Noah M. Jorgenson

Bioinformatics

Lab 8

1. When running the viterbi.py python script with verbose=True in the viterbi function, we are informed of the time step in the observation sequence and the probabilities at that timestep. With verbose set to False we instead see the final probability with the observed and predicted sequences. Verbose would be useful if we wanted to see the probabilities part way through the observation sequence. T is a dictionary with key values corresponding to states and the matching value is a tuple containing the probability of that state, the sequence up to and including that state and the total probability of that sequence. The output of the algorithm is showing us the sequence with the highest final probability. It predicts that the next day will be rainy. Adding ‘walk’ to the sequence, causes the prediction to shift to sunny since walking is correlated with a high chance of it being sunny outside.
2. The HMM I constructed was with the exon, 5’ splice site and introns as the hidden states and the ACGT nucleotides in the observed DNA sequence were the observed sequence. Using the given probabilities to determine the emission probabilities and the state transition probabilities given in the diagram. The result was that given the sequence it never quite reached the ‘end’ but it did happen to get to the 5’ splice site. The final sequence of hidden states was:

['E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', 'E', '5p', 'I', 'I', 'I', 'I', 'I', 'I', 'I', 'I']

1. The total probability as you add more and more states is naturally going to approach very small values or unlikely probabilities. To remedy this problem we can leverage the log value of that probability and use a specialized python class to hold the decimal values such as numpy.Decimal. In order to predict the secondary structure states of a protein sequence with a HMM, we’d need an observed sequence of proteins and the probability that a residue to appear, or maybe group of residues, would be in a particular structure (alpha helix, beta sheet or coil).