

BMI/CS 776
Spring 2009
HW #1

(1) Consider using an HMM to model sequences that have either zero or one occurrence of a particular motif. Suppose that the motif is three bases long, the sequences are of arbitrary length, and the motif can occur at any position in a given sequence. Draw the topology of an HMM for this task. You do not need to show any emission or transition probabilities, but clearly show the states in the model (including begin, end states) and all of the transitions.

(2) We can think of the MEME ZOOPS model as an HMM. Draw a diagram of the HMM structure for such a model for the case in which all sequences are five bases long and the motif width $W=3$. Show all states, transitions, and transition probabilities.

(3) The Gibbs sampling method for motif finding we discussed makes the OOPs assumption. Show how you would modify the algorithm pseudocode in order to implement the ZOOPS model.

(4) Suppose we are learning a CRM-modeling HMM, using the method of Noto and Craven. At some point in the search, suppose our model is a conjunction of two motifs, X and Y (i.e. X **AND** Y).

- a) Draw the HMM topology. Show all of the states and the transitions.
- b) Show how the HMM topology would be changed after applying the **OR** operator. Show the result for all possible applications of the **OR** operator.

(5) Consider the following alignment:

AAGCG-C
AT--GTC

- a) Write out the expression describing how it would be scored using a global alignment algorithm with an affine gap penalty (as described on p. 29 in Durbin et al.),
- b) Write out the expression describing how it would be scored using a pairwise HMM (as described in the lecture notes).
- c) Are these two methods equivalent? If not, indicate how they differ.

(6) Which of following HMM pieces could be represented exactly using a semi-Markov model?
 Assume that the emission probabilities will be set to the fixed values shown, but we want to learn arbitrary transition probabilities. Justify your answer for each.

