CSE 582 Final Exam

Olcay Taner YILDIZ

I. Question

Hidden Markov models can be used in algorithms of protein secondary structure prediction. One rather straightforward approach uses the secondary structure conformations α-helix, β-strand, and turn as the hidden states emitting observable amino acids. It is assumed that the frequencies of appearance of each of twenty aminoacids in either conformation have been determined from analysis of proteins with the three-dimensional structures known from the experiment. Draw the state of the HMM and describe the Viterbi algorithm that could be used for predicting the protein secondary structure.

II. Question

A 4200 nt long DNA sequence is used as a training set for parameter estimation of the DNA statistical model. The observed counts of sixteen dinucleotides, N_{XY}, are as follows.

	T	C	A	G
T	510	380	210	190
C	240	170	360	230
A	370	200	220	210
G	190	170	220	220

Find the maximum likelihood estimates of (a) the transition probabilities P_{TT}, P_{AG} of the first order Markov model of the DNA sequence; (b) the transition probabilities P_{TT}, P_{AG} of the first order Markov model for the DNA sequence complementary to the given training sequence.

III. Question

The pair HM shown in Figure 4.2, generates two aligned DNA sequences x and y. State M emits aligned pairs of nucleotides with emission probabilities $P_{x_i y_j}$ defined as follows:

$$P_{TT} = P_{CC} = P_{AA} = P_{GG} = 0.5$$
$$P_{CT} = P_{AG} = 0.05$$
$$P_{AT} = P_{GC} = 0.3$$
$$P_{GT} = P_{AC} = 0.15$$

The insert states X and Y emit symbols from sequences x and y respectively. The emission probabilities are the same for both insert states:

$$q_A = q_C = q_G = q_T = 0.25$$

No symbols are emitted by begin and end states. The values of transition probabilities are as follows:

$$\delta = 0.2$$
$$\tau = 0.1$$
$$\epsilon = 0.1$$

Use the Viterbi algorithm for pair HMM to find the optimal alignment of DNA sequences x = TTACG and y = TAG.

IV. Question

Estimate the parameters of a profile HMM for the following multiple alignment of DNA sequences:

G	C	A	G
G	-	-	G
G	-	A	G
G	C	T	G
A	-	A	C
G	-	A	C
G	-	G	G
A	-	A	C

Draw the state diagram.

V. Question

Solve exercise 7.6.