15-110 Recitation Week 14

Reminders

- HW6 due TOMORROW 12/08 at Noon! No Revisions!
    - Make sure your code is running on the autograder and that you stop by OH if it is timing out/breaking
    - Double check the writeup and videos to ensure that your code passes the manual grading portions; check your edge cases!

- Final Exam:
    - andrewIDs A-I: Rashid Auditorium (GHC 4401)
    - andrewIDs J-Z: McConomy Auditorium (CUC)
    - 8:30-11:30 am
    - Monday, December 11
    - Review sessions:
        - Friday 12/08 at 5pm in DH 2210 on Recursion, Efficiency, Encryption, and Tractability
        - Sunday 12/10 at 3pm in GHC 4401 on Helper Functions, Data Analysis, and Simulation
    - TAs will hold small groups
    - You can bring your note sheets: 20 pages

Overview

- Runtime/Big-O and Tractability

- Recursion

- Simulation - MVC and Monte Carlo

- Data Analysis

- Machine Learning and AI

Problems

Big-O and Tractability

What is the Big-O of the following mystery function? List the Big-O of each line in the function to help you find the overall runtime

```
def mystery(matrix): # matrix is an n by 10 2D list of integers
    total = 0
    for row in matrix:
        row.sort() #sort runs in nlogn time
        for elem in row[::2]:
            total += elem
    return total
```

Overall Runtime:

Consider the following function:

```
def roundOne(n): # n is an integer which has value n
    newN = n**2
    result = []
    for i in range(0, newN, n):
        print(i * n)
        for j in range(i):
            result.append(i * j)
    return result
```

What is the Big-O complexity of this function if we only count the **print statements**?

What is the Big-O complexity of this function if we only count **appending to result**?

Some basic Tractability/P vs. NP questions:

What are the Useful NP Problems?

If you try to solve the subset sum problem by finding the sum of all subsets of the input, is this tractable or intractable? Explain.

If someone gives you a potential solution to the subset sum problem, is verifying whether the solution is correct tractable or intractable?

What complexity class is the subset sum problem in and why?

Recursive Functions

Function 1:

Write a recursive function that returns all capital letters in a string in reverse order of appearance

```
#reverseCaps("EnjoY a rEaLlY Long wintEr brEak") returns "EELYLEYE"
```

Function 2:

Write a recursive function `evaluate(numbers, operations)` that takes in a list of numbers of size n and a list of operations of size n - 1 and recursively evaluates the expression `numbers[0] operations[0] numbers[1] … operations[n-1] numbers[n]`.

For example, if `numbers = [5, 6, 4]` and `operations = ["+", "-"]` then `evaluate(numbers, operations)` will return 5 + 6 - 4 = 7. Note the only operations you will get are "+" and "-" so you don't have to worry about order of operations. You can assume numbers will always be at least length 1 and operations will always be the length of numbers minus 1.

Recursive Family Trees

Write a recursive function `getOldestAncestor(tree)` that takes in a family tree and returns the name and age of the oldest ancestor of the person represented by the root node as a list. A family tree can be thought of as
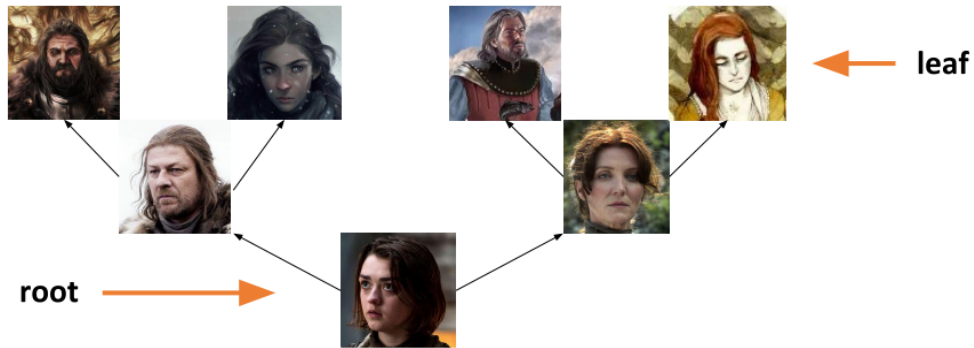
where the root represents a child and the leaves represent their oldest ancestors. You are guaranteed that the oldest ancestor in a given tree will be in a leaf node. The "value" key for a given node stores a list where the 0th index is the name of the person and the 1st index is the age of the person.

Simulation - Model, View, Controller

1. Imagine we have the following parts of a simulation:
   Set the starting position and radius of a circle and draw it in the center of the canvas.
   > What functions would we use for this?

   The circle moves horizontally left and right across the canvas based on time (when it reaches one end it switches directions)
   > What functions do we need?

