

Problem 1: Change This! [9 marks]

For the following questions in this section, if applicable, you can leave your answer in equation form if you wish. For example, you can leave your answer in the form of $2^2 + 2^2$ instead of calculating that out to become 8. Don't forget you have a **binary conversion table** that may be of use!

1. [4 marks] Write an algorithm to convert a number from hexadecimal to decimal.

You can assume that you have already written an algorithm that details how to convert from binary to decimal and can refer to it inside your hexadecimal to decimal algorithm if necessary (using the binary to decimal algorithm in your algorithm is not mandatory).

Version 1 (hexadecimal to decimal)

- a. Have a variable that helps keep track of the power of 16 that we need to multiply by. Set the initial value of the variable to 0.
- b. Have a pointer point at the rightmost digit in the hexadecimal number. The pointer will tell us which number we are currently trying to convert to decimal.
- c. Have a variable that stores the decimal sum of the hex digits converted so far.
- d. Look at the digit the pointer from step b is referring to and determine its decimal value.
 - i. If it is from 0-9, keep the value of the number unchanged.
 - ii. If it is a number from A to F, use the following conversion: A = 10, B = 11, C = 12, D = 13, E = 14, and F = 15.
- e. Multiply the value in step D by 16 to the power of the variable in step A.
- f. Add the product to the variable from step c.
- g. Increment the variable from step a by 1.
- h. Move to the pointer from step B one digit to the left and repeat steps d to h. When there are no more digits left in the hexadecimal number (or you have reached an x), use the value in the variable from step c as the answer.

Version 2 (hexadecimal to binary to decimal)

- a. Use a pointer to point at the leftmost digit in the hexadecimal number. The pointer will tell us which number we are currently trying to convert to binary.
- b. Convert the hexadecimal digit (the one the pointer is referring to) to binary by using the binary conversion table.
- c. Have a variable store the binary conversion of each hexadecimal digit. Each subsequent binary conversion should be added to the right.
 - a. E.g., If the variable originally stored 1101 and you are trying to add 1000 to the number, you will create 1101 1100.
- d. Move the pointer from step A one digit to the right.
- e. Repeat steps A – D until the pointer there are no more hexadecimal digits.
- f. Use the binary to decimal algorithm to convert the binary number stored in step b to decimal.

2. [1 mark] What is the sum of $1111_2 + 1111_2$? This question requires an answer that is not in equation form (i.e., you will not receive marks for having $1111_2 + 1111_2$ as your answer).

$$\begin{array}{r} 1111_2 \\ + 1111_2 \\ \hline 11110_2 \end{array}$$

3. [1 mark] Convert the binary number 100111011 to hexadecimal.

Divide the binary number into groups of four starting from the right. Convert each group of four binary numbers into a hexadecimal number.

0x13B

4. [1 mark] Convert the binary number 1101 1100 to decimal.

$$\begin{aligned} & 1 \times 2^7 + 1 \times 2^6 + 0 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0 \\ & = 128 + 64 + 0 + 16 + 8 + 4 + 0 + 0 \\ & = 220 \end{aligned}$$

The calculations are shown to demonstrate how the answer was derived and is not necessary for full marks.

5. [2 marks] If I had an image that was 4 pixels long and 10 pixels wide and I wanted it to fill up a screen that is 20 pixels long and 40 pixels wide, how many screen pixels would I need to display each image pixel? Show your work.

The first thing we notice is that the image is smaller than the screen which means that the image must be stretched to fill the screen (i.e., there needs to be more than 1 screen pixel required to show each image pixel).

