

DATA 5322 Statistical Machine Learning II

Spring Quarter 2024

Written Homework Guidelines

Overview

Written homework will be assigned with each chapter and will explore the full data science methodology of collecting, processing, and analyzing data as well as communicating results. These assignments require the use of computational and coding skills and will emphasize visualization and communication skills. Typical writing assignments will be exploratory and are meant to reflect real-world problems. Assignments are built for students to showcase mastery of course content, but will be based on real data and may require additional resources, data cleaning, or other tools that are not directly presented in class. **As such, it is expected that these assignments will take longer to complete.** Each written homework will have two deliverables: (1) your code and (2) a written report of your work. The reports should be written to explain the concept/theory, your methodology, and your findings. They should be professional and polished and will be graded on the presentation and communication (neat, complete, clear language, sufficient plots that are correctly labeled and discussed, etc). Expectations for written communication are high—students can seek extra support at the Writing Center and the English-Language Learning Center.

Format and Contents

Reports should be formatted nicely in a word processing software. The report should not take more than 7-10 pages, including plots—try your best to convey only the most important details and results. Language should be clear and concise. All equations should be typeset. All plots should be carefully considered (i.e. not simply the output of the function you used!), must of good image quality, and should be captioned and labeled. Each report must include the following sections:

Title and Abstract: A descriptive title and short (100 word) abstract

Intro and Overview: Setting the goals of the report; introducing the topics/methods; description of the data set to be investigated

Theoretical Background: A summary of the theory behind the concept you will use (does not need to be too detailed/rigorous); description of appropriate uses of the method; critical details like tuning parameters, equations, run time, etc.

Methodology: A short description of data processing/cleaning; description of the computations you ran, including selection of variables, error comparisons, cross-validation, etc

Computational Results: Presentation of your computational results (nice plots!)

Discussion: Interpretations of computational results; key findings and their relevance; discussion of improvements to or extensions you could make to this work

Conclusions: Discussion of the broader impacts of your findings; summary of the report

Bibliography/References: Numbered citations (i.e. not just URLs) to any sources you used, should be directly connected to specific content in the body

Appendix: Code Cleaned version of the code you used