Information you may find useful

Powers of two

2^9	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
512	256	128	64	32	16	8	4	2	1

Hexadecimal digits

Binary representation	**Decimal representation**	**Hexadecimal representation**
0000	0	0
0001	1	1
0010	2	2
0011	3	3
0100	4	4
0101	5	5
0110	6	6
0111	7	7
1000	8	8
1001	9	9
1010	10	A
1011	11	B
1100	12	C
1101	13	D
1110	14	E
1111	15	F

Problem 1: Multiple Choice – circle the correct answer [4 marks]

i. When do most experts think that superintelligence will arrive?

a. In the next 10 years

b. In the next 10-25 years

c. In more than 25 years

d. never

ii. What is the best way of combining these DNA sequences? Peerwise question 2419517

1.GTTAA
2.AACGT
3.TCCGA
4.GAACG

a. TCCGAACGTTAA

b. GTTAACGTCCGAACG

c. GTTAACGTCGAACG

d. All of the above

e. None of the above

iii. If I had a bag of beans that were all different sizes and I wanted to know how many small beans I have, which of the following would be the better solution for my question? [reading quiz question]

Solution A: Take a bean out of the bag. If the bean is small, then put the bean down and increment the current value of the number of small beans. If the bean is big, then put the bean down. Repeat all the steps until the bag is empty.

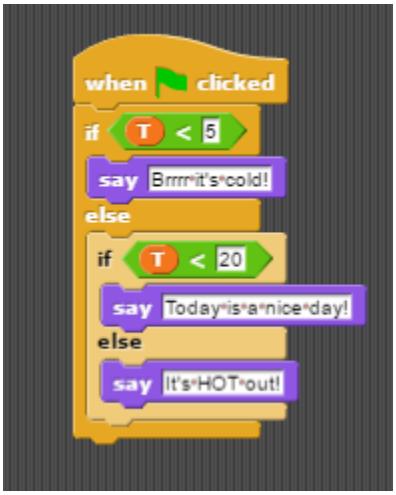
Solution B: Take a sieve that allows all smaller beans to pass through while trapping the big beans. Pour all the beans from the bag into the sieve and count all the beans that pass through the sieve.

a. Both solutions are equally good as they both manage to do what we want.

b. Solution A is better because it can find our solution faster.

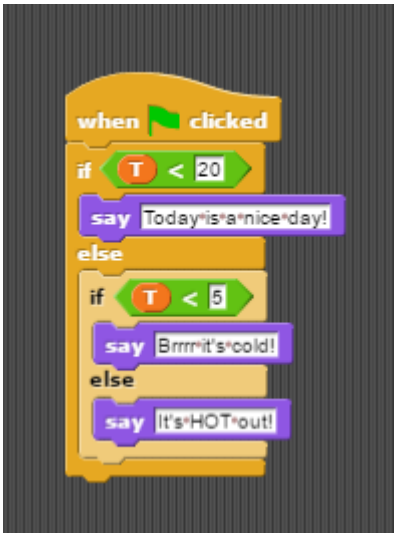
c. Solution B is better because it can find our solution faster.

d. Both solutions are bad.

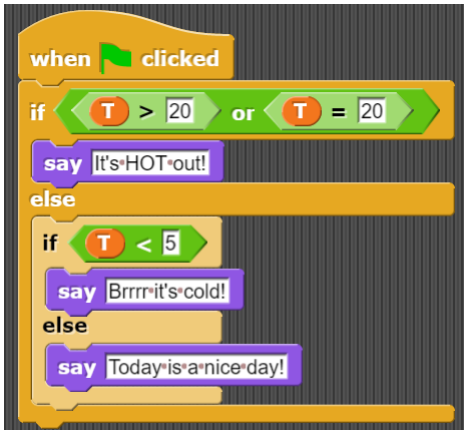


iv. Circle the algorithm that is equivalent to the one above Modified from a PeerWise question

a.



b.



c.

d.

e. None are equivalent.