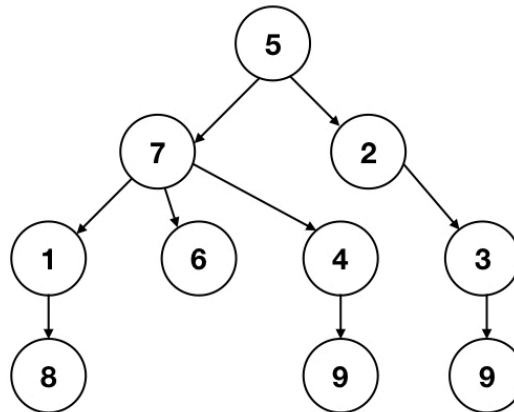
When it comes to length, we need to stretch one pixel of the image across 5 rows of the screen in order to display the image evenly ($20/4$). When it comes to width, we need to stretch one image pixel across 4 columns in order to evenly spread our image across the screen ($40/10$).

To determine how many screen pixels are required to display one image pixel, we multiply 5 by 4 to get 20 pixels.

Problem 2: True or False? [12 marks]

Clearly indicate the answer to the following questions by circling True or False and then writing a 1-2 sentence justification for why you chose that answer. Ambiguously answered questions will be considered incorrect.

1. [2 marks] The following picture is a valid example of a rooted tree.



True

False

Justification:

To qualify as a tree, the following conditions must be met:

- One node is the designated root
- A node can have zero or more children
- All non-root nodes have a single parent

The above tree fulfills these conditions so it is a valid example of a tree

2. [2 marks] If you use the k-means algorithm to cluster the same dataset with the same number of clusters, you will always obtain the same grouping. For example, if you choose to cluster 100 data points into 3 groups, running the k-means algorithm on these data points multiple times will result in the same data points clustered together.

True

False

Justification:

The assignment of data points to clusters will depend on the locations of the centroids at the beginning of the k-means algorithm. Given the same data sets and the same number of clusters, you could potentially end up with a different cluster arrangement.

3. [2 marks] In order to load a webpage, the server has to go to the client to fetch the webpage.

True

False

Justification:

The client has to fetch information from the server- not the other way around.

4. [2 marks] 0x100 and 100 evaluate to the same number.

True

False

Justification:

The 0x in 0x100 indicates that a number is in hexadecimal. 100 indicates that the number is in decimal form. 100 in decimal and 0x100 are different numbers. 0x100 is equivalent to the decimal number 256 while 100 in decimal is equivalent to 0x64.

5. [2 marks] If I zoom in on a picture and it becomes pixelated, the picture has been saved using a lossless algorithm.

True

False

Justification:

If you use a lossless algorithm to store a picture, zooming in will not affect picture quality. The computer will take care of calculating how much the picture should be scaled by and will adjust the pixels in the picture accordingly.

Storing a picture using a lossy algorithm will cause loss of information. When you zoom in on the picture, it will start becoming pixelated at a certain point.

6. [2 marks] A computer stores the value of each pixel by storing the amount of red, gray, and blue that the colour is composed of.

True

False

Justification:

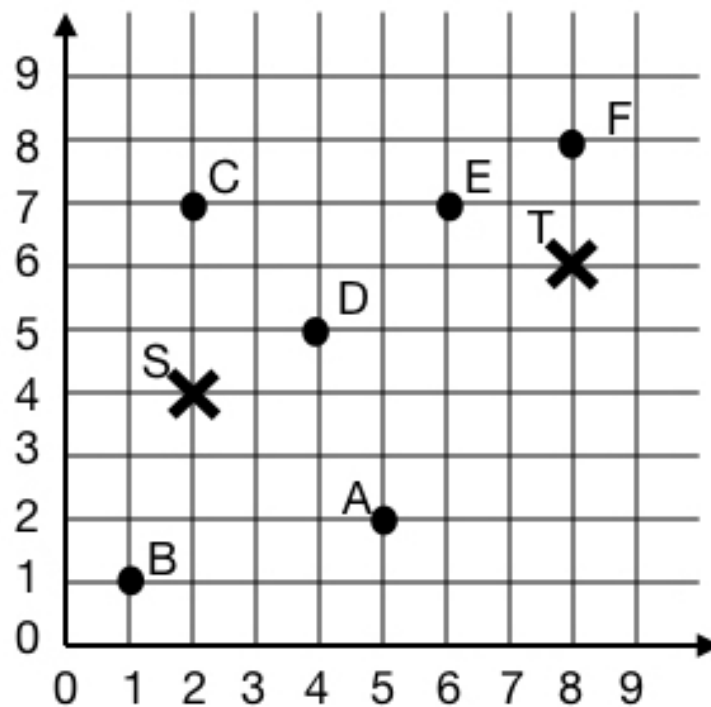
Each pixel is made up of 24 bits. The first 8 bits (when you start counting from the left) represent red, the next 8 bits represent green, and the 8 bits after that represent blue.

