STAT 337-437 (QUANTITATIVE BIOINFORMATICS)

PROJECT GUIDELINES

Goal: This project is intended to demonstrate to students the widespread use and applicability of probabilistic and statistical methods in bioinformatics practice and research. It permits students to observe the uses, abuses and limitations of some of the techniques discussed in class. Lastly, it gives the students an opportunity to explore their originality and apply statistical methods to real data.

Step 1: Find an Article for your Project

Students are to find a published journal article which uses sophisticated probability or statistical methods preferably in the broad field of bioinformatics. The article that you choose can neither be "too easy" nor "too difficult", and each group is responsible for shortlisting several articles to make sure the chosen one(s) is/are at the right level. Note that interesting research articles can be obtained via the Web – either through Loyola's Library e-Journals or from the Archives pages of most journals themselves. Discuss as a group, what would you like to work on? A project on cancer? Genetic diseases? Modelling data using a clustering technique? Search to the best of your ability and send me choices (3 or 4 articles) within 2 weeks from today, by **March 25th**. If you cannot find anything in your topic of your interest, let me know and I will send you some options.

Step 2: Read and Understand the Article

This should be done as a group and will help you in the next steps of the project. Remember I am always here to help!

Step 3: Find a Dataset to Apply the methods you learned

This step may be done in tandem with Step 2 and involves applying the methods of the paper on a real dataset. There are several data resources available in the web (Eg: https://web.stanford.edu/~hastie/ElemStatLearn/). There are tons of datasets available in R too! Try to finalize this early on as you will present it in your project.

Step 4: Let's get creative (We are statisticians! We can think ahead!)

Here, students are expected to go one step further and perform some original analysis on the chosen dataset. This may be anything and maybe discussed with the professor.

Step 5: Lightning Presentations

The last week of class 04/27, 04/29 would be scheduled for short presentations, where each group would present their findings (as a group) to the class. Each group will have around 10 minutes for their presentation.

Step 6: Project Report

Each group would also turn in a report (max 5 pages) of their project. The objective here is to provide an outline and critique of the probability/statistical methods of your selected article. This would be followed by a description of your application on real data and original analysis. You may make sections such as 'Introduction'/ 'Background', 'Methods' 'New Analysis' 'Discussion' 'Conclusion' etc. (these are suggestions only). Please include your code at the end of your project report and attach a copy of your chosen article separately for easy and fair grading. For fair grading, clearly outline the contribution to this project by each individual member at the end of the report.

Grading: I will grade on choice of article (20), application on dataset (50) (based on presentation + report) and original analysis (30) (based on presentation + report). Groups of size 2 will be weighed accordingly.

Dates: Here is a timeline for the project

03/25/2021: Last date to get your chosen article approved. It is the responsibility of each group to email me with options/questions/discussion etc. so we can finalize an article that meets the criteria for this project.

04/27/2021, 04/29/2021: Short Class presentations

04/30/2021: All reports are due on this date by 5 PM.