

MATH 5610, Project Guidelines and Ideas

The project for this course involves doing in-depth work on some topic related to computational biology. Below is a list of possible topics. You are free to choose from this list or come up with your own project idea. Whatever you decide to do, your project should involve gaining in-depth knowledge about some area of bioinformatics as well as some creative thinking on your part. You can use the textbook as a starting point for learning about your topic, but you should also plan to read journal articles.

There are three deliverables for your project. Guidelines for these deliverables are given below:

- Project Proposal: Due March 8 before class.
- Draft Report: Due Friday, April 19 before class. (Must be double spaced).
- Final Report: Due Friday, May 6 at noon.

1 Ideas for Projects

Programming Projects Find a description of an algorithm from a recent journal article. Implement the algorithm in a programming language of your choice (I suggest MATLAB). Compare your program to an existing package, or to other algorithms in the literature. Discuss the strengths and limitations of the algorithm. Comment on the size of problems you can tackle with the algorithm. Here are some possible topics for programming projects:

1. EM-algorithm for training a Profile Hidden Markov Model for sequence alignment.
2. Sequence Alignment with Tandem Duplication.
3. Sequence Database search algorithms: BLAST, PSI-BLAST, FASTA, CLUSTAL W, or any of their variants.
4. Genetic Algorithms (or other heuristics) for multiple sequence alignment.
5. Algorithms for fragment assembly.

6. Physical Mapping of DNA.
7. Reconstructing phylogenetic trees based on pairwise distances, neighbor joining, maximum parsimony, or maximum likelihood estimation.
8. Gene Prediction—identifying regions of DNA that encode proteins.
9. Protein Structure Prediction.
10. RNA Secondary Structure Prediction.
11. Clustering gene expression data.

Survey Papers A survey paper involves reading several (5 or more) papers on a single topic. Give detailed summaries of the papers, and compare how they differ. Comment on ideas for possible extensions. Possible topics for survey papers:

1. Protein visualization tools
2. Probabilistic Graphical Models (applied to bioinformatic problems)
3. Proteomics
4. Ligand docking
5. Post-translational Modification Prediction

2 Project Proposal

Your project proposal should clearly state what you are planning to do. Be careful not to promise too much. (I suggest proposing something modest, with possible extensions if time permits). If you are doing a programming project, state the programming language you intend to use, as well as the paper(s) you are basing your project on. If you are doing a survey paper, identify papers you intend to read.

3 Draft Report

The Draft Report should follow the same format as the final report (see below).

4 Final Report

Your final report should be prepared using LaTeX using 12 point font. The following is a suggested organization for the report

1. title page (use `\maketitle`)
2. abstract
3. introduction—your goal here is to communicate what the report is about so that the reader knows whether or not to read the report.
4. background – review the relevant literature.
5. description of your work—use several sections here with appropriate titles to describe what you did, how you collected data, how you analyzed results, etc.
6. summary/conclusions
7. Appendices
 - (a) Code listing.
 - (b) Instructions on how to use your program.
 - (c) others.
8. Bibliography (use BibTeX to create your bibliography).