Problem 3: Clustering [11 marks]

1. [2 marks] Using one sentence, describe what the k in the k -means algorithm stands for.

The k stands for the number of clusters you want to create using the algorithm. Once k is set, it will not change for the duration of the algorithm running.

2. Consider the following clustering example:



This example has six data points (A, B, C, D, E, and F) that have been represented in the above example by using circles. Daphne wants to use the k -means algorithm to create two clusters. S and T (the two X's in the above example) have been chosen as the initial centroid locations of the two clusters.

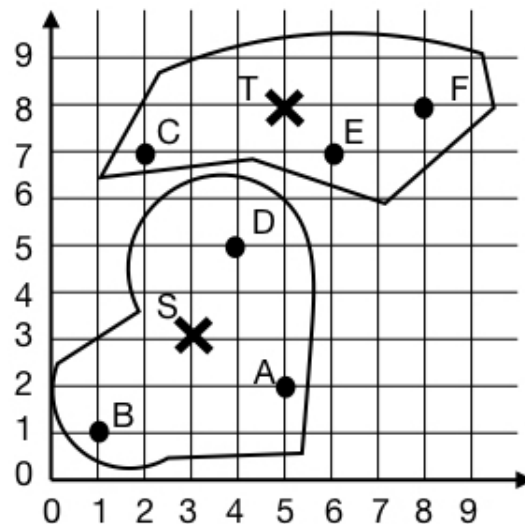
The following chart contains the distances between the various points and the two centroid locations which may or may not be helpful to you as you solve the rest of the questions in this section.

	A	B	C	D	E	F
S	3.6	3.2	3.0	2.2	5.0	7.2
T	5.0	8.6	6.1	4.1	2.2	2.0

- a. [6 marks] Circle which cluster (S or T) points A, B, C, D, E, and F should belong to. Ambiguously circled answers will be considered incorrect.

Point	Cluster S	Cluster T
A	Yes	Yes
B	Yes	Yes
C	Yes	Yes
D	Yes	Yes
E	Yes	Yes
F	Yes	Yes

- b. [3 marks] After the k-means algorithm has iterated for some time, the following clustering appears. The circles indicate which data points have been assigned into which cluster.



Can we stop repeating the k-means algorithm now? Why or why not?

We can't stop repeating the k-means algorithm yet. First, we have to recalculate the new locations of the centroids based on the data points inside the cluster. Then, we have to calculate the distance between each data point and the new location of centroid S and centroid T. After that, we have to determine which cluster a data point belongs to based on the distance between the data point and the centroid. If there are no changes to the clusters, we are finished. If there are changes, we have to repeat the k-means algorithm.

Problem 4: Internet [6 marks]

1. [2 marks] The exam schedule has come out and you can finally start to book a vacation for yourself at the end of April! You start shopping around on the Internet for the best deal on hotels at your ideal destination place but being the savvy shopper that you are, you don't want to buy the first package that you see. You want to shop around on different websites to see if you can get a better deal.

As you sit there browsing, you realize that the price of the hotel seems to go up each time you refresh the page. Assuming that no one else has booked a stay at the hotel you are interested during the time you are browsing for deals online, use the terminology we have learned in class to **briefly** explain (in 2-4 sentences) why the displayed prices are going up.

The cookies saved by the website indicate that you have refreshed the page several times so the price goes up each time you refresh.

2.
 - a. [1 mark] Think about the times when you have posted an article for the individual portion of In the News. What type of communication is it? Circle your answer from the following four options. Ambiguously circled answers will be considered incorrect.
 - i. Synchronous
 - ii. Asynchronous**
 - iii. Both synchronous and asynchronous
 - iv. Neither
 - b. [1 mark] In 2-3 sentences, justify your answer to question 2a.

Sending and receiving are not occurring at the same time. You can post an article without anyone replying for a while (much like email).

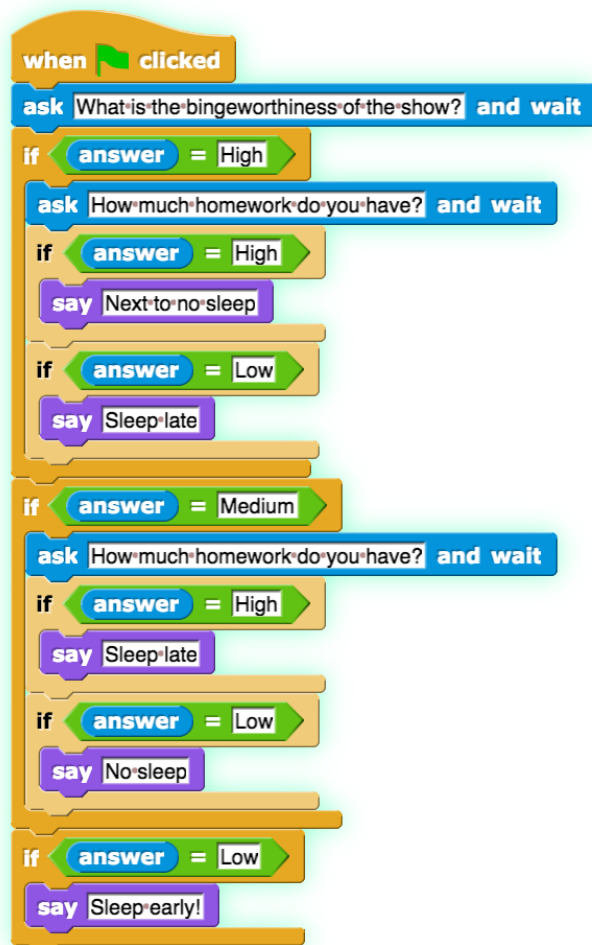
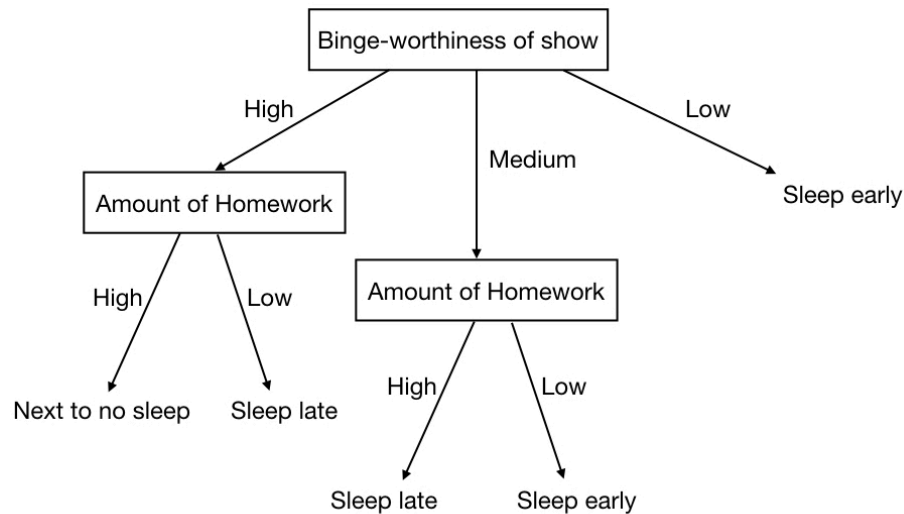
3. [2 marks] If I use the TCP/IP protocol to send a packet from one computer to another, can I predict the packet's path with absolute certainty? Why or why not?

