

Group Members: Jinghan Lu, Tianyuan Fu

We changed our topic of option 4 to detecting the CpG islands for the prediction of latter nucleic acid based on a given sequence using two Hidden Markov Models, the Forward and Viterbi algorithms. So far, we have implemented the I/O and calculated the probability of “A,” “C,” “G,” “T” occurring in the dataset and the transition probability of 16 combinations between “A,” “C,” “G,” “T.” For the input data, we chose the genetic information of the chromosome 1 of homo sapiens. Since the whole sequence is large, we chose 85,000 among the whole sequence to speed up the testing. Since we regard the input data as the state sequence, we have implemented the algorithm for the disjoint probability and transition probability. However, our group are still a bit confused about the definition of observations in our case. Further work would be proceeded as the definition is clarified.