6.034 Quiz 2
24 October 2012

Name	ASH KETCHUM
email	

Circle your TA **(for 1 extra credit point)**, so that we can more easily enter your score in our records and return your quiz to you promptly.

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

Problem number	Maximum	Score	Grader
1	50		
2	50		
Total	100		

Problem number	Maximum	Score	Grader
3	5		

We recommend you reserve a few minutes for problem three.

There are 14 pages in this quiz, 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 quiz. As always, open book, open notes, open just about everything, including a calculator, but no computers.

Problem 1 : Pokémon (50 points)

The tools you learn in 6.034 can help you in many situations. Suppose, for example, you are taking a quiz with the following multiple choice questions:

1. Which of the following Pokémon is slowest?

B

a. Abra
b. Butterfree
c. Celebi
d. Diglett
e. Elekid

2. Which of the following Pokémon is heaviest?

B

a. Abra
b. Butterfree
c. Celebi
d. Diglett
e. Elekid

3. Which of the following Pokémon is shortest?

D

a. Abra
b. Butterfree
c. Celebi
d. Diglett
e. Elekid

4. Which of the following Pokémon is fastest?

C

a. Abra
b. Butterfree
c. Celebi
d. Diglett
e. Elekid

5. Which of the following Pokémon has this footprint  ?

C

a. Abra
b. Butterfree
c. Celebi
d. Diglett
e. Elekid

Part A, Constraints (5 points)

You can treat the multiple choice questions as a constraint satisfaction problem, where each question is a variable, and the values are the possible answers A, B, C, D, E.

You decide to represent the domains of each question in the following table. Note that a friend knowledgeable about Pokémon has told you to eliminate some answers right away, so D is not a possible answer for Question 1, A and E are not possible for Questions 2, 3, and 4, and B and D are not possible for Question 5:

Question 1	A	B	C		E
Question 2		B	C	D	
Question 3		B	C	D	
Question 4		B	C	D	
Question 5	A		C		E

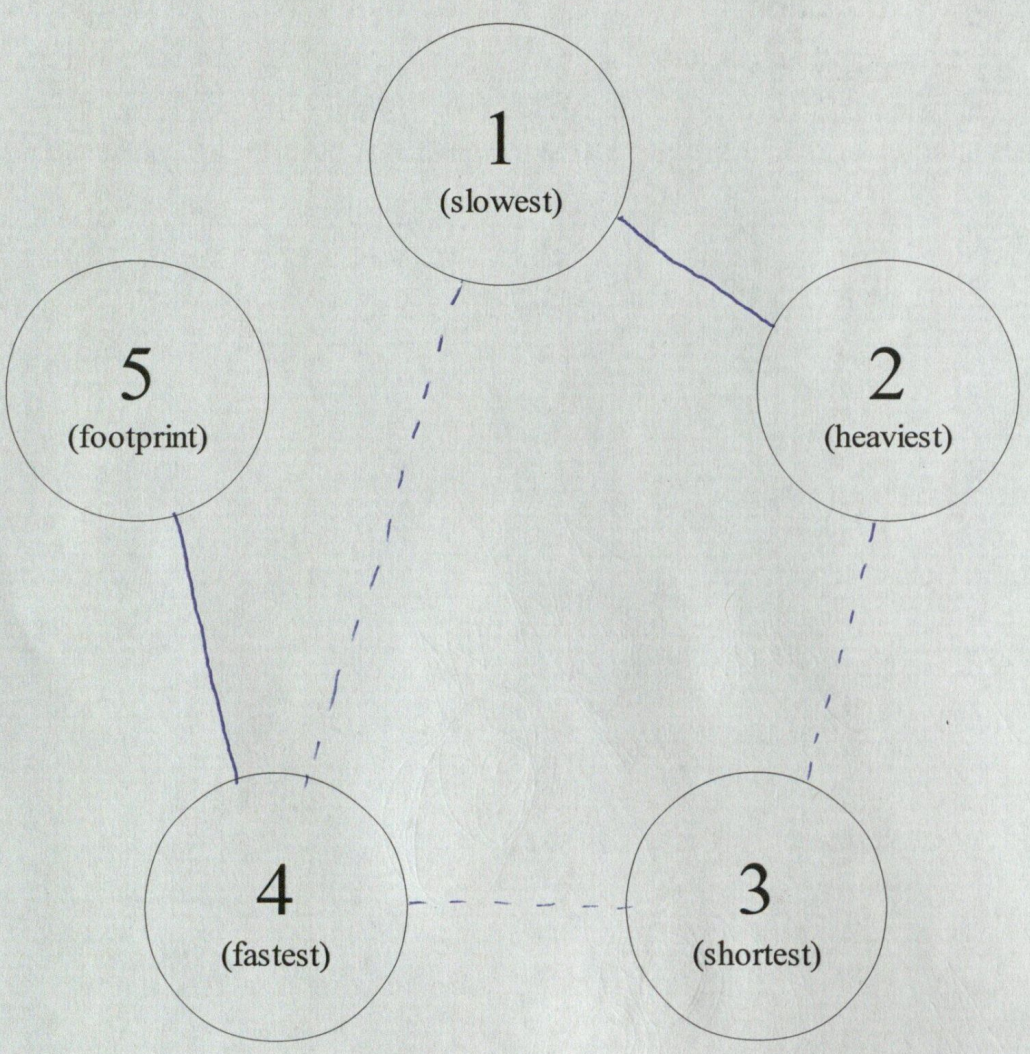
This domain table is also printed on a tear-off sheet located after the final page of the quiz for your convenience.

