

Exercise 6

Deadlines: Tuesday 2008.10.07 (copy) and Tuesday 2008.10.14 (corrected)

PROBLEM 1

The Viterbi algorithm is a deterministic algorithm for solving the Decoding problem. Design a randomized algorithm for solving the Decoding problem that start from a randomly chosen hidden path and tries to improve it using coin tossing. Provide pseudo-code.

Can you say anything about the time complexity of your algorithm?

PROBLEM 2

a) Design a greedy algorithm for doing pair-wise global sequence alignment. Assume you have a scoring matrix δ available. Provide pseudo-code.

What is the time complexity of your algorithm?

b) Extend you algorithm in a) to alignment of multiple sequences. You do not have to provide pseudo-code, only a short explanation.

What is the time complexity of your algorithm?

PROBLEM 3

The following motif finding algorithm takes as input a $t \times n$ table \mathbf{DNA} containing t sequences of length n , and outputs a set of start positions (one for each sequence) that defines a set of motifs of length l . To evaluate the set of motifs, the algorithm uses $\text{Score}(\mathbf{s}, \mathbf{DNA})$ that returns the number of matches between the motifs given by the start positions in \mathbf{s} and the corresponding consensus motif. $\text{Score}(\mathbf{s}, i, \mathbf{DNA})$ returns the partial score for the i first sequences. The parameter *random* is a number between 1 and n .

```
YetAnotherMotifSearch( $\mathbf{DNA}$ ,  $t$ ,  $n$ ,  $l$ , random)
1  bestMotif  $\leftarrow (1, 1, \dots, 1)$ 
2  bestScore  $\leftarrow 0$ 
3  for start  $\leftarrow 1$  to  $n - l + 1$ 
4     $\mathbf{s} \leftarrow (1, 1, \dots, 1)$ 
5    srandom  $\leftarrow$  start
6    for  $i \leftarrow 1$  to  $t$ 
7      if  $i \neq \textit{random}$ 
8        bestScoreStart  $\leftarrow 0$ 
9        for  $s_i \leftarrow 1$  to  $n - l + 1$ 
10       if  $\text{Score}(\mathbf{s}, i, \mathbf{DNA}) > \textit{bestScoreStart}$ 
11         bestScoreStart  $\leftarrow \text{Score}(\mathbf{s}, i, \mathbf{DNA})$ 
12         bestPosition  $\leftarrow s_i$ 
13        $s_i \leftarrow \textit{bestPosition}$ 
14     if bestScoreStart  $>$  bestScore
15       bestScore  $\leftarrow \textit{bestScoreStart}$ 
16     bestMotif  $\leftarrow \mathbf{s}$ 
17  return bestMotif
```

- What is the general idea behind this algorithm (explain shortly what the algorithm does)?
- To which of the algorithm design techniques we have covered in this course would you classify YetAnotherMotifSearch? Why?
- What can you say about the running time of YetAnotherMotifSearch? Explain.
- How would you use the parameter *random* to create results that are more reliable?

PROBLEM 4

We have seen many different algorithms for the motif finding problem. Design your own algorithm that is different from any of the one we have discussed so far. An explanation of your idea is enough (you don't need to provide pseudo-code). What type of algorithm design method is your algorithm an example of? What is the time complexity of your algorithm?