

CSE 582 Midterm Exam

Olcay Taner YILDIZ

I. QUESTION

Find the optimal pairwise global alignment of the sequences TACGAGTACGA and ACTGACGACTGAC with the condition that **G** nucleotides shown in bold font must be aligned together. The scoring parameters are defined as +2 for match, -1 for mismatch, and $d = -2$ for a linear gap penalty.

II. QUESTION

Prove that the running time of the DP algorithm for optimal pairwise alignment of two sequences of lengths n with a gap penalty function of a general form is $O(n^3)$.

III. QUESTION

In the herpesvirus genome, nucleotides C, G, A, and T occur with frequencies 35/100, 35/100, 15/100, and 15/100 respectively. Assuming the independence model for the genome, what is the probability that a randomly selected 15 nucleotide long DNA fragment contains eight C's or G's and seven A's or T's?

IV. QUESTION

Suppose we consider two independence models of nucleotide sequence. The first model, M_1 , has the same probabilities of nucleotides as defined in Problem 1.6. The second model, M_2 , assigns to each nucleotide type the probability 1/4 to appear in any given position. Given the observed sequence $x = \text{ACTGACGACTGAC}$, compare the likelihoods of these models.

V. QUESTION

Write an algorithm that, given an n -element multiset, generates all m -element subsets of this set. For example, the set $\{1, 2, 2, 3\}$ has four two-element subsets $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, and $\{2, 2\}$. How long will your algorithm take to run?

VI. QUESTION

How many permutations on n elements have a single breakpoint? How many permutations have exactly two breakpoints? How many permutations have exactly three breakpoints?

VII. QUESTION

The SBH problem is to reconstruct a DNA sequence from its l -mer composition

- Suppose that instead of a single target DNA fragment, we have two target DNA fragments and we simultaneously analyze both of them with a universal DNA array. Give a precise formulation of the resulting problem (something like the formulation of the SBH problem).
- Give an approach to the above problem which resembles the Hamiltonian Path approach to SBH.
- Give an approach to the above problem which resembles the Eulerian Path approach to SBH.