University of Wisconsin-Madison
BMI/CS 776: Advanced Bioinformatics
Prof. Anthony Gitter

Spring 2018
Project ideas

Overview

The goals of the class project are to gain experience working on a (small) bioinformatics research problem using real biological data. In order to encourage creativity and ambitious projects, you will not be graded based on whether or not your idea works. Grades will be based on whether your project was completed, thoughtful, executed carefully, and built upon sound computational and biological principles. Contributions to the bioinformatics research community (e.g. releasing your new method as open source software; sharing trained models; or contributing patches, new features, or documentation to existing software) will be viewed favorably.

In contrast to the homework, you are encouraged to use existing software and packages for the project and may use any programming language(s). You are required to reference all data sources, software, and other resources in your project reports. Reusing code, text, or images without proper attribution is unacceptable and will be treated as academic misconduct.

Each student will implement their own project individually. However, multiple students may elect to work on the same topic as long as they work independently. You do not need to run your code on the biostat server, but your project must be reproducible. This means that you must provide the source code and sufficient documentation for running your code.

Types of projects include:

- Applied projects where one or more computational methods are run on a single dataset of interest and the evaluation focuses on the biological interpretation of the results.
- Benchmarking projects where multiple computational approaches are run on multiple datasets to assess their relative strengths.
- A computational project where a new algorithm is implemented or an existing algorithm is extended and compared with the original.
- An extension of your current research, in which you use a dataset that you are already studying but analyze it with new types of computational methods. *The project should not be something that you were already planning to do as part of your research.*

Specific ideas

- BMI/CS 776 Spring 2015 project ideas https://www.biostat.wisc.edu/bmi776/spring-15/project.html
- BMI/CS 776 Spring 2016 project ideas https://www.biostat.wisc.edu/bmi776/spring-16/project_ideas.pdf
- BMI/CS 776 Spring 2017 project ideas https://www.biostat.wisc.edu/bmi776/spring-17/project_ideas.pdf
- Implement a solution to a closed DREAM challenge http://dreamchallenges.org/project-list/closed/ Be sure to only use the training data when building your model, and choose a challenge for which the test set is already available so that you can evaluate your performance. Avoid reimplementing methods that have already been described in the challenge summary papers.

- Create a GenomeSpace recipe and associated tool to teach users how to conduct a particular bioinformatics analysis. http://www.genomespace.org/blog/2016/02/01/announcing-the-new-genomespace-recipe-resource An example recipe is http://recipes.genomespace.org/view/8
- Participate in a current CAMDA Contest Challenge http://contest.camda.info/
- Extend an existing tool for deep learning in genomics. Convolutional or recurrent neural networks for regulatory genomics data and autoencoders for gene expression data are a few specific ideas, but there are many possibilities. http://kipoi.org/ provides pre-trained models that you can modify or use for benchmarking. https://greenelab.github.io/deep-review/ describes approaches for a wide variety of biological problems. However, there are extremely limited GPU resources available on campus for training neural networks.
- Benchmark single-cell RNA-Seq trajectory inference algorithms, building upon the tools and assessments described in https://doi.org/10.1101/276907
- Run multiple network algorithms to predict pathways from genome-wide datasets and compare the results. Potential network algorithms include:
    - ResponseNet, which we will learn in class
    - Forest from Omics Integrator https://github.com/fraenkel-lab/OmicsIntegrator
    - PathLinker https://github.com/Murali-group/PathLinker
    - TieDIE https://github.com/epaull/TieDIE
- Benchmark multiple algorithms for reconstructing gene-gene networks from gene expression data.