

Information you may find useful. This sheet will NOT be graded.

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 Answers [9 Marks]

- 1) [1 mark] The percentage of Bachelor's degrees conferred to women in the U.S. has increased significantly in the past 30 years in all but one of the following fields, where it has decreased significantly. Which field is the one exception?
- a. Computer Science
 - b. Engineering
 - c. Mathematics and Statistics
 - d. Physical Sciences
- 2) [5 marks] Facebook collects personal data on its users from which of the following sources? Circle all that apply.
- a. Personal data that users provide, e.g., when setting up a profile on Facebook.
 - b. Data that describes devices, ISP providers etc. that users use to access Facebook.
 - c. Data that describes users' behaviour on other websites, while logged on to Facebook (e.g., ads clicked).
 - d. Data on some websites that you visit while *not* logged on to Facebook.
 - e. Data from trusted third-party partners.
- 3) [1 mark] 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

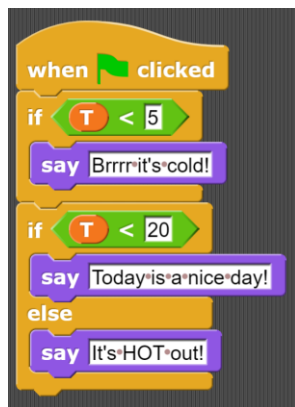
4) [1 mark] What is the best way of combining these DNA sequences?

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

- a. TCCGAACGTTAA
- b. GTTAACGTCCGAACG
- c. GTTAACGTCGAACG
- d. All of the above
- e. None of the above



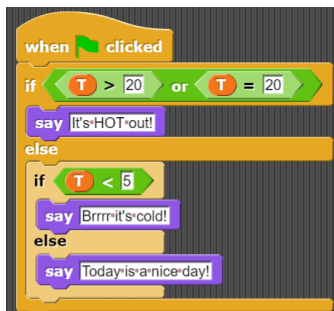
5) [1 mark] Circle the algorithm that is equivalent to the one above



a.



b.



c.



d.

e. None are equivalent.

Problem 2: 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”.

1. [1 mark] DOS attack -Denial of Service. Distributed Denial of Service is also okay.

2. [1 mark] DNS -Domain Name System. I'd also accept Domain Name Server

3. [1 mark] ML - Machine Learning. I'd also accept Markup Language, but I doubt anyone will give that as an answer.

4. [1 mark] NLP - Natural Language Processing

5. [1 mark] OS - Operating System

6. [1 mark] RAM - Random Access Memory

Problem 3: The Internet [8 Marks]

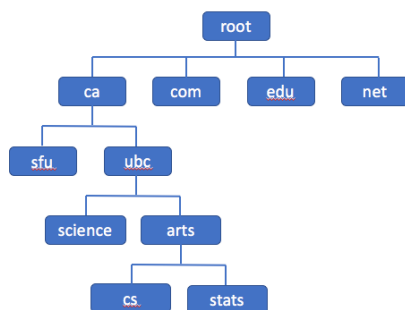
1) [1 mark] What information does a TCP/IP packet contain?

A: A unit of information, the destination IP address and the return IP address

2) [3 marks] Suppose you are sending an email, which is broken into packets for transmission over the internet. State whether each of the following is true or false.

- | | | |
|--|-------------|--------------|
| a) All of the packets are sent along the same path to the destination. | True | <u>False</u> |
| b) The packets may arrive out of order. | <u>True</u> | False |
| c) Some packets may not arrive at all. | <u>True</u> | False |

3) Following is a representation of a domain name hierarchy:



a) [1 mark] Add the domain stats.ubc.net to the above representation

The student should add a node for ubc below the .net node and a node for stats below that.

b) [2 marks] Which domain is more related to stats.arts.ubc.ca: stats.ubc.net or science.ubc.ca. Why?

Science.ubc.ca. They have many more common ancestors in the hierarchy. 2 marks; one for answer, one for reasoning.

c) [1 mark] Give an example of a top-level domain shown in this diagram. ca, com, ... net

Problem 4: Data Representation [5 Marks]

1) [1 mark] The RGB setting for blue is: Circle one:

- a. 00 00 00
- b. FF 00 00
- c. 00 FF 00
- d. 00 00 FF
- e. FF FF FF

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

The answer is: 0b101010111

3) [2 marks] Translate the following hexadecimal number to decimal: 0xA866

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

Problem 5: Algorithms [4 Marks] [adapted from AP exam]

A music critic completes an article for a magazine, and realizes that he has confused two musicians: Jay-Z and Kanye. Before submitting the article, he needs to change all occurrences of “Jay-Z” to “Kanye” and all occurrences of “Kanye” to “Jay-Z”. The critic will use the fact that the word “Hozier” does not appear anywhere in the article.