After some thought, you come up with the following constraints on the questions:

- The fastest Pokémon **cannot** also **be** the slowest.
- The heaviest Pokémon **is** the slowest.
- The shortest Pokémon **is** neither heaviest nor fastest.
- The footprint **belongs to** the fastest Pokémon.

Draw the constraints between the questions on the graph below. Use a **dashed line** (- - - - - - - -) between questions to show that they **cannot** have the same answer. Use a **solid line** (———) to show that two questions **must** have the same answer.

This constraint graph is also printed on a tear-off sheet located after the final page of the quiz for your convenience.

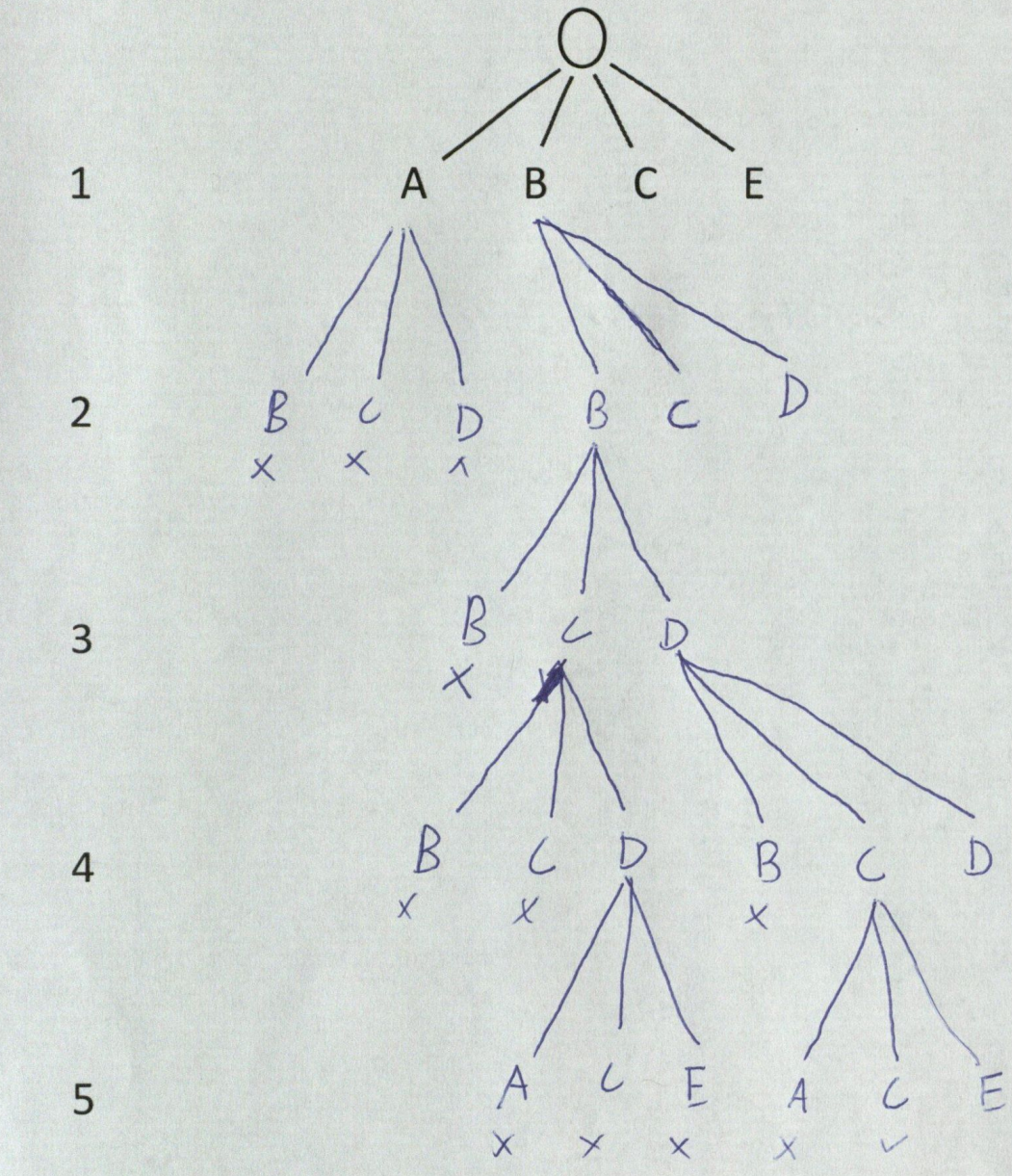


Part B, Searching for answers (15 points)

Find a consistent set of answers to the multiple choice questions **starting from the reduced domains from part A** and using **depth first search** with **no constraint propagation**, checking only constraints on assignment.

- Break ties using alphabetical order.
- Of course, questions can be assigned the same answer as long as no constraints are violated.

