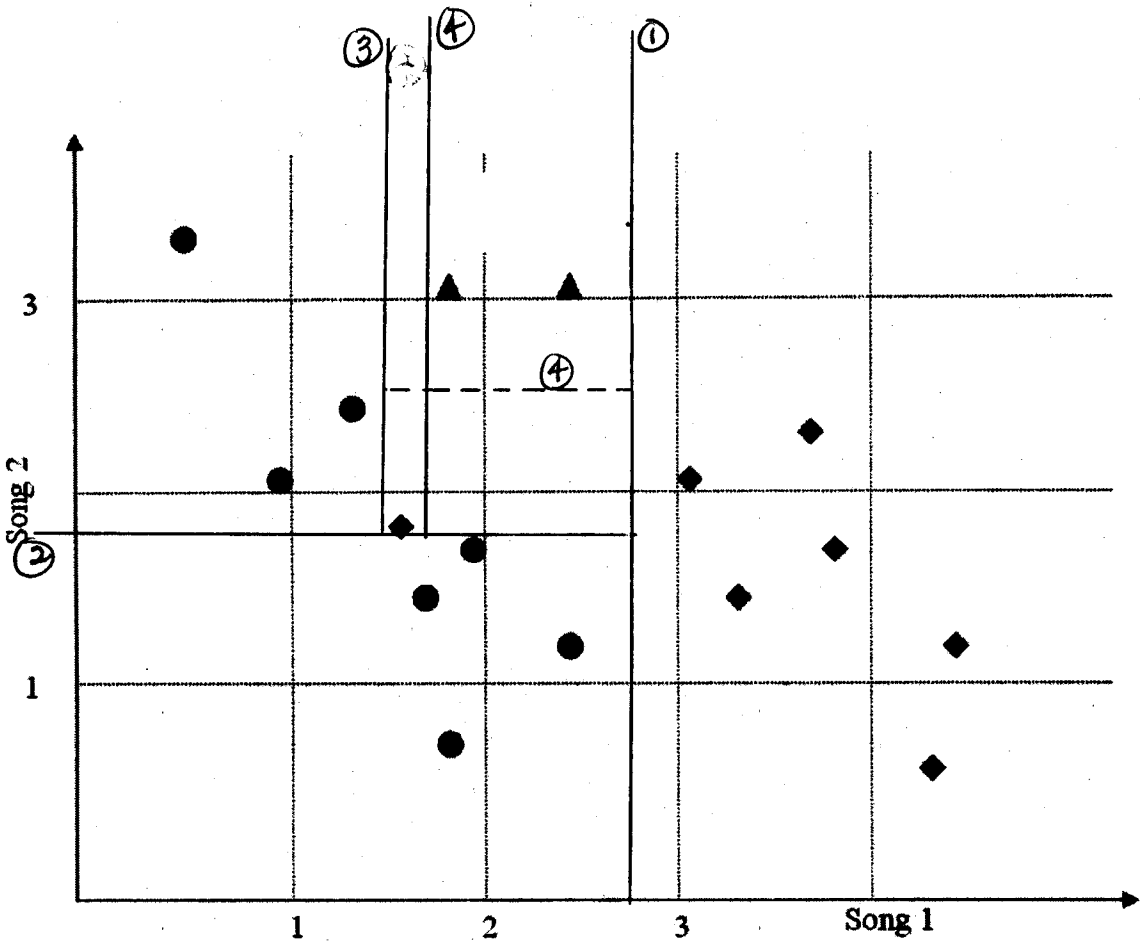
On the chart, draw the decision boundaries produced by a 1-nearest neighbor classifier on the samples. Later on, we will refer to this as option A.

Extra copies of this chart are on the tear-off sheet at the end of the quiz.



2.2.2: Standard identification tree (4 points)

Repeat, but this time, draw the decision boundaries found by an identification tree builder. Assume that all tests are of the form $x < T$ or $y < T$, where T is a threshold value. Later on, we will refer to this as option B.



