

Special Topics in Data Science: Algorithms in Computational Biology

specialtopics.deblasiolab.org/s22/

We will through the duration of the semester examine common algorithmic solutions to domain specific data science problems, and how to distill computational problems from questions asked in other domains (i.e. computational biology). While the specific applications we will use as examples are in biology, the approaches discussed are applicable to many interdisciplinary fields. That said, this course is self contained and no previous knowledge of biology is needed. The solution techniques include dynamic programming, computational optimization/integer linear programming, and machine learning to name a few.

Some of the specific biological problems to be discussed are:

- * **pairwise/multiple sequence alignment** which has applications not only in biology but in aligning time-series data.
- * **hashing and sketching** which has its roots in web page similarity measurement and plagiarism detection but is now used for genome assembly.
- * **genome assembly** which uses techniques developed for other purposes such as database searching.
- * **sequence database searching** using tools like BLAST which have applications to non-biological databases.
- * **phylogenomics**, that is the study of finding ancestry from sequences, which can be used in exploring unknown history of things like viruses (both biological and digital).

While these are some of the topics we will discuss they are in no way exhaustive of the field; as with previous versions of the course I am open to suggestions of topics of interest to the students enrolled. The only prerequisite is CS 3 (CS 2302), only the minimum amount of biology will be included to understand the underlying question and the extraction of the computational problem, but it will be self contained and none is expected ahead of time. This semester this class is cross listed as CS 4364/5364 CRNs 27123/28040.

I am always happy to discuss any questions anyone has about the course, dfdeblasio@utep.edu.