ST 516: Midterm Project

Guidelines for Project Reports

This document provides guidance for content of the midterm project reports. Reports should be no more than five pages in length (single-spaced, 10-point font or larger) and follow the format provided below. When writing your report, begin with the end in mind, i.e., lay out the story you want to tell to describe and support your key findings. Since you have space limitations, this is an effective way to make sure you have a coherent message that is clearly communicated. And while you will have to be judicious in your use of graphs and tables, include the right visualization aids to help clearly supporting your findings.

Executive Summary

This section should be no more than 2-3 paragraphs and should contain the following

- One or two sentences summarizing the problem statement.
- Identify the model you chose (type of model and why --accuracy, interpretability, etc.).
- Key findings, e.g., "ridership was heavier on sunny days by xx%." You can have several of these. See section below "4. Results" for other elements to include.

1. **Introduction**

   A short description on what you are trying to model and why. You can borrow heavily from the problem statement in the assignment. You should include the modeling approaches you are going to use to solve the problem, and describe any methods you plan to employ to compare results of the different methods. It is often easiest to write the introduction after you have finished the body of the report so it can serve as a roadmap for the reader regarding the story of your analysis you are going to tell.

2. **Data**

   In this section, describe your data set (variables, number of observations) and the origins of the data. You can borrow heavily from the UCI descriptions. You should also include any exploratory analysis here, like scatterplots and correlations among the variables. While you may not have to do any pre-processing for this project, in the real world this is also the place to mention any pre-processing you may have performed (filling in missing values, etc.).

3. **Methods**

   This section describes the meat of your analysis. Here you identify the types of models you decided to fit to the data (and why you chose those approaches), summarize the

fits, and discuss how the models compare to each other (you choose the types of comparisons you want to make). You should also describe the results of any diagnostics you choose to use, and, if any problems exist, describe any approaches you took to remedy those problems.

4. **Results**

   In this section, you will identify the model (or models) you chose to use and the reasons for your decision. You will also identify any key findings in terms of importance or relationships between the predictor variables and the responses, and well as point out differences in these relationships between the two response variables. You should address both prediction errors and, if necessary, interpretation, quantifying relationships between specific predictors and response if and where appropriate for the latter. Also mention any shortcomings of the model(s) you selected.

5. **Conclusion**

   A short recap of the body of the report (no more than 2-3 paragraphs), and include any future work you think might be useful.