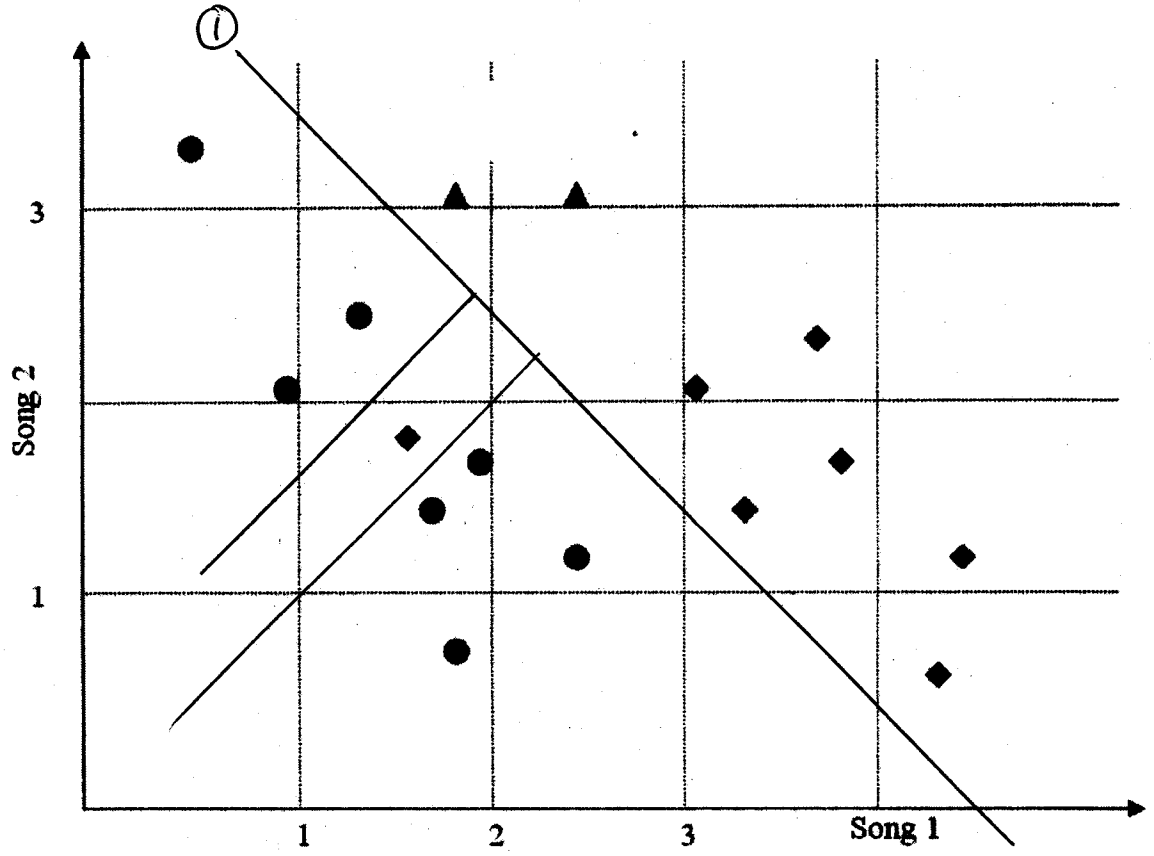
2.2.3: Disorder Calculation (4 points)

What is the entropy (disorder) associated with the first test in the decision tree found by an identification tree builder? You may write your answer as a number or a variable-free expression.

$$\frac{6}{16} \cdot 0 + \frac{10}{16}\left(-\frac{1}{10}\log_2\left(\frac{1}{10}\right) + -\frac{2}{10}\log_2\left(\frac{2}{10}\right) + -\frac{7}{10}\log_2\left(\frac{7}{10}\right)\right)$$

2.2.4: Augmented identification tree (4 points)

Again, find the decision boundaries found by an identification tree builder, but this time suppose that your test options include only $x - y < T$ and $x + y < T$, where T is a threshold value. Later on, we will refer to this as option C.



2.2.5: Disorder Calculation (4 points)

What is the entropy (disorder) associated with the first test in the decision tree found by an identification tree builder described in 2.2.4? You may write your answer as a number or a variable-free expression.

① $\frac{1}{2}\left(-\frac{1}{4}\log_2\left(\frac{1}{4}\right) + -\frac{3}{4}\log_2\left(\frac{3}{4}\right)\right) + \frac{1}{2}\left(-\frac{1}{8}\log_2\left(\frac{1}{8}\right) + -\frac{7}{8}\log_2\left(\frac{7}{8}\right)\right)$

2.2.6: Picking a theory (2 points)

Which of the three options do you consider best?

