STAT/CSCI 4990 Final Project Guideline

General Rules

1. **Academic Honesty:** This final project is an open book and take-home project. Though the discussion is encouraged, each group should finish the project with their own efforts. Students can read and search related materials from both offline and online sources. Another author's intellectual contributions (e.g. language, codes, figures, thoughts, ideas, expressions, etc.) should be properly cited when they appear in your project report. Students are expected to **abide by the UGA academic honor code** and must not copy from the works of others. This includes published or unpublished articles, Wiki documents, etc. Plagiarism will result in failure and, very likely, will have more severe consequences.

2. **Goal:** The goal of this project is for students to demonstrate their abilities to analyze a data science problem with suitable statistics and machine learning techniques, produce a written report explaining statistical analyses that solve the stated problem, and present their findings and analysis results to the class. There is not necessarily a uniquely correct answer for each problem. A successful report/presentation should produce not only correct analysis methods but also explanations that would be comprehensible to someone with only a basic knowledge of data science.

3. **Deadline and Submission:** The due date of the written report is **06:00 PM of April 24th, 2022 (Sunday)**. No late submission is allowed. The project should be submitted **electronically in eLC** in terms of a formal research report. The students can write up the report in any word processing software, however, the submission should contain **only one single PDF file per group**.

Written Report

1. **Grading:** The **written report** contributes **30% of the final score**. The grading is based on the **overall quality of the report**. The following aspects are considered important for a high-quality report: (a) the methodology used should be suitable; (b) the implementation and results should be correct and clear; (c) the explanation should be comprehensible; (d) the mathematics involved should be rigorous; (e) the presentation should be precise and concise; (f) the report format should be correct.

2. **Format:** The report should be prepared on **A4/US letter-sized paper**. The main report should be **no more than 8 pages** including everything (e.g. text, equations, tables, figures, etc.) except the list of references. The **list of references** should be added at the end of the report. The **font size** should be no smaller than 11 pt and the line space should be at least single-spaced. The **margins** should be at least 1 inch on top and bottom and 1.25 inches on left and right. All **tables and figures** included in the report should be properly numbered. You may only number the **equations** that are referenced somewhere else in the report. The **sections and subsections** should also be properly numbered. Remember that you have a limited budget on pages, please use that wisely by choosing what to include in your report.

3. **Implementation and Coding:** Students can implement any suitable and reasonable method to solve the problem. For each method, students should give a clear description of its implementation including the details like how to define the objective function, how to compute the estimator, how to choose the tuning parameters, etc.. The standard of clarity is that someone else can replicate the method based on your description. Implementing methods beyond the scope of lectures are neither awarded nor punished. Students can code in any programming languages. Please do not include the codes in the report. If necessary, you can summarize your program in an algorithm format (like the ones shown in lecture slides). Again, remember the page limits.

In-class presentation

1. **Grading:** The **in-class presentation** contributes another **30% of the final score**. The grading is based on the **overall quality of the presentation**. The instructors and TA will serve as the judges. The following aspects are considered important for a high-quality presentation: (a) organization of the content; (b) the clarity and correctness of the presentation; (c) the preparation of slides; (d) handling of the questions; (e) balanced teamwork.

2. **Format:** Each group should prepare and deliver an in-class presentation in **Week 15**. Each presentation will be of length **20 minutes**. Each group member should present no less than **5 minutes for a group of 3 people** and no less than **4 minutes for a group of 4 people**. After each presentation, there will be a **10 - 15 minutes** session for questions and discussions. The instructors and TA will ask questions regarding the content in the presentation. The students in the other groups are also encouraged to ask questions.

Additional information

1. **Group meeting with instructors:** We encourage each group to meet and discuss their final project with the instructors. You can join in the regular office hour to have a quick discussion or make an appointment. The role of instructors in these meetings

is multifold. First, we will act as instructors to explain any part of the rule or project problem that seems ambiguous to you. Second, we will serve as "project managers" to oversee the progress of your project and provide some high-level suggestions. We will not help you with detailed technical issues (mathematics, coding, writing report, etc.). Third, we will serve as "clients" who communicate the demands with you from a non-expert point of view.

2. **Poster session:** The statistics department is organizing a student poster session, where the students from various classes can demonstrate their projects to the audience outside their classes. This event is likely to happen at the end of this semester. You are encouraged to attend. The instructors will provide you with necessary assistance such as the template to make a poster. The statistics department will cover your cost for printing the poster. We may provide the team who successfully participated in the poster session with a few bonus points. We will announce more details when they are available.