You may want to use the tear-off sheet with the domain table and constraint graph to guide your search.



Part C, The right start (15 points)

You realize that the efficiency of your search depends on the question you try to answer first. Out of the five questions, which one should you try to answer first in order to make your search most efficient?

4

Why?

Most Constrained.

Part D, Improved search for answers (15 points)

Now find a consistent set of answers to the multiple choice questions **using the reduced domains from part A** and **depth first search with forward checking**.

Start with the question you chose in part B, then do the rest of the questions in **numerical order.**

4

Your answer to part B: C

ABXE

1

BXD

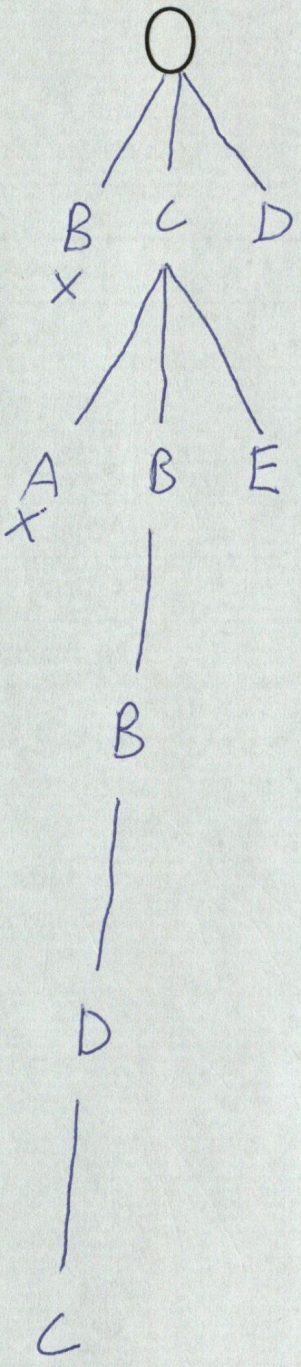
2

BXD

3

AXE

5



Problem 2: Nearest Neighbors and ID Trees (50 points)

The Doctor and his companions, Amy, Rory, and River, have found themselves in a city full of statues. They have to keep their eyes on the statues—some of them are actually quantum-locked Weeping Angels that can only move when they're not being watched. If a Weeping Angel touches one of them, they will be sent back in time and never found again.

