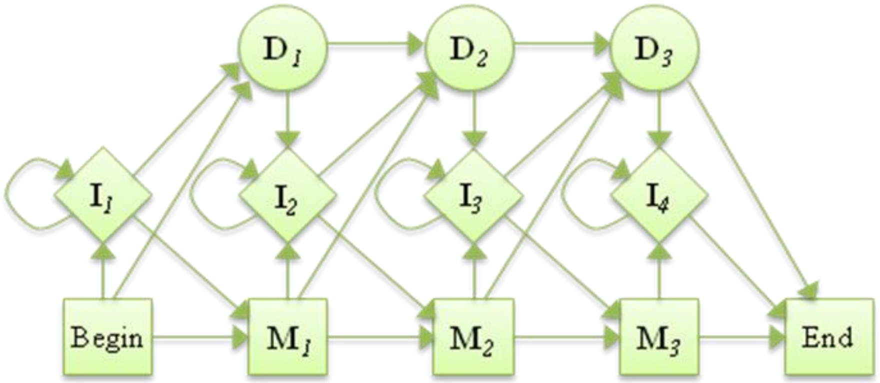
CENG 465
Fall 2013-2014

Due Date: November 29, 2013 via COW or hardcopy in class

Assignment #2

(Written Assignment)

Problem: Consider the following profile Hidden Markov Model (HMM) constructed for a small DNA sequence motif.



$\pi_{Begin} = 1.0$, $\pi_{M1} = \pi_{M2} = \pi_{M3} = \pi_{End} = 0.0$

Transition probabilities:

	Begin	**M_1**	**M_2**	**M_3**	**I_1**	**I_2**	**I_3**	**I_4**	**D_1**	**D_2**	**D_3**	**End**
Begin	0	**0.7**	0	0	**0.2**	0	0	0	**0.1**	0	0	0
M_1	0	0	**0.8**	0	0	**0.1**	0	0	0	**0.1**	0	0
M_2	0	0	0	**0.5**	0	0	**0.4**	0	0	0	**0.1**	0
M_3	0	0	0	0	0	0	0	**0.2**	0	0	0	**0.8**
I_1	0	**0.6**	0	0	**0.2**	0	0	0	**0.2**	0	0	0
I_2	0	0	**0.4**	0	0	**0.4**	0	0	0	**0.2**	0	0
I_3	0	0	0	**0.5**	0	0	**0.4**	0	0	0	**0.1**	0
I_4	0	0	0	0	0	0	0	**0.3**	0	0	0	**0.7**
D_1	0	0	0	0	0	**0.5**	0	0	0	**0.5**	0	0
D_2	0	0	0	0	0	0	**0.5**	0	0	0	**0.5**	0
D_3	0	0	0	0	0	0	0	**0.6**	0	0	0	**0.4**
End	0	0	0	0	0	0	0	0	0	0	0	0

Note: Each row sums up to 1, except the row for the End state.

Emission probabilities:

M_1: $\omega_A = 0.1$, $\omega_C = 0.7$, $\omega_G = 0.1$, $\omega_T = 0.1$

M_2: $\omega_A = 0.1$, $\omega_C = 0.1$, $\omega_G = 0.7$, $\omega_T = 0.1$

M_3: $\omega_A = 0.7$, $\omega_C = 0.1$, $\omega_G = 0.1$, $\omega_T = 0.1$

I_1, I_2, I_3, I_4: $\omega_A = 0.25$, $\omega_C = 0.25$, $\omega_G = 0.25$, $\omega_T = 0.25$

Note: Begin, End, and delete states do not emit any symbols.

Align the following sequence to the profile HMM using the Viterbi algorithm. In other words find the sequence of states which is most likely to emit the following DNA sequence. Show the contents of the partial probability table you construct. What is the final probability associated with this path?

Note: Initialize the partial probability table with v_{Begin}(“”)=1.0. Do not use log of probabilites. You may use some programming or tools like Excel to help in your calculations.

`TAACTTGAGT`

Submission:

Submit your solution via COW or submit a hardcopy on November 29. You solution may be handwritten or written on computer.