

Bioinformatics Computing
CSE40532/60532
Homework #3

Reading:

1. Read Chapters 5 and 6 (week of 10/7)
2. Read select journal article handouts (week of 10/7)

Problems: (due 10/7)

1. Compute the global and end-gap free alignment of the two mitochondrial sequences from HW #1: NC_001807 and AF254446. Use an affine gap penalty (and the associated additional matrices) instead of a constant penalty. Traceback is optional, but you must report the scores for each type of alignment obtained using the following scoring scheme used (+4 match, -2 mismatch, -3 to open a gap and -1 to extend a gap) in your write up. Please also submit the code in your drop box (10 points).
2. Implement the dishonest casino Hidden Markov Model (HMM) described on page 54 of Durbin. Assume the casino is always fair at the start, and shifts to being loaded with probability 0.05 as described (i.e., at $t=0$, $\Pr(F) = 1$ & $\Pr(L) = 0$). Generate a random sequence of 300 rolls based on this model. Please submit the code and instructions how to run in your drop box; a particular instance of 300 rolls in the drop box is optional (5 points).
3. Download the two particular instances available from the course website.
4. Determine the probability of the benchmark files given the dishonest casino model of #2 using an implementation of the Forward algorithm. Please report the probabilities in your report and submit your source code in the drop box with instructions how to run it (10 points).
5. Determine the most likely state sequence of the benchmark files given the dishonest casino model of #2 using an implementation of the Viterbi algorithm. Please save the result as files "viterbi.1.txt" and "viterbi.2.txt" and label the states "F" and "L" as used in Figure 3.5 in Durbin. Please submit these result files and the source code in your drop boxes (10 points).