CS494/594
Bioinformatics computing

Challenge #1

1. In your groups of 3-4 students, make a random DNA sequence (string) of size 12-16 characters. Try and make it more complicated by using all four options: A, G, C, T and include it below (shared Google Doc) on a zoom white board

2. Reverse complement your sequence from #1

3. What are the multinomial model probabilities using the sequence from #1 as your training sequence?

   P(A) = P(T) = P(C) = P(G) =

4. Using your model from #3, compute the probability of AGCTCG

5. What are the Markov order 1 probabilities using the sequence from #1 as your training sequence? For example, if our training string was only ATC, P(T|A) = 1, P(C|T) = 1, and all other 14 probabilities would be zero.

6. As asked in the lecture #3 video, derive an alignment of:

   VIVALASVEGAS
   VIVADAVIS

7. Perform a simple "by eye" alignment of the two strings on slide #14/lecture 3 using the same scoring system, i.e., matches get +1 and all mismatches/gaps get a -1. Report the score of your "by eye" alignment. The strings are "CATCAC" and "CTCCAGC"