BENG 420 Bioinformatics for Engineers
Fall 2018
Course Project Description

The purpose of the individual course project is to apply the material learnt in class lectures, homework assignments and readings to solve a biomedical or public health problem. It is expected that you will formulate a hypothesis relevant to the problem you pick and test it using data mining approaches. You can either implement a data mining method from recent literature or compare several standard approaches to examine their performance on the same dataset. You can implement your program in Matlab or other programming languages you prefer. You will test methods using either real data or simulated data. Looking for proper biomedical data is non-trivial so before you finalize your topic in the proposal, please make sure you have found the data already when you write your proposal. The emphasis should be on analysis and interpretation, and validation of the choice of method(s) used and any underlying assumptions.

Evaluation: The individual project is worth 25% of the grade. The project will be evaluated based on (1) a final project report, (2) project presentation, and (3) several intermediate reports.

Final Project Report: The final project report should be formatted as a two-column, 4-page IEEE conference paper, with appropriate references in IEEE format. The template file can be found here:
https://www.ieee.org/conferences_events/conferences/publishing/templates.html

The report should have the following four mandatory sections:

1. Introduction: This section should clearly describe the biomedical problem that is being addressed in this project, the significance of the problem for public health or basic science, and the benefit of using a data mining approach to understand this problem. The hypothesis needs to be clearly stated, including applicable alternate hypotheses. This section should also include a brief review what has been done previously with appropriate references.
2. Methods: This section should clearly describe the data mining methods that were used, a description of the data analysis method that was developed with appropriate equations and flow diagrams, a description of the validation method that was used to evaluate the results. Any improvement of modification over previous methods needs to be specified. The data used should be described in details including data source, dataset size, attributes etc. This section should demonstrate that the student successfully find data for analysis, understands the data mining techniques involved and any statistical methods used to interpret the results.

3. Results: This section should describe the results of the data analysis. Graphs, tables and figures should be properly labeled and captioned.
4. Discussion and Summary: This section should interpret the results of the analysis and connect them back to the introduction. The discussion should clearly address the following questions:
    a. How would you compare your results to previous studies? Did you come up with a conclusion agreeable with existing studies? If not, explain what might be the reasons.
    b. What did you learn from the results? Were the results expected or unexpected based on your judgment? Why?
    c. What are the limitations of the method and analysis? What simplification and assumption did you use?
    d. How do the results improve your understanding of the biomedical problem described in the introduction?

Presentation: A project presentation summarizing the problem you addressed in your introduction, your methods, your main results, and a brief discussion. Your presentation should be about 8 minutes. Presentations will be given in a mini-conference format on 12/3 and 12/5. Presentation dates will be available to sign up at least a week prior to your presentation. This portion of the project is worth 5% of your total grade.

Deliverables: Students will develop sections of the project report throughout the semester, submit cumulative reports on intermediate due date and receive feedback. The following is the schedule for the deliverables.

Due date	**Deliverable**	**Grade percentage**
9/24	Submit a one-page project proposal, including a clearly articulated statement of what exactly is the hypothesis you are trying to examine, what the expected results might be and a review of related work in the literature, including a properly formatted bibliography.	4%
10/10	Submit properly formatted Introduction and Methods sections.	5%
11/19	Submit first draft of Results section and revisions to Introduction and Methods.	5%
12/3 & 12/5	Presentation	5%
12/10	Final project report and source code due with all four sections including all revisions.	6%

Some online biomedical data repositories (feel free to use data from other sources):

http://www.physionet.org/

http://www.oecd.org/els/health-systems/health-data.htm
http://datam.i2r.a-star.edu.sg/datasets/krbd/
https://archive.ics.uci.edu/ml/datasets.html
http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html