Part A: ID Trees (30 points)

The Doctor is looking for a way to distinguish Weeping Angels from other statues. Luckily, River's journal has some information about statues from her previous adventures with the Doctor:

Statue	Classification	Height in ft	Shape	Material
1	Angel	7	Human	Stone
2	Angel	2.5	Human	Stone
3	Not Angel	7	Human	Copper
4	Not Angel	3	Animal	Copper
5	Not Angel	8	Animal	Stone
6	Angel	305	Human	Copper

The table is also printed on a tear-off sheet located after the final page of the quiz for your convenience.

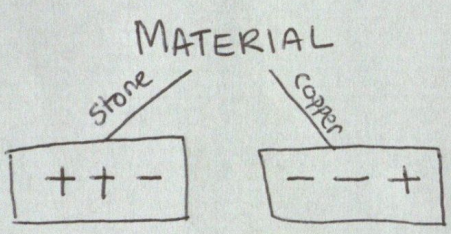
Part A1 (~~30~~ 24 points):

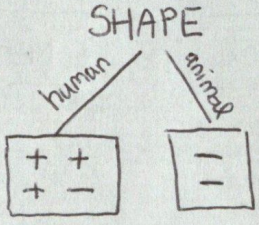
The Doctor looks at this information and decides to make an ID Tree to classify the statues as Angel or Not Angel. Compute the disorder of the following tests when they are applied to the data. (You may leave your answer in terms of logarithms.)

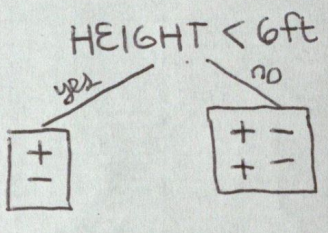
Number	Test	Disorder
Test 1	What material is the statue? (Stone/Copper)	$-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}$
Test 2	What shape is the statue? (Human/Animal)	$-\frac{2}{3}\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)$
Test 3	Is the height less than 6 ft? (Yes/No)	1

Which test does he choose?

TEST 2 (SHAPE)







$$\text{DISORDER} = \left[\frac{3}{6}\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right) + \frac{3}{6}\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right)\right] \times -1$$

$$= -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}$$

$$\text{DISORDER} = \left[\frac{4}{6}\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right) + \frac{2}{6}\left(\frac{2}{2}\log_2\frac{2}{2}\right)\right] \times -1$$