Which of the following algorithms can be used to accomplish the music critic’s goal? Circle all that are correct.

- (a) First, change all occurrences of “Jay-Z” to “Kanye.”
Then, change all occurrences of “Kanye” to “Jay-Z.”
- (b) First, change all occurrences of “Jay-Z” to “Kanye.”
Then, change all occurrences of “Kanye” to “Jay-Z.”
Last, change all occurrences of “Hozier” to “Kanye.”
- (c) First, change all occurrences of “Jay-Z” to “Hozier.”
Then, change all occurrences of “Kanye” to “Jay-Z.”
Last, change all occurrences of “Hozier” to “Kanye.”
- (d) First, change all occurrences of “Jay-Z” to “Hozier.”
Then, change all occurrences of “Hozier” to “Kanye.”
Last, change all occurrences of “Kanye” to “Jay-Z.”

Problem 6: Algorithms [4 Marks] [taken verbatim from the AP Principles sample exam]

There are 32 students standing in a classroom. Two different algorithms are given for finding the average height of the students.

Algorithm A

Step 1: All students stand.

Step 2: A randomly selected student writes his or her height on a card and is seated.

Step 3: A randomly selected standing student adds his or her height to the value on the card, records the new value on the card, and is seated. The previous value on the card is erased.

Step 4: Repeat step 3 until no students remain standing.

Step 5: The sum on the card is divided by 32. The result is given to the teacher.

Algorithm B

Step 1: All students stand.

Step 2: Each student is given a card. Each student writes his or her height on the card.

Step 3: Standing students form random pairs at the same time. Each pair adds the numbers written on their cards and writes the result on one student's card; the other student is seated. The previous value on the card is erased.

Step 4: Repeat step 3 until one student remains standing.

Step 5: The sum on the last student's card is divided by 32. The result is given to the teacher.

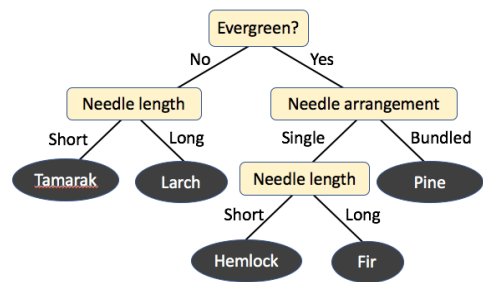
Which of the following statements are true? Circle your answers.

- (a) Algorithm A always calculates the correct average, but Algorithm B does not.
- (b) Algorithm B always calculates the correct average, but Algorithm A does not.
- (c) Both Algorithm A and Algorithm B always calculate the correct average.
- (d) Neither Algorithm A nor Algorithm B calculates the correct average.

Problem 7: Decision Trees on Trees [7 Marks]

Below is a table of training data identifying some coniferous trees, along with a corresponding decision tree and some Snap! code fragments.

| Evergreen? | Needle length | Needle arrangement? | Tree type |
|------------|---------------|---------------------|-----------|
| No | Short | Tufted | Tamarak |
| No | Long | Tufted | Larch |
| Yes | Short | Single | Hemlock |
| Yes | Short | Bundled | Pine |
| Yes | Long | Single | Fir |



```
ask [Is the tree evergreen?] and wait
if [answer = Yes]
  ask [How are the needles arranged?] and wait
  if [answer = Bundled]
    say [This is a pine]
  if [answer = Single]
    ask [Are the needles short or long?] and wait
    if [answer = Short]
      say [This is a hemlock]
    else
      say [This is a fir]
  else
    ask [Are the needles short or long?] and wait
    if [answer = Short]
      say [This is a tamarak]
    if [answer = Long]
      say [This is a larch]
```

Left

```
ask [Is the tree evergreen?] and wait
if [answer = No]
  ask [Are the needles short or long?] and wait
  if [answer = Short]
    say [This is a tamarak]
  else
    say [This is a larch]
else
  ask [How are the needles arranged?] and wait
  if [answer = Single]
    ask [Are the needles short or long?] and wait
    if [answer = Short]
      say [This is a hemlock]
    else
      say [This is a fir]
  else
    say [This is a pine]
```

