

ECONOMICS 690: ECONOMICS OF MACHINE LEARNING
UNIVERSITY OF WISCONSIN - MADISON
FALL 2020
PROJECT

INSTRUCTIONS:

- The purpose of the assignment is to have you demonstrate that you can apply the techniques learned in the course to a novel setting in order to obtain answers to economic questions of interest
- You will work in groups. In light of the logistical challenges this semester, we will assign you to groups of 5-6 people by the middle of September. In particular, to facilitate interaction, we will try to make sure that students in a group are in similar time zones.
- You can either choose a topic from the list on page 3, or design your own project. In case you choose one of the projects we outline, the questions that we list are minimal guidelines you should follow, but you are free to expand your analysis beyond those.
- You'll receive a Google form to fill out with your group and topic information by **Sept. 30, 2020**.
- Your group will submit a final report and relevant code by **December 11, 2020**.
- For computational advice, you should first meet with Greg.
- Projects will incorporate strategies from some of the major topics in the course (i.e. data visualization, supervised learning, unsupervised learning) in a way that is appropriate to answer the question of interest. While technical skills in implementing ML techniques are important, the best projects will also motivate carefully the rationale for the analysis, and provide a compelling explanation of the results.

FINAL REPORT:

- The report should contain the following sections
 1. Abstract summarizing the question and results (no more than 100 words)
 2. Introduction
 3. Data section summarizing data set, providing relevant summary statistics, etc.
 4. Methodology section explaining the techniques that you used, why you used those techniques, etc
 5. Results section presenting and explaining your main findings, *including a discussion of the caveats in interpreting the results*
 6. Conclusion
 7. Do not include printouts of your code, but attach to your submission the programs necessary for replication
- **The report should be at most 10 pages long (not including the appendix)**
- Follow all of the style guidelines that appear at the end of this document
- Only include material that is relevant to the argument in your paper. Some students have a tendency to include as much course material as possible in their answers on exams in the hopes of showing every fact learned from lecture. True mastery of course material involves not only being able to recall facts and figures from lectures, but knowing when they are actually relevant and applying them appropriately. You should attempt to build a tight argument in the paper, without extraneous material.
- Similarly, when deciding the methods you want to use, more sophisticated does not always mean best. Part of demonstrating knowledge is being able to choose the best tool available to answer the research question, even if it is not the most sophisticated. You should not aim to include every technique you have learned in the course, instead choose the ones appropriate for answering your question.

PROJECT TOPICS

1. SUPERMARKET SCANNER DATA

- Dataset: Dominick's Database available at:
<https://research.chicagobooth.edu/kilts/marketing-databases/dominicks>
- Choose a product / market (ie beer, ice cream, etc)
- Describe the promotions you observe in the data
- Given the types of promotions you observe in the data, analyze the relative efficacy of each. Some dimensions to consider:
 - Does the efficacy vary by type of product?
 - Does it depend on time of year?
 - Does it differ across store locations? For what reasons?
- For the largest firm in your chosen product market, design a detailed promotion strategy and provide a prediction of the campaign's effects that justifies your strategy.

2. AUCTIONS

- Dataset: Greg Lewis's Ebay auction data available at:
<https://www.aeaweb.org/articles?id=10.1257/aer.101.4.1535>
- **Do not contact Greg Lewis**
- Describe the types of auctions you observe in the data (number of bidders, revenue)
- Predict winning bid based on auction and product characteristics
 - Do revenues vary by number of bidders?
 - Do revenues depend on time of year?
 - Do revenues depend on auction characteristics such as buy it now, text, photos, etc?
- For the most commonly observed car model in the data, provide a recommendation for the auction format including but not limited to: buy it now, length as auction, photos, text description.

3. STUBHUB

- Dataset: Andrew Sweeting's Stubhub database available at:
<https://www.journals.uchicago.edu/doi/suppl/10.1086/669254>
- **Do not contact Andrew Sweeting**
- Predict game attendance and ticket prices based on characteristics
 - Do the prices vary by team (home, visiting) record, time of year, day of week, time of day, stadium, section, etc.?
- For your favorite team, provide a recommendation on seller strategy on Stubhub, including which tickets are best to resell and when you should post your tickets.

4. DESIGN YOUR OWN QUESTION

- If you choose this option, you must submit a written proposal **by Sept. 30** that discusses the following:
 - (a) The question you will answer
 - (b) The dataset you will use (by the time you present your proposal, you should have a dataset, or at the very least a concrete strategy for obtaining it)
 - (c) The features of the dataset that will allow you to answer the question of interest (i.e. variables in the data, number of observations, etc.)
 - (d) A *detailed* plan describing the estimation strategy you will employ
- When designing your project, the project outlines described above can serve as a template.
- We will evaluate proposals based on whether we feel that the proposed question is appropriate for the course and can be feasibly answered with the data. *At our discretion, in case we feel your chances of writing a successful paper would be greatly improved by working on one of the above topics, we will assign one of those instead.*

STYLE GUIDELINES

These style guidelines are a slightly modified version of those for Econ 706.

In reading the papers, we pay considerable attention to style (correct spelling and grammar, clear exposition, good organization). So, too, do most people: reports that are difficult to read routinely get ignored, even if they contain good ideas. Thus, it will pay to develop the habit of working hard to craft a clear explanation of your ideas.

Many of the following suggestions are standard good practice. Others are matters of taste.