Problem 2: True or False [4 marks]

Circle the correct answer to each question

Newer methods of automatic language translation attempt to translate entire sentences at once.	True	False
Most of the improvements in the cost of assembling genomes has been from hardware speedups.	True	False
An association rule X→Y holds with support sup% if sup% of transactions that contain X also contain Y. False: that's confidence	True	False
It is possible to have an election system that has independence of irrelevant alternatives and does not result in a dictator. True: you need to also have Pareto efficiency for it to be impossible	True	False

Problem 3: Define this [6 marks]

Expand the following acronyms to their full version as discussed in class. Note that adding additional definitions may result in your losing points. No need to provide anything beyond just listing the expansion of the acronym. E.g., if the question was "URL" the answer should be "Uniform Resource Locator".

a. DOS attack –Denial of Service. Distributed Denial of Service is also okay.

b. DNS – Domain Name System. I'd also accept Domain Name Server

c. ML – Machine Learning. I'd also accept Markup Language, but I doubt anyone will give that as an answer.

d. NLP - Natural Language Processing

e. OS – Operating System

f. RAM – Random Access Memory

Problem 4: Decision Tree Algorithm [3 marks]

Assume that you have the following data and you are trying to make a decision tree to decide whether or not you should keep studying or not.

New Netflix	Weather	Next exam	Study?
Westworld	Snowy	Tomorrow	Yes
Mr. Robot	Rainy	Next week	Yes
Westworld	Sunny	Tomorrow	Yes
Mr. Robot	Snowy	Two weeks	No
Westworld	Rainy	Next week	No
Mr. Robot	Sunny	Next week	Yes
Mr. Robot	Snowy	Tomorrow	Yes
Mr. Robot	Rainy	Two weeks	Yes
Mr. Robot	Sunny	Next week	No
Dr. Who	Snowy	Tomorrow	Yes
Westworld	Rainy	Next week	No
Dr. Who	Sunny	Next week	Yes

Which is the best first column to split on based on the method we discussed in class? Why?

We need to consider entropy, first look at New Netflix:

Westworld – yes, yes, no, no
Mr. Robot – yes, no, yes, yes, yes, no
Dr. Who – yes, yes

Weather:

Snowy – yes, no, yes, yes
Rainy – Yes, no, yes, no
Sunny – yes, yes, no, yes

Next exam:
Tomorrow – yes, yes, yes, yes
Next week – yes, no, yes, no, no, yes
Two weeks – no, yes

So the best attribute to split on is “Next Exam” since it reduces entropy the most given our definition.

1 point for answer
1 point for mentioning entropy
1 point for reasonable explanation

Problem 5: Number conversion [4 marks]

[2 marks] Translate the following decimal number to binary: 343

The answer is: 101010111

[2 marks] Translate the following hexadecimal number to decimal: A866

The answer is: 43110 (did you try turning this upside down on a calculator?!)

Problem 6: Artificial Intelligence [3 marks]

a. What is an argument from lecture or the reading against the claim that artificial intelligence will lead to higher levels of unemployment?

Possible answers include:

- Most jobs are very difficult to automate
- All this has happened before: the loss of jobs in one sector (for example, agriculture, laundering, and other tasks that are already partly automated) led to an increase in employees in other sectors
- Artificial intellience can create jobs as well, balancing out the jobs that it automates

b. List of responsibilities of data entry clerks:
    - Read source documents such as checks, sales reports, or bills
    - Operate data entry device, such as keyboard or scanner
    - Compile, sort, and verify data
    - Compare data with source documents
    - Preparing materials for printing
    - Report errors

    a. What in the list of responsibilities are computers currently capable of undertaking?

    All that do not require physical movement of items/documents, or deciphering handwriting. Computers can now decipher and parse handwriting to some degree, but there could be errors.

