Project Description
Bios 835-Fall 2023

Empirical evidence suggests that experiential learning, or learning through direct experience, is more efficacious than mere didactic instruction. The origins of this principle are nebulous, but its perpetuation throughout historical contexts suggests its enduring relevance. Drawing from personal observations as a parent to three children, there is a clear disparity in knowledge acquired through verbal directives as opposed to lived experiences. For instance, repetitive insistence regarding neat eating habits may yield minimal behavioral adjustments. However, a single peer-driven event, such as negative feedback during lunch, can catalyze profound and enduring behavioral transformations.

The goal of this project is to further explore some aspects of the machine learning methods covered in the course. Any machine learning topic is legitimate with the approval of the course instructor. Your project must fit in either (or both) of the following two forms:

1. A thorough analyses of a challenging dataset with machine learning methods. For this type of project, justifying the need for machine learning methods will be critical. The dataset does not need to be original. Prediction/classification cannot be the only outcome of the analyses. One of the following much be addressed: selective inference, predictive inference, or interpretable ML.

2. Learn how to extend a "standard" machine learning topic taught in the class. For this type of project, the goal will be to learn about extensions of one of the methods that we discussed in class. It will be required to run a simulation study to demonstrate the impact of this extension over the standard approach (for or against). You also must apply your method to a dataset.

Project abstract (due on 10/27)

One-page proposals are due in class by the deadline stated above. These proposals should spell out briefly the main goal of the study, your basic approach, and the dataset you plan to use. Pick a subject that interests you. The problem you choose should not be something that you have already done. This project needs to show careful planning, good logic, and use the fundamental concepts learned in the course.

Final report (due on 12/1)

The final report will consist of a paper of at most 8-pages. This does not include tables, figures and code, all of which should be included at the end of your report (code can be included separately). You should make a thorough but concise report of your entire investigation.

An outline for the paper for style #1 is:

- Introduction
  - Background
  - Goal of the project

- Methods section
  - Description of the data
  - Description of model(s) in light of the data example and motivation
    * Methods
    * Model tuning
    * Performance metrics
    * Model comparison
  - Brief description of software
    * Useable code that can be used to reproduce all results must be included. Code will be graded by how easily it can be understood and efficiency.
- Results section
  - Exploratory analysis and assumption checks
  - Model comparisons
    * What did you see
    * Supporting tables/figures
  - Selected model
    * Describe the final selected model
    * Model performance (estimated unbiasedly)
  - Model inference or interpretability
    * selective inference, predictive inference, or interpretable ML.
- Discussion/Conclusion
  - Model comparison
    * Be sure not to over-state your conclusions
    * Anything unexpected?
  - Final model
    * What are the pros and cons of this model in light of the data example?
    * What are some methods that should be developed to address the cons?

An outline for the paper for style #2 is:

- Introduction
  - Background
  - Goal of the project
- Methods section
  - Description of the 'standard' method and it's extension

  - Other comparison models (if any) and optimization
    * Methods and optimization
    * Model tuning
    * Performance metrics
    * Model comparison
  - Brief description of software
    * Useable code that can be used to reproduce all results must be included. Code will be graded by how easily it can be understood and efficiency.

- Simulation section
  - Description of data generation
  - Model comparisons

- Results section
  - Description of the data (including motivation for using the method)
  - Exploratory analysis and assumption checks
  - Model comparisons
    * What did you see
    * Supporting tables/figures
  - Selected model
    * Describe the final selected model
    * Model performance (estimated unbiasedly)

- Discussion/Conclusion
  - Model comparison
    * Be sure not to over-state your conclusions
    * Anything unexpected?
  - Summary of method
    * What are the pros and cons of this method? (these should be demonstrated in the analysis and simulations)
    * Are their any further extensions that address the cons.
    * What are some future methods that haven't been delevoped?

Oral Presentation

On the last two days of class (Tuesday, December 5th, Thursday, December 7th), each person will give an oral presentation of 15 minutes. The order of the presentations will be randomized.

Grading

Scores for these projects will be assigned based on the following rubric:

- Introduction 10%
- Methods Section 20%
- Simulation Section 10% (style #2 only)
- Results Section 20% for style #1, 10% for style #2
- Discussion/conclusion 5%
- Code 5%
- General readability of report 10%
- Seminar presentation 30%