CSE 20, Fall 2020 - Homework 1

Due: Monday 10/12 at 11am PDT

Instructions

Upload a single file to Gradescope for each group. All group members' names and PIDs should be on each page of the submission. Your assignments in this class will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should always explain how you arrived at your conclusions, using mathematically sound reasoning. Whether you use formal proof techniques or write a more informal argument for why something is true, your answers should always be well-supported. Your goal should be to convince the reader that your results and methods are sound.

Reading Rosen Sections 2.1, 4.1, 4.2 and 5.3

Key Concepts Set Definitions, Division Algorithm, Algorithms, Tracing Pseudocodes for Different Inputs, Base Expansion and Conversions

Problem 1 (20 points)

In this question we think about recommendation systems. A key concept about recommendation system is clustering, which refers to the process of grouping data with similar patterns together. For example, each YouTube user's watching/liking/disliking history can be represented as an n-tuple indicating their tastes of videos. Users with similar tastes can then be clustered to provide future recommendation for one another. Mathematically, clustering is based on the notation of distance between pairs of n-tuples.

Consider n = 5. Define the following functions whose inputs are ordered pairs of 5-tuples whose components come from the set {0, 1, 2}.

$$d_{max,5}((x_1, x_2, x_3, x_4, x_5), (y_1, y_2, y_3, y_4, y_5)) = \max_{1 \le i \le 5} |x_i - y_i|$$

$$d_{rmse,5}((x_1, x_2, x_3, x_4, x_5), (y_1, y_2, y_3, y_4, y_5)) = \sqrt{\frac{\sum_{i=1}^{5}(x_i - y_i)^2}{5}}$$

(a) Give a recursive definition of the max function whose input is a sequence of integers and whose output is the maximum value in the sequence. Include a description of the domain and codomain of the function, along with the basis step and the recursive step of the function definition.

(b) Repeat (a) for the RMSE function.
(c) Provide possible examples input for which the function has a preset value.
    (i) Give an example input to $d_{max,5}$ for which the output of the function is 0. Is your example the only possible example? Why?
    (ii) Give an example input to $d_{max,5}$ for which the output of the function is 1. Is your example the only possible example? Why?
    (iii) Give an example input to $d_{rmse,5}$ for which the output of the function is 0. Is your example the only possible example? Why?
    (iv) Give an example input to $d_{rmse,5}$ for which the output of the function is 2. Is your example the only possible example? Why?

Problem 2 (20 points)

DNA is made up of strands of four different bases that match up in specific ways. The bases are elements of the set B = {A, C, G, T}.

Definition The set of DNA strands S is defined (recursively) by:

Basis Step: $A \in S, C \in S, G \in S, T \in S$

Recursive Step: If $s \in S$ and $b \in B$, then $sb \in S$

A function $dnalen$ that computes the length of DNA strands in S is defined by:

$$rnalen: S \rightarrow \mathrm{Z}^{+}$$

Basis Step: If $b \in B$ then $dnalen(b) = 1$

Recursive Step: If $s \in S$ and $b \in B$, then $dnalen(sb) = 1 + dnalen(s)$

Each of the sets below is described using set builder notation. Rewrite them using the roster method. For example, the set described in set builder notation as

$\{s \in S$ | the leftmost base in s is T and $dnalen(s) = 2\}$

is described using the roster method by

$$\{TA, TC, TG, TT\}$$

Please solve the 2 problems below:

(a) $\{s \in S$ |the leftmost base in s is the same as the rightmost base in s, the second-from-left base of s is not A and is not C, and $dnalen(s) = 4\}$
(b) $\{s \in S$ |s has at most one T and there are twice as many As as Gs in S, and $dnalen(s) = 4\}$

Problem 3 (20 points)

We focus on **The Division Algorithm** (Rosen 4.1 Theorem 2, p. 239) in this problem:

Let n be an integer and d a positive integer. There are unique integers q and r, with $0 \leq r < d$, such that $n = dq + r$. In this case, d is called the divisor, n is called the dividend, q is called the quotient, and r is called the remainder. We write $q = n \textbf{ div } d$ and $r = n \textbf{ mod } d$.

In display industry, any color can be mathematically represented as a 3-tuple (r, g, b) where r, g, b represent the red, green and blue component respectively. Each of (r, g, b) must be from the collection $\{x \in N \mid 0 \leq x \leq 255\}$.

(a) Convert the number expressed in base 16 as $(FF)_{16}$ to binary, octal and decimal. Justify your answer in a clear way that anyone who has just taken the course can easily follow. What is the color (x, x, x) where x =$(80)_{16}$? You can use a web color tool (include the URL of the tool you use as part of your assignment write-up) .

(b) Suppose you were told that the colors that will work best for your web app are (r, g, b) where $r \textbf{ mod } 16 = 6$ and $g \textbf{ div } 16 = 7$ and $b \textbf{ mod } 16 = 8$. Give three distinct examples of such colors. For each example, specify its red, green, and blue components both in decimal and in hexadecimal. Justify your answer in a clear way that anyone who has just taken the course can easily follow.

Problem 4 (20 points)

Consider the following pseudocode:

```
procedure f1 (m: real number; n: positive integer)
total := 0
a := m
b := n
while b > 1
    if (b mod 2 = 1) then total := total + a
    a := 2 · a
    b := b div 2
return total + a
```

(a) Trace through an example of this pseudocode when $m = 312$ and $n = 32$. It might be helpful to create a table showing the value of each variable after each step. What does this algorithm do?

(b) Repeat part(a) for $m = 32$ and $n = 512$. What is the difference between this and (a)? Given 2 integers x and y, can you come up with a general principle for choosing which number to pick as m and which number to pick as$n$?
(c) Convert $m = 312$ and $n = 32$ to binary.
(d) Multiply the binary representation of m and $n$using the standard long-form multiplication technique that you use for decimal multiplication.
(e) Convert your answer from (d) to decimal and compare it with your answer from (a). Intuitively, how do you think this algorithm corresponds to the long multiplication algorithm?

Problem 5 (20 points)

In class, we have dealt with several examples of recursive definitions of sets. These will form a basis of several algorithms that you will use throughout this class and in the future. Recursive definitions involve 2 steps: the Basis Step and Recursive step.

(a) Let's consider a recursive definition for the set S_4 of positive integer multiples of 4.

Basis Step: If $x \in$ __________, then $x \in S_4$
Recursive Step: If $x \in S_4$and $y \in S_4$, then ______________ $\in S_4$

(b) Let's say that you now want to make a set Z_4 of all (positive and negative) multiples of 4. Modify the recursive definition given above that will allow you to do this.

(c) Similar to how we define a recursive set, we can also define a recursive algorithm for computing certain functions.
Let's try and construct a recursive algorithm to calculate the factorial of a number. The factorial (defined as the product of all integers from 1 to n) of a number of n is defined as:

$$n! = n \cdot (n-1) \cdot (n-2) \ldots 3 \cdot 2 \cdot 1$$

1. **procedure** $factorial$(n: non-negative integer)
2. **if** ___________ then **return** 1
3. **else return** ___________

Problem 6 (Bonus: 10 points)

The base-14 expansion of a number n has d digits. Find an upper bound on the number of bits in the binary expansion of n.