BMI/CS 776
Exam Practice Questions

(1) 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.

(2) 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).

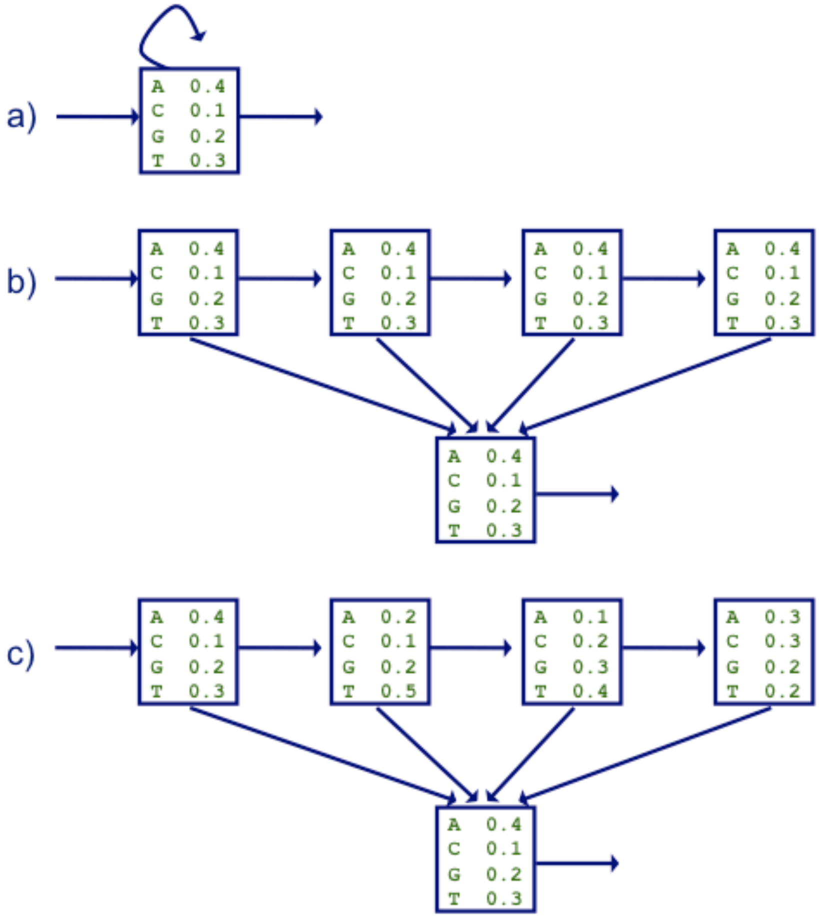
a) Draw the HMM topology. Show all of the states and the transitions.
b) Show how the HMM topology would be changed after applying the **OR** operator. Show the result for all possible applications of the **OR** operator.

(3) 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.

(4) 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.



(5) For the following SCFG, show all of the possible parse trees for the sequence **cggcg**. Which parse is the most probable one? Note that the probability of each production is shown in parentheses following it.

S → B S	(0.1)
S → B	(0.1)
S → **g** S **c**	(0.4)
S → **c** S **g**	(0.4)
B → **c**	(0.5)
B → **g**	(0.5)

(6) Show how the Inside algorithm would determine the probability of the string **ccgg** under the following grammar.

S → D D	(0.3)
S → B T	(0.7)
T → B D	(0.6)
T → D B	(0.4)
D → B B	(1.0)
B → **c**	(0.5)
B → **g**	(0.5)

(7) Consider a simple threading problem in which we have a template with three segments (*i, j, k*). We are given a sequence for which there are two possible starting positions for each segment.

Given the following values for the scores of the individual segments and the scores for segment interactions, show how the branch-and-bound method would find the optimal threading.

$g_1(i, 2) = 4$	$g_1(j, 8) = 2$	$g_1(k, 13) = 1$
$g_1(i, 3) = 3$	$g_1(j, 9) = 5$	$g_1(k, 14) = 10$

$g_2(i, j, 2, 8) = 6$	$g_2(i, k, 2, 13) = 1$	$g_2(j, k, 8, 13) = 3$
$g_2(i, j, 2, 9) = 0$	$g_2(i, k, 2, 14) = 0$	$g_2(j, k, 8, 14) = 12$
$g_2(i, j, 3, 8) = 1$	$g_2(i, k, 3, 13) = 9$	$g_2(j, k, 9, 13) = 5$
$g_2(i, j, 3, 9) = 0$	$g_2(i, k, 3, 14) = 0$	$g_2(j, k, 9, 14) = 11$

Use the "simple lower bound" presented in class. When splitting a threading, split the segment having the minimal g_1 value for some position (e.g. split on *k* first since $g_1(k, 13) = 1$). To split a selected segment, divide it into two intervals of length one.