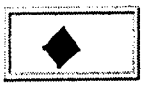
B or C

Which of the three options do you consider worst?

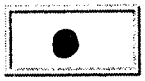
A

2.2.7: Identification (2 points)

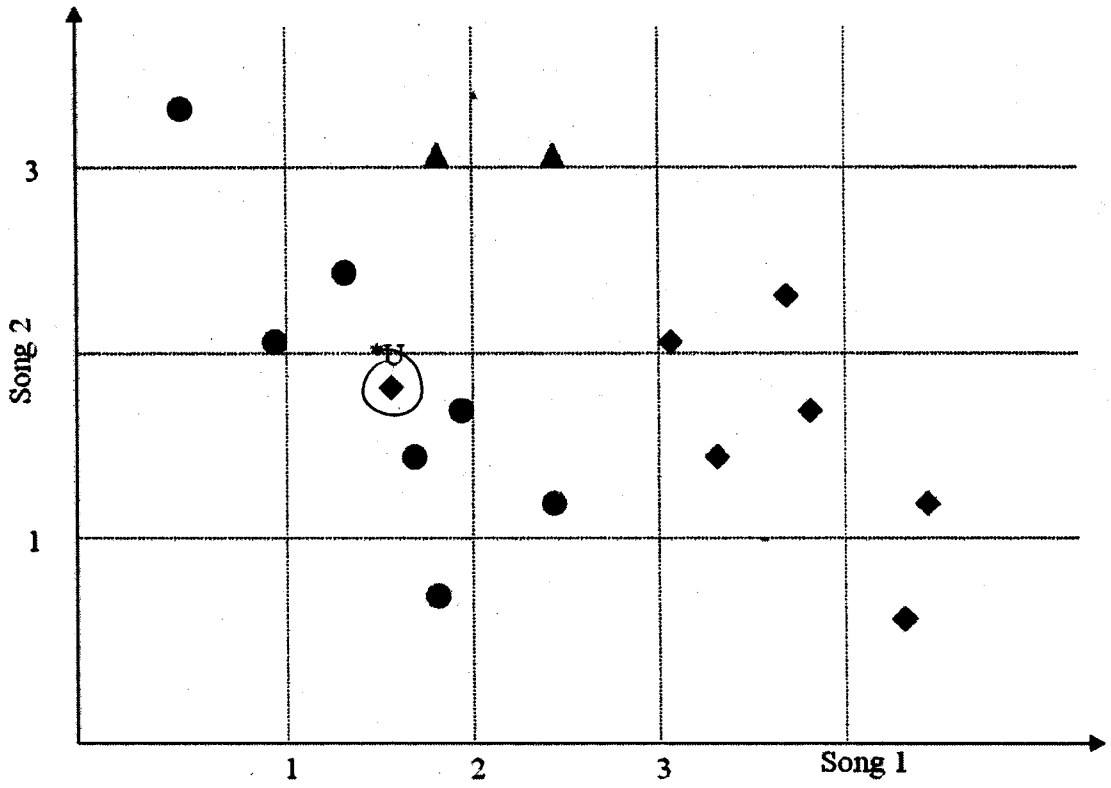
In the following chart, what is the label of the unknown person *U, using the 1-Nearest Neighbor method and the Euclidean distance as the distance metric?



What is the label using the 3-Nearest Neighbor method and the Euclidean distance?



