HMM (MAF)

14. Assume an dinucleotide-HMM where in a DNA sequence

$X_i = \text{nucleotide } i \text{ and nucleotide } i+1$

That is, X_1 is nucleotide 1 and 2, X_2 is nucleotide 2 and 3 and so on. Assume further that the only possible dinucleotide sequences in the model are {AA,AG,AT,CA,GA,TC}.

a) Draw a schematic picture of the state space showing only the possible transitions.

[2p]

b) Which of the following sequences could not be generated by this HMM: CAAGAT, AGATAG, GAGATC, TCAAAG?

[1p]

15. Both the Viterbi and the forward algorithm use dynamic programming to relate the observed sequence to the state sequence. For both the Viterbi and the forward algorithm, answer the following questions (label clearly which answer belongs to which algorithm):

a) What are we trying to compute with each algorithm? Describe with a short sentence and with a probability expression.

[3p]

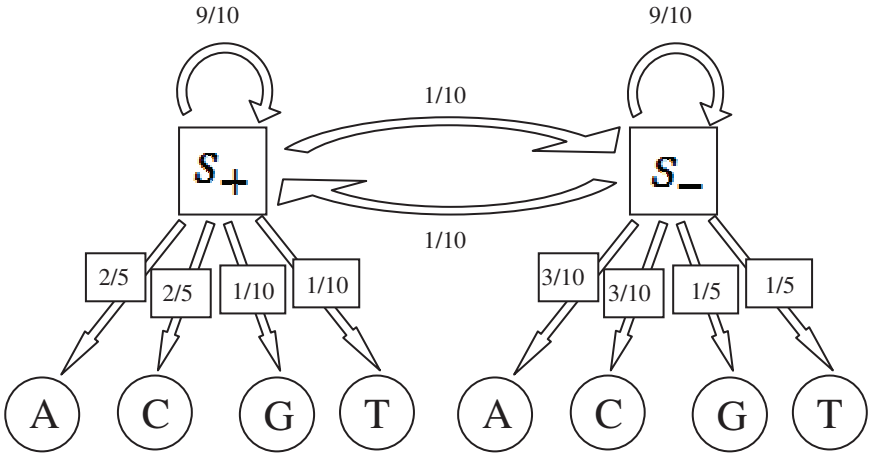
b) What does the recurrence relation look like for each algorithm?

[2p]

c) Briefly describe the traceback for each algorithm, including where it starts and ends.

[2p]

16. Assume that we have a two-state HMM as illustrated in the figure, for instance corresponding to CpG-island and non-CpG-island.



(a) Describe the log-odds ratio and how it can be used for classification.

[2p]

(b) Using, the Viterbi algorithm, what is the most likely sequence of states for the sequence GTCT?

[4p]