1. In writing up a research report, one should have an audience in mind. We suggest that you take the audience to be your fellow students in this course. They'll know most of the relevant economics and methodology, but won't know your data set or your approach.
2. Include a cover page with the following information: Title; date; your name; your e-mail address; the word "Abstract"; an abstract of 100 words or fewer. If you have acknowledgments to make (thanking a fellow student for helpful comments, for examples), put these on the bottom of the cover page. The text of the paper begins on the next page.
3. Your paper should be divided into sections, to help guide the reader. You should have the following sections: Abstract, Introduction, Data, Methodology, Results, Conclusion, Appendix
4. Number the pages. No plastic covers or binders, please.
5. In the introduction, present an overview of your paper and summarize your findings. In the conclusion, give suggestions for future research.
6. Be explicit about your data set. State the sample size. For time series, state whether the data are monthly, quarterly or annual, and whether or not they are seasonally adjusted. State the units of measurement. For example, if ?income? is a variable, state whether it is measured in current dollars or constant 2008 dollars, and if it is per capita, say so. For data in logs or log differences, you will usually want to multiply by 100 so that the units will be percent or percent change. Also explain the choice of the sample: why does it start in a given year, or you use only a cross-section from a given year instead of a panel, and so on.

7. You will want to include a plot and/or a table with basic statistics (means, standard deviations) of the data.
8. Number the equations. You can number the equations by section, if you have sections in the paper. That is, the third equation in section 2 can be numbered 2-3, the second equation in section 4 can be numbered 4-2, etc., if you prefer doing this to numbering sequentially through the paper.
9. Tables:
 - (a) Number the tables, and on each include a descriptive header (“Means and Standard Deviations of Data,” or “Variance Decompositions,” for example).
 - (b) Tables may appear in the text in the appropriate place, or at the end of the paper.
 - (c) Tables should not run over page boundaries, unless they are too long to fit on a single page. That is, if you include a table in the text, you should insure that you place it so that it does not run from one page to the next.
 - (d) Make every effort to make each table self-contained, even though this will require you to redundantly present information that is also stated in the body of the paper itself. This is now the standard in the profession and you should look at a paper published recently to see how much detail is included in tables.
 - i. In notes at the bottom of each table, define the symbols that are in the table, or give a precise reference to where the definition may be found. It is not adequate to simply state “definitions are in the paper” or “see section 2 of the paper for definitions”. Instead say something like “Variable definitions: y =log per capita income in 1992 dollars, r =interest rate on 3 month Treasury bills (end of quarter),” and so on. Alternatively, for many of you it might be best to include a table that defines the symbols, and in subsequent tables say “see Table x for variable definitions” where “x” is the number of the table that defines the symbols. (You will also present such information in the text itself.)
 - ii. In tables that present regression results, include a note that describes the estimation technique (“The probit was estimated by maximum likelihood, assuming normality,” for example.) (You will also present such information in the text itself.)
 - iii. If a given set of variables appears in more than one table—as is often the case—there is no need to repeat the variable definitions. Instead one of the notes to (say) Table 2 can say “Variable definitions are given in notes to Table 1.”

- iv. Be sure to include the name of the dependent variable somewhere in the table.
 - v. When possible use words to describe the variables in your model. For example, if years of schooling is a regressor in your model write out “Years of schooling” not “*YRSCH*” if that is the name of the variable in your statistical software.
- (e) In all but the simplest tables, number the rows and columns. When the text references a result in the table, cite the row and column: “the t-statistic is 2.12 (row (2), column (4))”
10. Figures:
- (a) Number the figures, and on each include a descriptive header (“Parental Income versus SAT Score,” for example).
 - (b) Figures may appear in the text in the appropriate place, or at the end of the paper.
 - (c) Figures should not run over page boundaries, and must always fit on a single page. That is, if you include a figure in the text, you should insure that you place it so that it does not run from one page to the next.
11. Reporting of estimates:
- (a) Do not report more than 3 or 4 digits. Example: report 0.412, not 0.4117678.
 - (b) Avoid long strings of zeroes at the beginning of a number. You can always retroactively rescale variables and coefficients.
 - (c) Report standard errors, not t-statistics. Standard errors belong in parentheses under the coefficients. Example:
- | | | | |
|--------|---------|-----|-------|
| Report | 0.412 | not | 0.412 |
| | (0.146) | | 0.146 |
12. Avoid the use of elaborate acronyms to denote variables (like $AUSGDP_{90}$ for Australian GDP in 1990 dollars). They are rarely helpful to the reader. A single letter, usually with a subscript, ordinarily suffices and is easier to read when used in equations.
13. References:
- (a) All references cited in the paper should be listed in a bibliography at the end of the paper. Cite these in the text as Walter (1995) or Walter (1995, p361).

- (b) When you reference a specific result, such as a point estimate of a parameter, or a theorem that establishes a particular claim, give the page number, such as Walter (1995,p361). When you reference a general result, for example noting other papers that have studied topics similar to yours, no page number is needed.
14. Computer code: You do not need to include your programs in the paper. We should be able to figure out what you did without seeing it explicitly.
 15. Miscellaneous reminders on terminology:
 - (a) Hypotheses (not tests) may be “accepted” or “rejected.”
 - (b) Hypotheses refer to the magnitudes of population parameters, not estimates, and not to statistical significance. The word “significant” should not appear in the statement of a hypothesis.
 16. It is a violation of scholarly ethics to repeat a passage, even a sentence, from another source without putting the passage in quotes and citing the source: the usual publication details in case of printed matter, the URL and date in the case of web-only material. This rule applies even when you are describing dreary facts: if you repeat a description from another paper of how data were collected, or the steps in computing an estimate, you must put the passage in quotation marks and cite the original source.