What data point is most likely to be noise? **Circle it on the graph.**

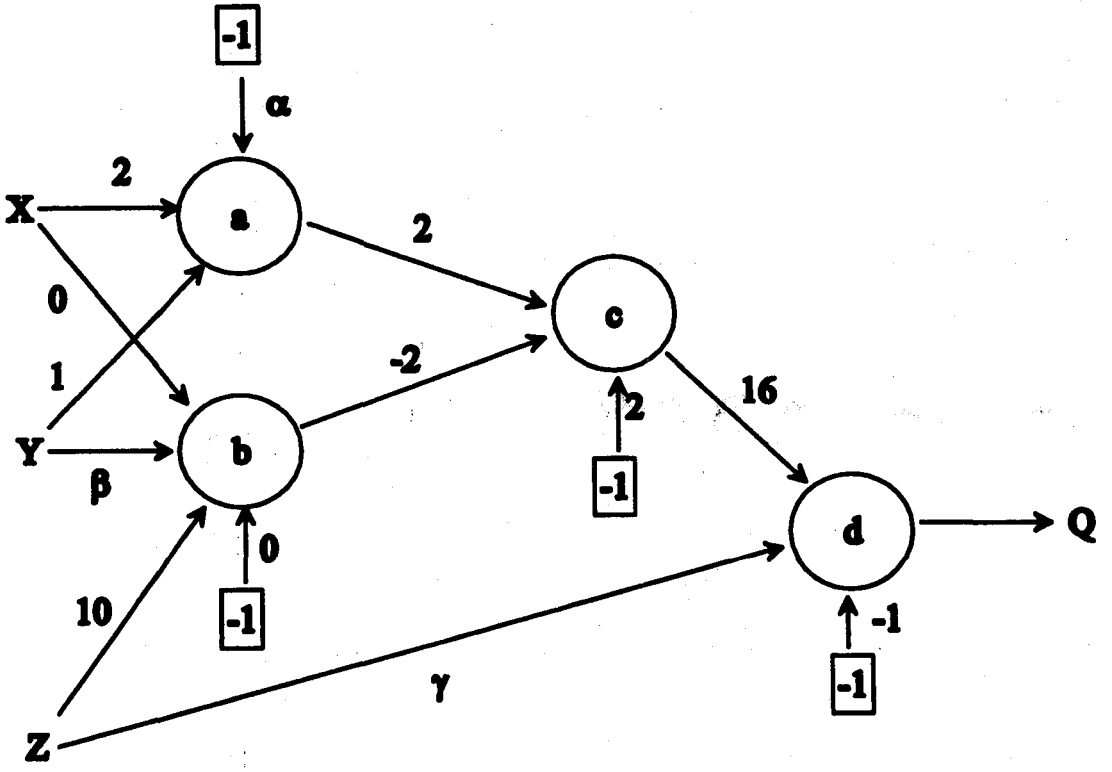


Question 3: Neural Nets (31 points)

Special note: You can get the correct answer to the second, easier part of this problem even if you do not find the correct answer to the first part.

Show your work so that we can try to find opportunities for partial credit.

Consider the following network (several copies are on a tear-off sheet at the end, for your convenience:



There are four neurons, **a**, **b**, **c**, and **d**, all of which use a sigmoid function.

You can use the following values to approximate the sigmoid function. The table is also on the tear-off sheet.

Input	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
Output	0.00	0.01	0.02	0.04	0.12	0.27	0.50	0.73	0.88	0.96	0.98	0.99	1.00

Part 3.1: Forward Propagation (21 points)

You are given all but three of the weights, along with this data, which provides the observed output, given three sets of input values:

X	Y	Z	Q
-2	0	0	0.96
0	1	0	0.96
0	1	0.1	0.73

Note: all of the numbers should come out nicely if you round to the nearest tenth.

Calculate the values of the three unknown weights:

3.1.1: Alpha (12 points)

-4 (see tear-off sheet)

3.1.2: Beta (6 points)

5 (see tear-off sheet)

3.1.3: Gamma (3 points)

-20 (or -19.2)
(see tear-off sheet)

Part 3.2: Backward Propagation (10 points)

Consider a situation when the inputs are **X=2, Y=7, Z=2** and the output is **Q=0.5**. If the desired value **Q*** is 0, and the rate constant is 4, then how much does gamma **change** after backpropagation? Note that we are not asking for the new gamma, but only the change.

-1

$\gamma' = \gamma - r\delta_d z$, $r = 4 + z = 2$

$\gamma' = \gamma - 8\delta_d$

$\delta_d = y_n(1-y_n)(y_n - y_n^*)$
$= \frac{1}{2}(1-\frac{1}{2})(\frac{1}{2}-0)$
$= \frac{1}{8}$

$\gamma' = \gamma - 1$

Question 4: Miscellaneous (20 points)

Circle the best answer for each multiple choice question. No penalty for wrong answers.

Perceptron units are less useful than sigmoid units because:

- Perceptron units are much more computationally expensive
- ⊙ Multilayer perceptrons are too hard to train
- Perceptrons cannot be used to differentiate more than two types of object
- All of the above
- None of the above

Neural networks ...

