Math 678 - Final Project

You only need to complete ONE of the following three options. Email me ANYTIME with your questions.

1 Real Data Analysis

Choose one data set and apply at least two machine learning methods to the data. The size of the dataset should not be too small or simple, so that the learning task is non-trivial. You need to make comparisons of different methods in terms of their model assumptions, methods, selecting tuning parameters, and algorithms. Write a report which should include the following elements:

- describe the data set and the background of this study.
- introduce the key idea of each method, its performances, advantages, and limits.
- summarize the results given by different methods and major discovery.
- address the questions like: Do all the methods give consistent results? Does one method perform significantly better than others? Why?

2 Creative Work

Develop a new method or idea that may improve existing methods or provide a new solution to a challenging machine learning problem. The method is not necessarily entirely novel, as long as it contains some new thoughts. The focus can be on theoretical, algorithmic, exploratory, or graphical. The report should

- introduce and define the statistical learning problem.
- review the literature briefly and your motivation.
- describe the main idea, the proposed algorithm, implementation of the algorithm, and its implications.

- discuss statistical properties (if possible).
- implement the method on some simulated or real examples.

3 Literature Review

Identify a machine learning topic you are interested. Read at least 2 relevant papers on the topic and write a literature review. Include the following elements in the report.

- clearly define the statistical learning problem and specify the goal of this analysis.
- summarize main ideas and results in each paper using the same mathematical notations defined by you.
- compare these methods regarding their formulations, model assumptions, computational cost (if relevant), statistical properties of the estimators, and etc.
- show some numerical examples or codes (which can be extracted from the papers).
- address the questions such as: What are their strengths and limitations? Is there room for further improvement? Which one do you recommend in practice?

Suggestion: The students are welcome to talk with me on identifying the topic and papers. The choice depends on your interest, background, and expertise.