2. Now go to your starter file where we have provided this simulation! Let's add the following elements
   Inside keyPressed function:
   - If we hit the return/enter key, we make the circle stop or start moving
   - If we hit the up arrow key, we move the circle up the canvas by 10 pixels
   - If we hit the down arrow key, we move the circle down the canvas by 10 pixels

   Recall that in keyPressed(data, event) we can use event.keysym to get the "name" for characters we can't show in strings
   > Enter/Return: event.keysym == "Return"
   > Up arrow key: event.keysym == "Up"
   > Down arrow key: event.keysym == "Down"

Simulation - Monte Carlo

Write a Monte Carlo Simulation to compute how long you have to listen to songs on your Spotify playlist until you hit "Headlines" by Drake. The runTrial function you write should take in a list of 2 element lists, where the 0th index of the inner list is the song name and the 1st index of the inner list is the length of the song in seconds. This 2D list of songs will be stored in a variable playlist in your getExpectedValue function. Remember that on each trial, the order of the songs in the playlist will be randomly shuffled. You can use the random.shuffle function to help you!

Data Analysis

Given the following types of data, name the best visualization method to display that data:

A. Numerical ✗ Ordinal
B. Numerical
C. Ordinal ✗ Ordinal
D. Categorical
E. Numerical ✗ Numerical ✗ Numerical

Write a function geneExpression(filename, sample) that takes in a text filename and a sample number and creates a bar plot visualization of the gene expression levels for the corresponding sample. First, read the file corresponding to the filename passed in. This file is several lines ("\n") of percent expression levels separated by tab characters ("\t"). For each string percentage in a line, convert the string percentage into a float and add it to a temporary list. Then for each line, add the temporary list to an overall list, creating a 2D list of gene expression percentages. Each row will correspond to a specific gene and each column will correspond to a specific sample. Then, plot the expression level for each gene for the input sample using the matplotlib bar plot function. On the x-axis, set the values to the string "Gene x" for every index x and on the y-axis, let the values be the expression percentage for each gene for the given sample.

Machine Learning and Artificial Intelligence

Why do we use validation data in the model creation process?

What is an advantage and a disadvantage of using a heuristic?

Categorize each of the following as either a classification, regression, or clustering problem:
1. Based on a set of symptoms, determine what illness a patient has.
2. Group a set of pictures into three groups, with similar pictures being in the same group
3. Using the number of people who show up to a movie premiere to predict how much money the movie will make

From this connect-four board, draw the next level of the game tree for the red player.
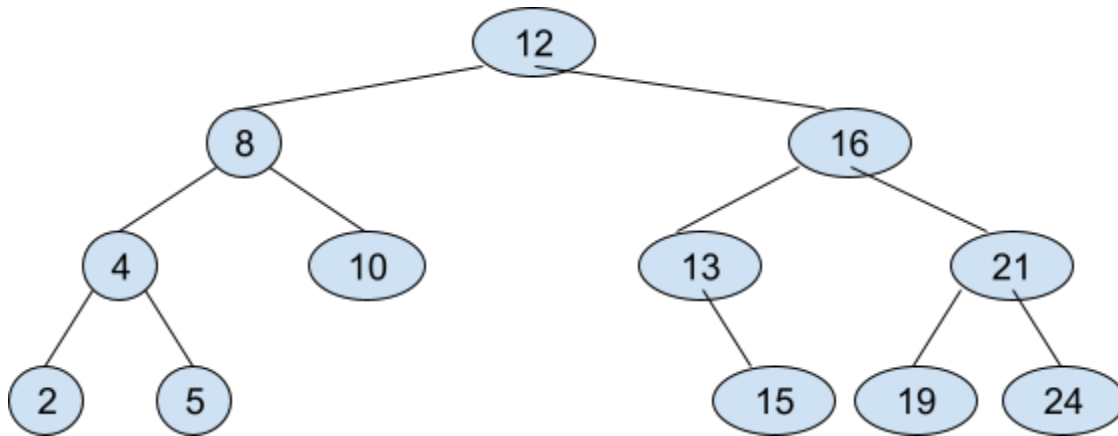(If you aren't familiar with connect-four, you can read about it here!
http://www.ludoteka.com/connect-4.html#:~:text=The%20aim%20for%20both%20players,be%20vertical%2C%20horizontal%20or%20diagonal.&text=Before%20starting%2C%20players%20decide%20randomly,made%20alternatively%2C%20one%20by%20turn)
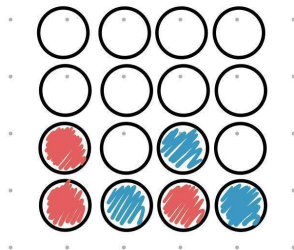
Search Algorithms

Perform binary search looking for the following elements in the BST. Write down all elements that are searched until the element in question is found or not.

15:

4:

18:

9: