FIT3155: Week 10 tutorial
Covering concepts from Weeks 8-9

Objectives: The tutorials, in general, give practice in problem solving, in analysis of algorithms and data-structures, and in logic useful in the above.

Instructions to the class: Prepare your answers to the questions **before** the tutorial. It will probably not be possible to cover all questions unless the class has prepared them all in advance.

Instructions to Tutors:

i. The purpose of the tutorials is not to solve the practical exercises.

ii. The purpose is to check answers, and to discuss particular sticking points, not to simply make answers available.

1. Design a Huffman code for a set of characters $\{a, b, c, d, e\}$ with their respective probabilities of $0.4, 0.2, 0.2, 0.1, 0.1$.

2. In the exercise above, you would have noticed that when performing greedy merging of two nodes/subtrees (of lowest probabilities in the current iteration), we encounter multiple choices (ways) to achieve this, and our strategy was to pick any two among the multiple choices rather *arbitrarily*.

   A useful **variation** of Huffman coding is to *always* merge **two shortest (in height) subtrees** whenever a multiple choice exists. This gives what is called the *minimum variance* Huffman coding.

   Design a minimum variance Huffman code on the set of characters and probabilities provided above. Compare (i.e., eyeball) the resultant code words between the original method and the this variant.

3. Does Huffman encoding always yield a prefix-free code words for the characters it is encoding? If yes, why? If no, why not?

4. Encode the following sequence into $\langle$`offset`, `length`, `character`$\rangle$ triples using LZ77 algorithm:

   $$barrayar\_bar\_by\_barrayar\_bay$$

   Assume the search window size and lookahead buffer sizes are both 15.

5. Decode the encoded triples from above, to recover back the original text.

6. A text is encoded using the LZ77 algorithm, which yields the following sequence of triples:
   $\langle 0, 0, r\rangle$ $\langle 0, 0, a\rangle$ $\langle 0, 0, t\rangle$ $\langle 2, 8, \_\rangle$ $\langle 3, 1, \_\rangle$ $\langle 0, 0, r\rangle$ $\langle 6, 4, t\rangle$ $\langle 9, 5, t\rangle$

   Assume the sizes of the search window and lookahead buffer are both 10.

7. Encode $N = 12345$ using the Elias variable-length encoding of integers.

8. Decode the above encoding to recover back $N = 12345$.

9. Encode, using the Elias variable-length encoding of integers, following sequence of integers:
   $$123, 100, 1, 23, 561$$

10. Concatenate the variable length binary encodings of the above sequence, and decode this bit string to recover back the full sequence of integers.

```
-=o0o=-
  END
-=o0o=-
```