CSCI 3104 Spring 2022
Instructor: Profs. Chen and Layer

Midterm 1 Standard 4 - Huffman Coding

Due Date ..... February 27
Name ..... **Julia Troni**
Student ID **109280095**

Contents

1 Instructions

- The solutions **should be typed**, using proper mathematical notation. We cannot accept hand-written solutions. Here's a short intro to LaTeX.
- You should submit your work through the **class Canvas page** only. Please submit one PDF file, compiled using this LaTeX template.
- You may not need a full page for your solutions; pagebreaks are there to help Gradescope automatically find where each problem is. Even if you do not attempt every problem, please submit this document with no fewer pages than the blank template (or Gradescope has issues with it).
- You **may not collaborate with other students**. **Copying from any source is an Honor Code violation. Furthermore, all submissions must be in your own words and reflect your understanding of the material.** If there is any confusion about this policy, it is your responsibility to clarify before the due date.
- Posting to **any** service including, but not limited to Chegg, Discord, Reddit, StackExchange, etc., for help on an assignment is a violation of the Honor Code.

2 Standard 4 - Huffman Coding

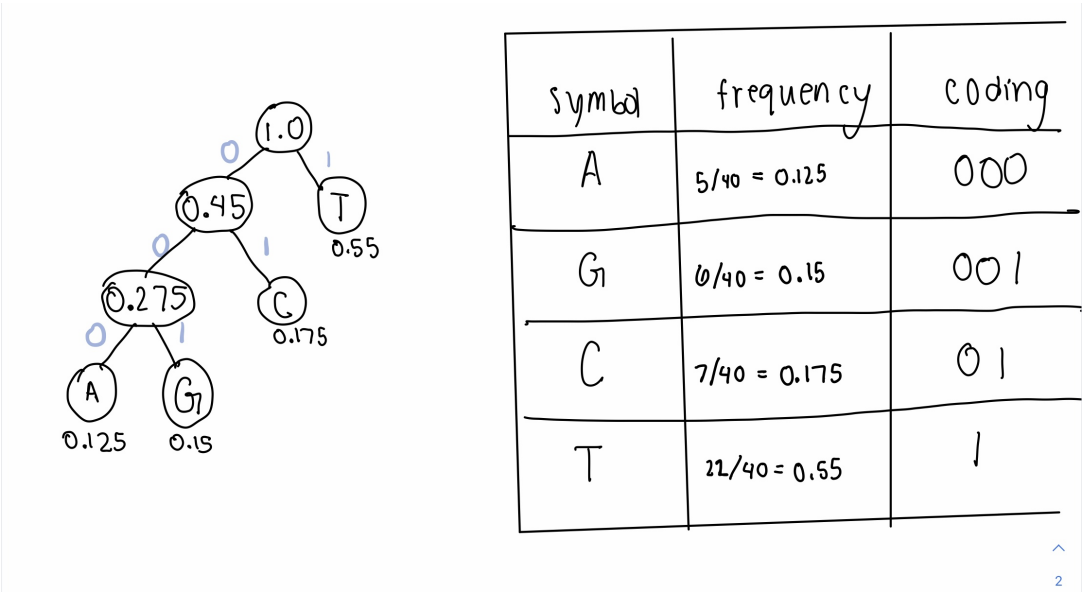
2.1 Problem 1

Problem 1. Give the Huffman encoding for the first 40 nucleotides of of COVID-19 spike gene:

```
ATGTTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTC
```

You need to first show the optimal binary tree you construct, and then write down the corresponding codes.

Answer. The optimal binary tree is as follows



And the corresponding codes are
A: 000
G: 001
C: 01
T: 1

And the encoding is
0001001111001111110111001111100011001010100001100000110110110000001101 □