- Use faithful models of neurons in the human brain.
- Always eventually converge to the correct answer.
- Work best with very little training data.
- All of the above
- ⊙ None of the above

Given an arbitrary binary expression (e.g. **(X AND Y) OR (NOT Z)**) the minimum neural network that can be used to compute it is:

- A network with 1 layer of neurons.
- ⊙ A network with 2 layers of neurons.
- ⊙ A network with 3 layers of neurons.
- A network with 4 layers of neurons.
- A network with an arbitrary number of layers.
- It may not be computable.

You have to develop a classification system given a n-dimensional sample set for which it is known that some of the dimensions are completely irrelevant. Unfortunately, you do not know which of the dimensions are irrelevant. You immediately conclude that the best approach of the following approaches is:

- A nearest-neighbor classifier.
- A neural-net classifier.
- ⊙ An identification-tree classifier.

The topological sorting algorithm ensures that:

- Precedence order is depth-first, up-to-join
- ⊙ Precedence order obeys the no-superclass before subclass principle.
- A precedence order is always found
- All of the above
- None of the above

Robotic arm control has been done successfully by calulating motor torques using:

- Using classical mechanics and carefully measured masses and moments of inertia.
- Genetic algorithms.
- (•) Table lookup with practice-derived table entries.
- Simple feedback mechanisms.
- All of the above
- None of the above

Generally speaking, genetic algorithms can be viewed as:

- Hill climbing with steps in random directions
- Hill climbing with an occasionally big step
- An embodiment of certain principles characteristic of real evolution
- Optimization programs
- (•) All of the above
- None of the above

It is easy to build a natural language interface to relational databases because:

- (•) Noun phrases correspond to relational select operations.
- Verb phrases correspond to relational project operations.
- Conjunctions correspond to relational join operations.
- Negation words (no, not, etc.) correspond to relational deletion operations.
- All of the above
- None of the above

The purpose of a trajectory frame is:

- To speed up relational operations.
- (•) To describe an object in motion.
- To describe a change in a variable.
- All of the above
- None of the above

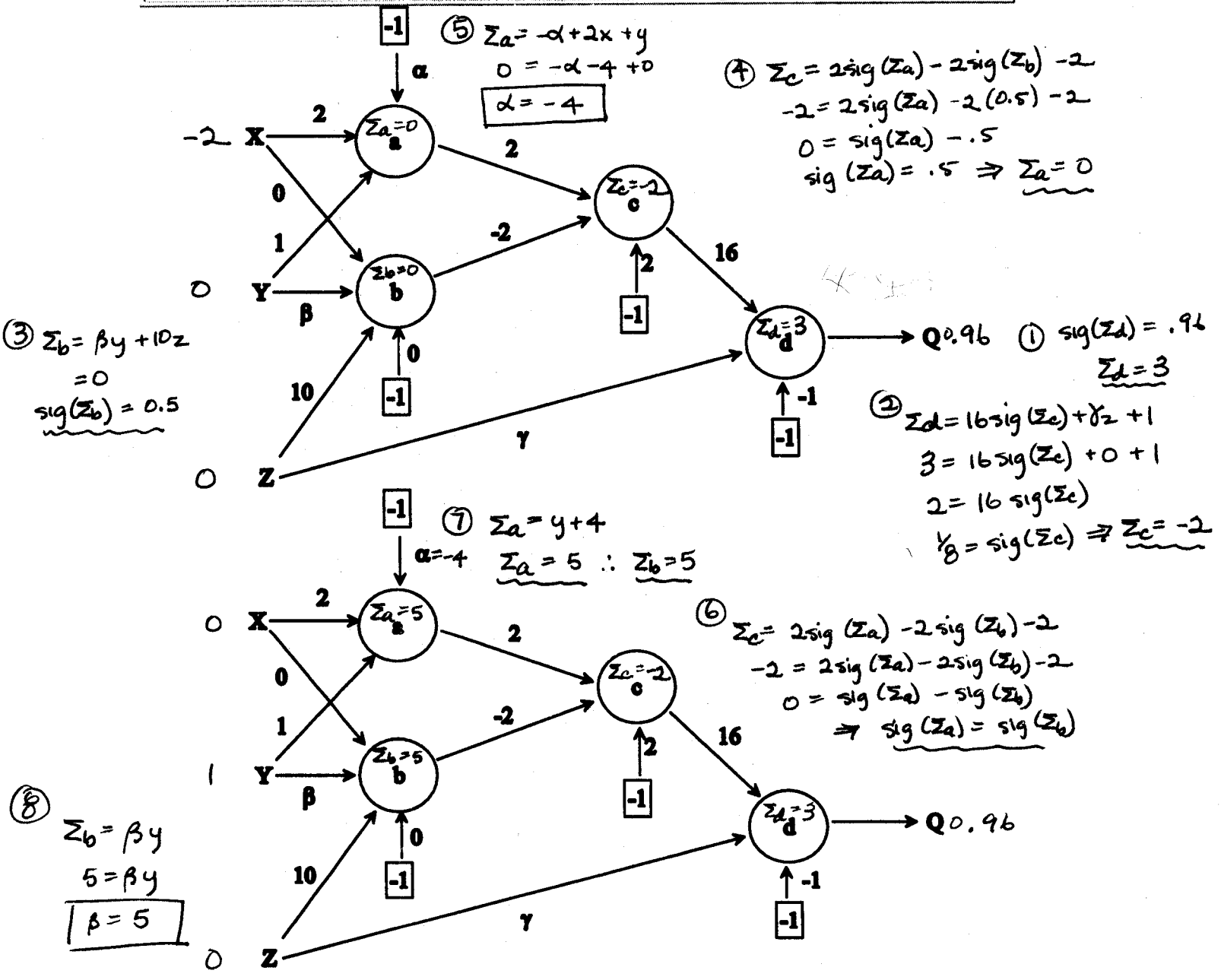
Transition frames can be used to:

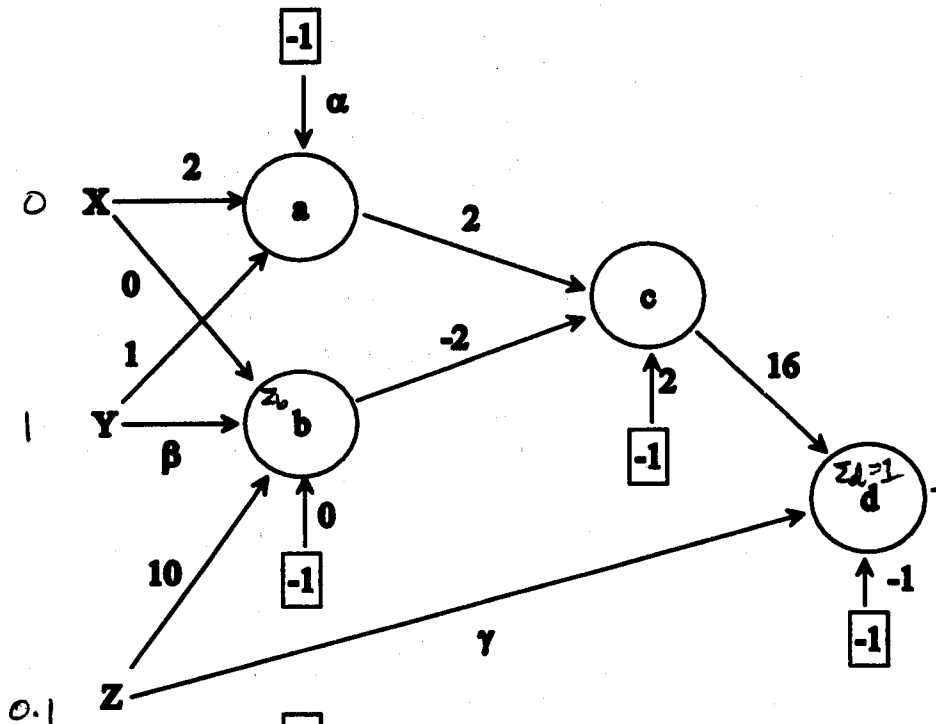
- Describe how variable values change over time.
- Describe events such as car crashes.
- Describe change, appearance, and disappearance.
- (•) All of the above
- None of the above

In this part of the exam, we provide extra, scratch copies of drawings for you to work with. You may tear these off; you need not hand them in.

Table of sigmoid function values:

Input	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
Output	0.00	0.01	0.02	0.04	0.12	0.27	0.50	0.73	0.88	0.96	0.98	0.99	1.00





$z_d = 1$

Q 0.73

⑨ $\text{sig}(z_d) = 0.73$

$z_d = 1$

⑩ $.1\gamma + 2 + 1 = 1$

$.1\gamma = -2$

$\gamma = -20$