Research Project Guidelines

PB HLTH 290
Targeted Learning with Biomedical Big Data
Spring 2019

The final project is your opportunity to apply what you have learned in this class to a real-world problem. During the first few weeks of class, you will identify a research project and select a small group of 2 to 3 people to work with. Consider a problem involving biomedical big data and apply the Targeted Learning framework as that has been the focus of this course. You may critically evaluate a previously investigated problem/publication or perform your own novel analysis. If you chose to critically evaluate a previously investigated problem/publication, it is preferred that you support your evaluation by simulations or obtain the data if it is available. Explicitly and thoughtfully apply the Targeted Learning framework to this problem. This does not mean you must have a perfect data analysis/critical evaluation. A good project, rather, is one in which you have clearly communicated the possible issues at each step, have done your best given your training so far, and have identified limitations and next steps.

The link to sign up for groups: here. Email the teaching staff if you have any trouble. Please form groups by February 10.

Project Presentation I – Proposal

Wednesday, February 20

This presentation is intended to introduce the class to your proposed project. The presentation should last a maximum of 5 minutes and all group members should speak. Expect questions after your presentation. To guide the class through your proposal, please use slides. During this presentation each group should:

- Give some background about a real-world problem that inspired your project.
- Briefly describe the data you will work with or simulate, including key variables.
- Define the proposed target parameter(s) you want to estimate and obtain inference for. Justify their importance/relevance.
- List any anticipated challenges and explain how you might address them.
- Introduce your group members and discuss each members' planned contribution.

Project Presentation II – Progress

Wednesday, April 03

This presentation is intended to discuss the progress of your project. The presentation should last a maximum of 5 minutes and all group members should speak. Expect questions after your presentation. To guide the class through your presentation, please use slides. During this presentation each group should:

- Describe what each group member has done since the last presentation (e.g., written code, ran the simulation study).
- Provide tangible evidence of their progress (e.g., figures, tables, EDA).
- List future developments and justify the importance of them.
- List issues that have arisen and explain how they might be addressed.
- Present specific areas where guidance is needed.

Project Presentation III – Comprehensive

Tuesday, May 07 – Thursday, May 09

This presentation is highlights the application of your education in this course. The series of presentations during the term are intended to improve the ultimate quality of your final presentation. The presentation should last about 30 minutes and all group members should speak. Expect questions during and after your presentation. During this presentation each group should:

- Have equal group participation.
- Define the statistical estimation problem. Additionally, give some background about why your question is interesting/important/relevant.
- Specify a valid loss function and candidate estimators. Provide justification.
- Obtain initial estimate with super learner. Provide results and interpret them correctly.
- Update initial fit using TMLE. Show that the score of your sub-model generates the canonical gradient.
- Provide Inference. What method did you use to calculate the standard error for the estimator of the target parameter?
- Interpret results. What additional assumptions are needed to interpret the target parameter as a causal parameter? Are they reasonable?
- Discuss limitations and next steps. How do your results fit with the existing literature? How do the limitations affect the interpretation or quality of the data? What new directions are supported by your findings?
- State the overall conclusion and impact of your study. What are the main "take-home" messages?
- Make effective use of tables, figures, and use clear notation.
- Explain the analysis clearly and justify your approach.