No. The route of a packet depends on how congested the network is at the time. The packet will try to avoid congestion if possible and the amount of congestion in a network depends on who else is using it/what they are doing at the time.

The rest of this page has been intentionally left blank.

Problem 5: Decision Trees [10 marks]

1. [2 marks] Amy has a decision tree below that she coded in Snap. Does the given Snap code snippet match the decision tree? Why or why not?



Write your answer to question 1 here:

The Snap code snippet does not match the decision tree above. The Snap code snippet when we go to the Medium branch for Binge-worthiness of show and then the Low branch for homework, the result is wrong. The result should be "Sleep early" instead of "No sleep".

2. Yelim would like to create a decision tree for the following data.

Weather	Free Time	Number of People Going	Go Hiking?
Sunny	Very little	Many	Yes
Rainy	Lots	Few	No
Snowy	Very little	Many	No
Sunny	Somewhat	Many	No
Sunny	Lots	Few	Yes

- a. [6 marks] What is the entropy for each of the attributes? List the name of each attribute and its entropy.

Weather 3

Free Time 4

Number of People Going 5

- b. [1 mark] Which attribute should Yelim split on? List all possibilities.

Weather

- c. [2 marks] Consider what happens when we add the following rows to Yelim's original data above.

Rainy	Lots	Few	No
Snowy	Very Little	Many	No

Should she still split her tree based on your answer to question b above? Why or why not?

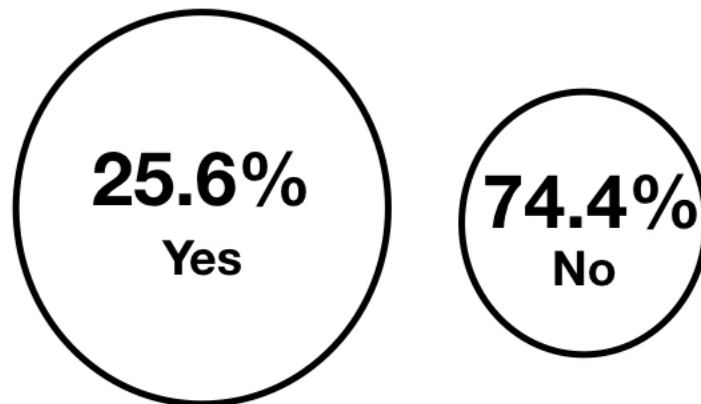
Entropies of each attribute after adding in the two rows are:

- Weather 3
- Free Time 6
- Number of People Going 7

Yelim should still split her tree based on Weather because that is the attribute with the lowest entropy.

Problem 6: Data Visualization [8 marks]

Imagine the following infographic was presented to you during an argument for why a pineapple and pickle mix should be added to all pizzas. The pickle grower association representative has said the following infographic is representative of the survey results from the student population.



1. [4 marks] Should you trust this infographic? List two reasons to support your answer.
 - Survey was conducted by someone from the pickle association and the issue being surveyed includes pickles (possible conflict of interest)
 - You do not know how the people who were surveyed were picked (e.g., if you go to a pineapple and pickle fan club to survey people, your results are going to be skewed)
 - There is no information about how many people were sampled. The percentages don't give us a sense of how the data can be extrapolated to the general student population.
 - There is no information about how the question was presented to the student population. Was there a bias in the way the question was worded? Was the question asked after providing students a choice between an even weirder option?
2. [4 marks] Are there any issues with the way this infographic is displayed? List two reasons to support your answer.
 - The Yes circle is much larger than the No circle even though the Yes percentage is lower.
 - There is no information about how many people were sampled. The percentages don't give us a sense of how the data can be extrapolated to the general student population.

Answers that related to suggesting the infographic should be a pie chart instead were not given marks because this question focused on issues with how the infographic is displayed, not what would make it better.