Right

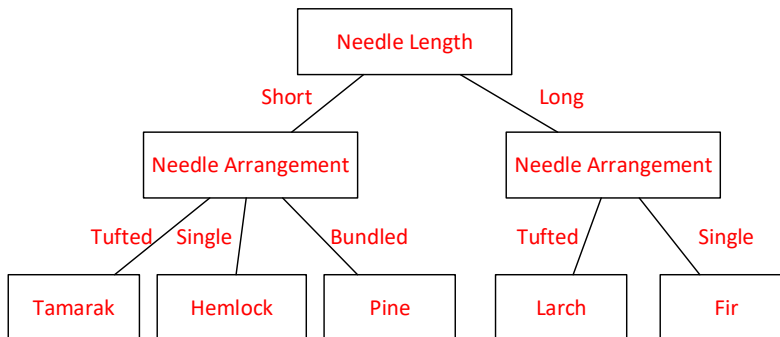
- 1) [1 mark] How many nodes does the decision tree have? 9 nodes
- 2) [1 mark] How many leaves does the decision tree have? 5 leaves
- 3) [1 mark] Which of the Snap! Code fragments correctly implement the decision tree algorithm? Circle one.
A. Left B. Right. C. Neither Left nor Right. D. Both Left and Right.

- 4) [1 mark] When building decision trees, why is it advantageous to have as few nodes as possible?

The fewer the nodes, the less time on average to arrive at the decision.

- 5) [3 marks] In the space below, draw a valid decision tree with *fewer* nodes than that shown on the previous page, for the same training data. (The table is repeated here for convenience.)

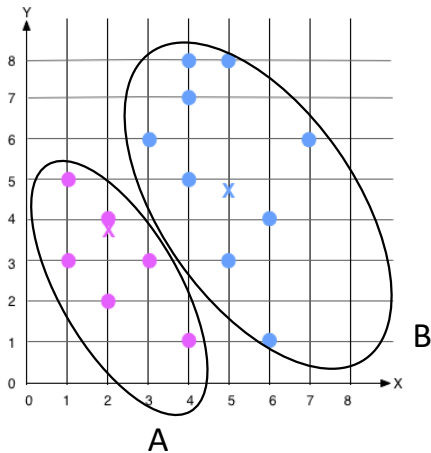
| Evergreen? | Needle length | Needle arrangement? | Tree type |
|------------|---------------|---------------------|-----------|
| No | Short | Tufted | Tamarak |
| No | Long | Tufted | Larch |
| Yes | Short | Single | Hemlock |
| Yes | Short | Bundled | Pine |
| | Long | Single | Fir |



Grading: 1 point for discussion of entropy, 2 points for tree

Problem 8: k-means clustering [6 marks] [practice exercise]

Consider the following diagram which shows a set of points (the circles), initial centroids (the x's), and cluster assignments (as shown by the ovals). For any necessary math, you may leave your answer as the expression rather than by doing the calculations, e.g., if your answer is $2^{10} = 1024$, you may leave your answer as 2^{10}



- 1) [4 marks] Choose the new centroids for Clusters A and B

Cluster A's X value:

$$\frac{(1+1+2+2+3+4)}{6}$$

Cluster A's Y value:

$$\frac{(1+2+3+3+4+5)}{6}$$

Cluster B's X value:

$$\frac{(3+4+4+4+5+5+6+6+7)}{9}$$

Cluster B's Y value:

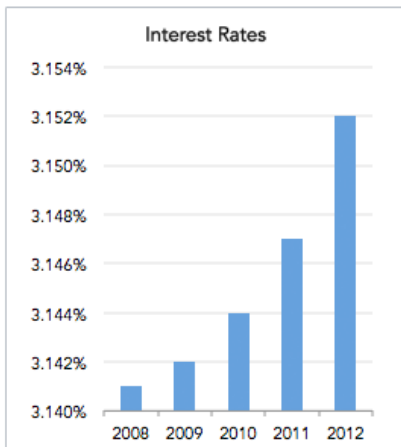
$$\frac{(1+3+4+5+6+6+7+8+8)}{9}$$

- 2) [2 marks] Will you need to recalculate the centroids again? (i.e., how do you know when you can stop?) If you need additional information, what information you need?

In order for the k-means algorithm to continue, it would need to be the case that (1) there are points that move between clusters and (2) we haven't reached our maximum number of iterations. In this case, you would not need to, but you would need to know that.

Problem 9: Visualization [8 marks]

- 1) [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/#>

- 2) [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

- 3) [2 marks] If you want to find out how you did relative to your classmates in an exam (i.e., which percentile you were in), 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.

- 4) [2 marks] What are two factors that we covered in class that you should consider to determine whether an infographic or statistic is misleading? Note: we will only grade the first two factors that you list.

