CSE182 Lecture 8 questions. HMMs

Vineet Bafna

October 30, 2007

The questions are open ended, but should help you understand lectures better. Do these questions make sense? Are they helpful in following the lecture? Constructive feedback is appreciated.

1. Consider a profile P of length m. For any position i, residue a, $P_i[a]$ is the score (not the frequency) of aligning residue a to the i-th position of the profile. Describe an algorithm that finds the highest scoring local alignment of a sequence $s[1..n]$ allowing for gaps and with an indel score δ.

2. Consider the "loaded coin" question posed in L8. Suppose your friend uses a 'loaded' coin in which the probability of HEADS is 1. Assume also that he will switch the two coins with probability 0.3.

   (a) Describe an HMM that models the string of coin tosses

   (b) Suppose you see the following sequence $HTHTHTTTTTHHTTHHTH$. Compute the maximum likelihood probability of number of times he cheated.

3. See L8. Construct an HMM for the 3 sequences in the family using the red ovals as match states. Compute the probability that ALIL is a member of the family using both the forward and the viterbi algorithm. (You can make up numbers for transition probabilities).

4. What is a gene?

5. What is the amino-acid sequence corresponding to the following DNA code? ATGGCCCGCAAATGACUA.

6. What do the following variants of BLAST do: blastx, tblastn, tblastx?