CSE 484 Final Exam

Olcay Taner YILDIZ

I. Question (12 points)

The k-Means clustering algorithm randomly selects an original partition into clusters and deterministically rearranges clusters afterward. Design a randomized version of the k-Means algorithm that uses coin tossing to rearrange clusters.

II. Question (15 points)

Let us say we have 2 states:
S_1: α, S_2: β with equal initial probabilities.
and the transition matrix is:

$$\mathbf{A} = \begin{vmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{vmatrix}$$

The emission probabilities for S_1 is:

$$\mathbf{e}(S_1) = \begin{vmatrix} \text{A: } \frac{2}{5} \\ \text{G: } \frac{2}{5} \\ \text{T: } \frac{1}{10} \\ \text{C: } \frac{1}{10} \end{vmatrix}$$

The emission probabilities for S_2 is:

$$\mathbf{e}(S_2) = \begin{vmatrix} \text{A: } \frac{1}{5} \\ \text{G: } \frac{1}{5} \\ \text{T: } \frac{3}{10} \\ \text{C: } \frac{3}{10} \end{vmatrix}$$

Decode the most likely sequence of states for sequence GGCT.

III. Question (16 points)

Determine the number of different ways to partition a set of n elements into k clusters for the cases below (solve 2 cases):

- Clusters have names, clusters can be empty and elements do not have names
- Clusters have names, clusters can be empty and elements have names
- Clusters have names, clusters can not be empty and elements do not have names
- Clusters do not have names, clusters can be empty and elements have names

IV. Question (16 points)

Design two algorithms for finding the shortest string in text t_1 that does not appear in text t_2.

a) The first algorithm will have $O(n^3)$ time complexity and compares substrings of two texts one by one.
b) The second algorithm will have $O(n^2)$ time complexity and uses Suffix trees.

V. Question (15 points)

What is the optimal local alignment for APPLE and HAPPE? Show all optimal local alignments and the corresponding paths under the match premium +1, mismatch penalty -1, and indel penalty -1.

VI. Question (14 points)

Perform the BREAKPOINTREVERSALSORT algorithm with π = 5 3 1 7 8 6 4 2 and show all intermediate breakpoints.

VII. Question (12 points)

Use a) the Hamiltonian path and b) Eulerian path approach to solve the SBH problem for the following spectrum:

$$S = \{AGT, AAA, ACT, AAC, CTT, GTA, TTT, TAA\}$$

Label edges and vertices of the graph, and give all possible sequences s such that Spectrum(s, 3) = S.