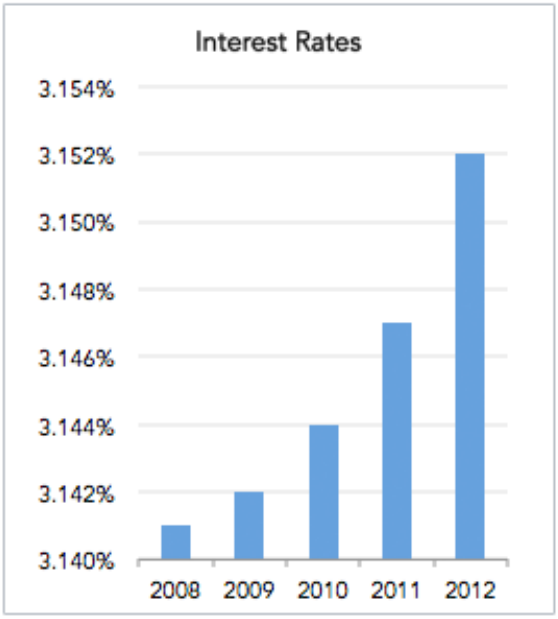
    Tasks that are currently able to be automated are: compiling, sorting, and comparing data, and reporting errors.

    b. Based on the number of tasks that computers are currently able to do, is it likely that this job will be automated in the next 20 years?

    Yes, very likely. In fact, this particular job has already lost more than 43000 positions from 2002 to 2014, a 16% decline.

Problem 7: Visualization [8 marks]

a. [2 marks] Is this graph misleading? Why or why not?



The graph is more misleading because the Y axis is truncated so that smaller changes appear to be more significant. Image from http://www.datapine.com/blog/misleading-data-visualization-examples/#

b. [2 marks] Which is more appropriate for showing the percentage of English native speakers in Vancouver, a line chart or a pie chart? Why?

A pie chart is more appropriate since it is used to show what makes up a whole, which this question asks for. A line chart is more appropriate if there is some reason for the ordering, which there is not in this scenario.

1 mark for answer, one for explanation

c. [2 marks] If you want to find out how you did relative to your classmates in an exam, which would you use: mean, median, or mode? Why?

Median because you want to know if you are in the top 50% or bottom 50%.

1 mark for answer, 1 for explanation.

d. [2 marks] What are two factors that we covered in class that you should consider when presented with an infographic or statistic?

Possible answers include:
Who did they survey? (Where did they find these people, what are their backgrounds/affiliations?)
How many people did they survey?
Is it possible that they manipulated the data? (for example, by omitting unwanted results)
Do they have an agenda they want to push?
Generally: where did the numbers come from?

[also a practice exercise]

Problem 8: In brief [6 marks]

a. [2 marks] You are going to the website https://connect.ubc.ca . You notice that to log on it took you to https://shibboleth.id.ubc.ca Should you click? Why or why not?

Yes. Since shibboleth and connect both have the same last two domain names, they are likely run by the same organization and thus are likely to be safe

b. [1 mark] Why can the time for an Internet packet to get to from one computer to another vary?
Network congestion can vary from time to time. Also acceptable would be a comment about how it might have taken

c. [1 mark] What is an application of bioinformatics as discussed in class?

Sequencing genes (segments of DNA) to identify genetic risks and cures.

d. [1 mark] What are potential reasons to use a database as covered in class?

Possible answers include:

- Storing large amounts of data
- Answering search queries by searching through data and returning what was requested
- Sorting/categorizing data
- Backing up data in case the system goes down

e. [1 mark] Give one technological reason presented in class why quantum computing is not likely to be coming to your laptop any time soon.

(should be because of temperature requirements). I may accept that there are only a small number of bits.

Problem 9: Image Representation [3 marks]

a. [1 mark] What is the difference between lossless and lossy compression?

You can recover all data from your original file when you uncompress something compressed with lossless compression. I think the students will say something like lossy compression = grainy images when you reopen it (since that's exactly what they did in lab).

b. [2 marks] Which of the following affects the size of an image file? (circle all that apply)

- Length of the image
- Width of the image
- Bit depth
- Type of file

[0.5 marks for each correct answer, minus .5 marks for each incorrect answer]