Midterm Exam

CS 4364/5364

Spring 2022

Name: _____

Instructions: Please read all of the instructions below before you begin:

- Read all of the questions in the exam before you begin.
- Questions marked with a dagger (†) are required for students in CS 5364, and bonus (optional) for those in CS 4364
- Figures (pictograms) can be included if they help describe a solution, and are encouraged if they are clear.
- Remember that unanswered questions will receive 0 credit, any reasonable try at a response will receive at least half-credit. If you feel you're unable to provide a reasonable answer to a question, you can answer with "I cannot provide a reasonable attempt for this question", which will be provided quarter-credit.
- Students are allowed to bring any written material with them to the exam, but no electronic devices can be used during the exam.
- if you need extra room, please use the *back of the previous page* rather than the current one (when it is folded open your whole answer should be showing).

question/part	score	4364	5364
1a		/10	/10
1b		/5	/5
1c		/10	/10
$1d^{\dagger}$		(bonus)	/10
2a		/10	/10
2b		/5	/5
2c		/5	/5
3a		/12	/12
$3b^{\dagger}$		(bonus)	/10
4a		/6	/6
4b		/5	/5
4c		/4	/4
5†		(bonus)	/10
6		/15	/15
Total		/82	/117

		Т	Т	Т	A	Т	G
	0	←-1	←-2	←-3	←-4	←-5	←-6
Т	↑-1	<u>~</u> 8	≿7	$\overset{\nwarrow}{\rightarrow} 6$	$\leftarrow 5$	<u>~</u> 4	$\leftarrow 3$
С	↑-2	$\uparrow 7$	$\stackrel{\uparrow}{\leftarrow} 6$	$\stackrel{\uparrow}{\leftarrow} 5$	$\stackrel{\uparrow}{\leftarrow} 4$	$\stackrel{\uparrow}{\leftarrow} 3$	$\stackrel{\uparrow}{\leftarrow} 2$
Т	↑-3	\uparrow 6	5	<u></u> ∼14	$\leftarrow 13$	<u></u> ∑12	←11
A	↑-4	<u>†</u> 5	<u>†</u> 14	$\stackrel{\uparrow}{\leftarrow} 13$	×22	$\leftarrow 21$	←20
Т	↑-5	\uparrow 4	\uparrow 13	×22	$\stackrel{\uparrow}{\leftarrow} 21$	₹ 30	$\leftarrow 29$

1. Use the Needleman-Wunch dynamic programming table for S = TTTATG and T = TCTAT below to the next questions.

- (a) (10 pts) How many co-optimal alignments of the two strings are there? What are they?
- (b) (5 pts) What is the optimal alignment of S[1....3] and T[1...5]? (note these are prefixes TTT and TCTAT)
- (c) (10 pts) What is the indel penalty used to construct the table? match score?
- (d) †(10 pts/3 pts bonus) Can you determine the mismatch penalty from the table above? Explain why or why not.

- 2. Below is an algorithm description for a given problem (not defined on purpose).
 - assume you are given a string $S = s_1 s_2 s_3 \dots s_n$.
 - set $S^R = s_n s_{n-1} s_{n-3} \dots s_1$, that is S^R is the reverse of S.
 - construct a new string $T = S \$ S^R$, where $\$ \notin \Sigma$.
 - compute the maximum prefix overlap, $M_i(T)$ for each $2 \le i \le (2n+1)$.
 - return $M_{n+2}(T) == n$.
 - (a) (10 pts) What is returned when (i)S = ABCBA? (ii)S = ABCDBA?
 - (b) (5 pts) What is the running time of the algorithm (in terms of n, the length of S)?
 - (c) (5 pts) What does this algorithm do?

3. Given the alignments below, determine the number of matches, mismatches, gaps, and indels; place the counts in the table. (2 pt each)

		Matches	Mismatches	Gaps	Indels
		mt	\mathbf{ms}	gp	id
A_1	GCT-AC-TTGC				
	-CTG-CA-TGC				
A_2	GCTACTTGC				
	-CTGCATGC				
A_3	GCTACTTGC				
	-CTGCATGC				

 $\dagger(10 \text{ pts}/3 \text{ pts bonus})$ Determine at set of values for α, β, γ , and δ such that each of the alignments A_3 is optimal under the following scoring:

 $\alpha \mathrm{mt} + \beta \mathrm{ms} + \gamma \mathrm{gp} + \delta \mathrm{id}$

assuming A_1, A_2 , and A_3 are the only possible alignments of the two sequences.



4. Use the generalized suffix tree below to answer the next questions.



Figure 1: Suffix tree for question 4

- (a) (6 pts) What are the 3 strings encoded in the tree?
- (b) (5 pts) What is the longest string contained in **at least two** strings?
- (c) (4 pts) Name two possible alphabets that could be used to produce the strings used to construct the strings?

5. †
(10 pts/3 pts bonus) True or False: For global pairwise alignment of two strings of size
 m and n

$$2\mathbf{mt} + 2\mathbf{ms} + \mathbf{id} + \mathbf{gp} = (m+n).$$

Where mt is the count of matches, ms is the count of mismatches, and id is the number of indels, and gp is the number of gaps. Explain your answer.

- 6. (15 pts) Given two sequences S and T (not necessarily the same length), let G and L be the scores of an optimal global alignment and an optimal local alignment, respectively.
 - (a) True or false $\forall S, T \in \Sigma^* : L \ge G$. Explain your answer.