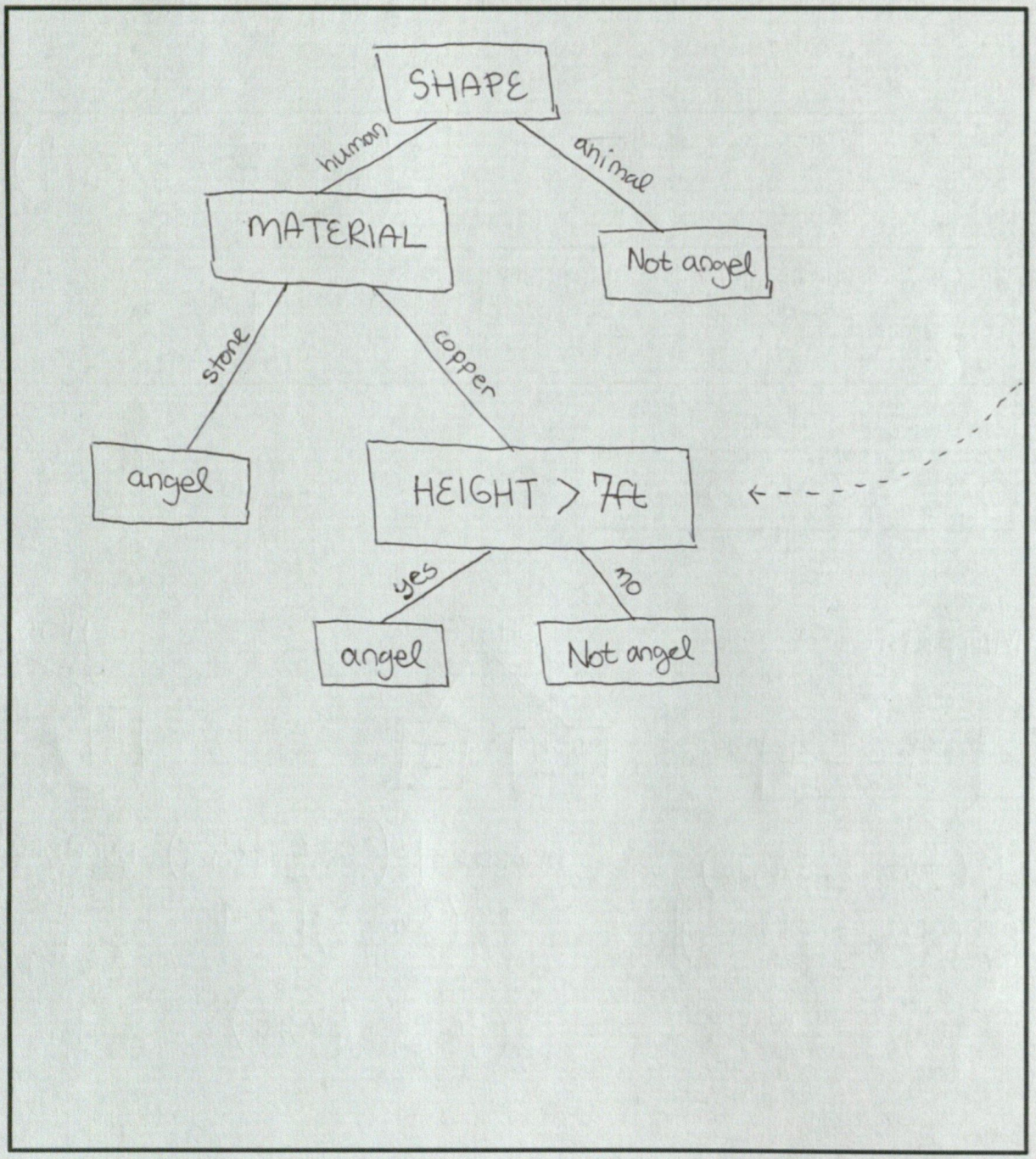
$$= -\frac{2}{3}\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)$$

$$\text{DISORDER} = \left[\frac{2}{6}\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) + \frac{4}{6}\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right)\right] \times -1$$

$$= 1$$

Now draw the complete identification tree for the given data, starting with the test you chose above. For subsequent tests, you may use tests with the following form: What shape is the statue; what material is the statue; and is the height less than a certain threshold value. You need to pick tests that minimize disorder, but you need not show any disorder calculations because you may be able to pick the right test using common sense.