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]

Commented [RAP1]: I'd skip this one if I were you

Problem 10: Artificial Intelligence [5 Marks]

- 1) [1 mark] Searle wrote that "A fair number of researchers in artificial intelligence (AI) believe [...] that by designing the right programs with the right inputs and outputs, they are literally creating minds."

Does Searle consider such researchers to be proponents of Strong AI or Weak AI? Circle one.

Strong AI

Weak AI

(See page 1 of Searle reading.)

- 2) [1 mark] True or False: With his "Chinese room argument", Searle aimed to establish that it is impossible to create artificial thinking systems.

True

False

A: False. The assigned article (top of page 5) quotes Searle as saying that "we might even come to be able to create thinking systems artificially."

- 3) [1 mark] Describe a difference that you observed in the responses of Eliza and Cleverbot. Be as concrete as possible.

- 4) [1 mark] Do you think that Turing would consider Cleverbot to be intelligent? Explain your answer.

Answer: Yes or No could be ok here, depending on whether the student thinks that Cleverbot could pass the Turing test, i.e., "play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning."

- 5) [1 mark] Do you think that Strong AI would consider that Cleverbot understands? Explain your answer.

Answer: No. Cleverbot simply manipulates symbols, which is not understanding.

Problem 11: Artificial Intelligence Impact [3 marks]

- 1) [1 mark] What is an argument from lecture 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 intelligence can create jobs as well, balancing out the jobs that it automates

- 2) [2 marks] Consider the 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

Is it likely that this job will be automated in the next 20 years? Why?

Yes, very likely because most of these tasks can already be automated. In fact, this particular job has already lost more than 43,000 positions from 2002 to 2014, a 16% decline.

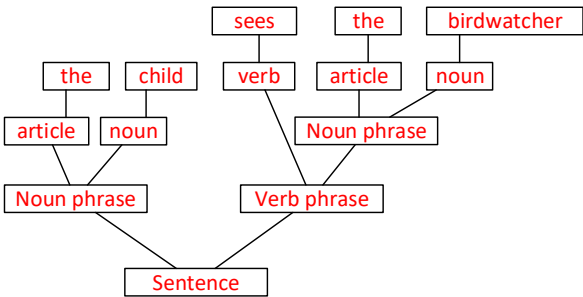
Problem 12: Natural Language Processing [7 marks]

| Lexicon | | Grammar | |
|-------------|-------------|---------------------|-----------------------------------|
| binoculars | Noun | Sentence | → NounPhrase, VerbPhrase |
| birdwatcher | Noun | NounPhrase | → Article, Noun |
| child | Noun | NounPhrase | → NounPhrase, PrepositionalPhrase |
| sees | Verb | PrepositionalPhrase | → Preposition, NounPhrase |
| the | Article | VerbPhrase | → Verb, NounPhrase |
| with | Preposition | VerbPhrase | → Verb |

1) [1 mark] What is parsing?

Breaking a sentence into its parts and describing their syntactic roles

2) [2 marks] Using the lexicon and grammar above, draw a parse tree representing a possible parsing of the sentence “The child sees the birdwatcher.”



3) [1 mark] What steps in natural language processing can help to resolve ambiguity that may arise in parsing? Circle one answer.

A. Semantic analysis.

B. Pragmatics.

C. Both semantic analysis and pragmatics.

4) [1 mark] Give an example of a way in which parsing was successfully used by Watson to answer Jeopardy questions.

Possible answers:

- When clues contain multiple facts about the answer, parsing helps break the clue into subclues, the answers to which can be searched independently
- Parsing helps determine When one subclue is nested in an outer clue.

- Finding "lexical answer types" (LATs) in clues: A LAT is word in the clue that indicates the type of the answer (independent of assigning semantics to that word).

5) [1 mark] Describe a key difference between traditional NLP approaches versus machine learning approaches (such as artificial neural networks) to natural language processing.

A: One possible answer: traditional approaches use formal rules such as parsing to make sense of language, while machine learning approaches learn from examples using models such as artificial neural networks.

6) [1 mark] What is an example of an adversarial attack in the context of machine learning models (such as neural networks)?

Answer (AI lecture slides): Adversarial examples are inputs to machine learning models that an attacker has intentionally designed to cause the model to make a mistake; they're like optical illusions for machines.

Problem 13: To infinity... and beyond! [3 marks]

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

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

2) [1 mark] What is a potential technical advantage to using 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

3) [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 would also accept that there are only a small number of bits or that it takes too much power.