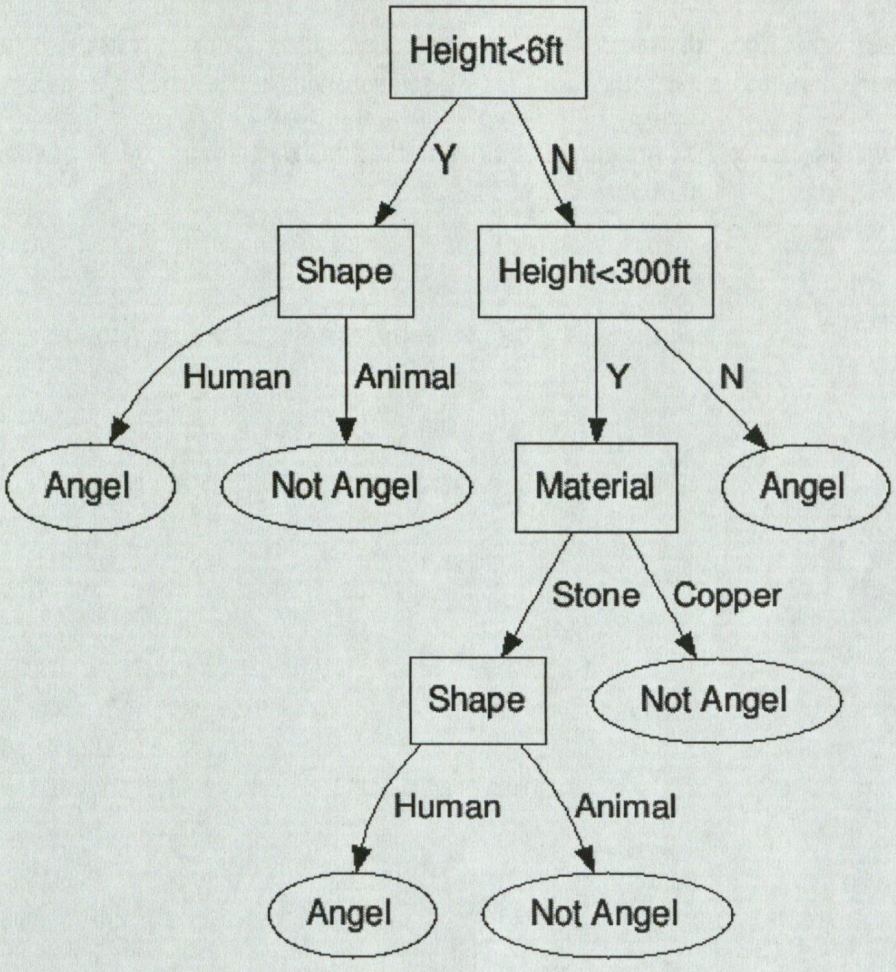
Note that you can use the Height test with any threshold.



TA Note:
Other thresholds were accepted here as long as they separated Statue #3 and Statue #6.

Part A2 (6 points):

River finds another page in her journal which contains the following ID tree:



Compare this tree to the tree you drew in part A1. Given that both ID trees correctly classify all 6 given statues, which do you think would be better at classifying statues in general, not just the given 6? Circle the best choice:

Your ID Tree River's ID Tree Same

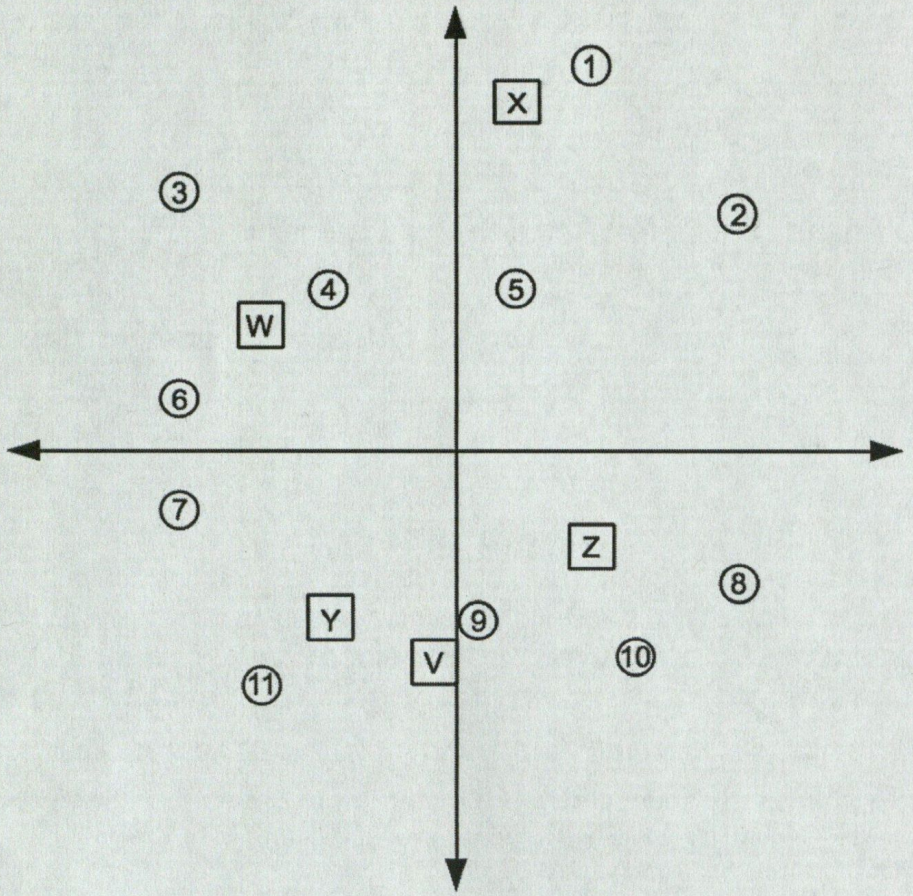
Explain your choice.

A simpler ID Tree is more likely to be better at classifying statues in general. (Occam's Razor). A more complicated tree like River's probably overfits the training data.

Part B: Nearest Neighbors (20 points)

Part B1: (15 points)

The Doctor soon becomes frustrated with ID trees, so he devises another strategy. He notices that Weeping Angels tend to be located near other Weeping Angels, and normal statues tend to be located near other normal statues. In the graph below, the Doctor has plotted the locations of several known Angels and Not-Angels (the points marked with circled numbers), as well as several unidentified statues (the square points V, W, X, Y, Z).



River flips through her journal and manages to recover a chart of nearest-neighbor classifications for the five unidentified statues (V, W, X, Y, Z), based on the 11 known statues. Unfortunately, some of the entries have been smudged:

Statue	Using 1 Nearest Neighbors	Using 3 Nearest Neighbors	Using 5 Nearest Neighbors
W	?	Angel	Not Angel
X	Not Angel	?	Angel
Y	Angel	Not Angel	?
Z	?	Not Angel	?
V	?	?	?

Your task is to use these incomplete nearest-neighbor results to determine the original classification of the 11 numbered points.

Designate each point (1-11) as Angel (A), Not Angel (N), or if the answer can't be determined from the given information, Unknown (U).

1. N
2. A
3. A
4. A
5. N
6. N
7. N
8. U
9. N
10. U
11. A

Part B2 (5 points)

Now that you have restored the classifications of some of the numbered points in part B1, how do you classify point V (Angel [A], Not Angel [N], or Unknown [U]) ...

Using 1 nearest neighbor? N

Using 3 nearest neighbors? U

Using 5 nearest neighbors? N

TA Note: We also accepted U here, although you can determine its answer is N.

Problem 3, Spiritual and Right Now

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.

1 Sussman's propagator architecture is motivated, in part, by his interest in problems such as:

1. Harnessing the power of cloud computing.
2. Harnessing the power of crowd sourcing.
3. Calculating the distances to galaxies.
4. Modeling the propagation of rumors spreading through social media.
5. All of the above.
6. None of the above.

3

2 Sussman's propagator architecture is best explained with a metaphor centered on:

1. A toolbox.
2. Program DNA.
3. Boxes and wires.
4. Baking a cake.
5. All of the above.
6. None of the above.

3

3 Evidently, Sussman's propagator architecture:

1. Makes strong ontological commitments.
2. Requires programs to be written in Propagator Scheme.
3. Ensures programs are provably correct.
4. Features random number generators to help model human thought.
5. All of the above.
6. None of the above.

6

4 Constraint propagation in line-drawing analysis

1. Is an example of forward checking.
2. Takes time exponential in the number of vertexes to converge.
3. Demonstrates that better description yields more constraint and faster performance.
4. Demonstrates the Goldilocks Principle.
5. All of the above.
6. None of the above.

3

5 Constraint exploitation in object recognition involves:

1. The ability of correlation to conquer noise.
2. The power of the Rumpelstiltskin Principle.
3. Determining vertex locations using a small number of standard views.
4. The fact that human heads are more or less spherical, like pumpkins.
5. All of the above.
6. None of the above.

3