Cogs 109 Final Project guidelines

Fall 2018

- Work in teams of at least 2 and no more than 4 students. Every student in the group will be expected to contribute substantially to the final product(s), and all students should be able to understand and explain all aspects of the project when you present your work in the final symposium.
- Your project should:
    1. Identify a real problem, challenge or scientific question which could benefit from data analysis and modeling. Your final report should explain why the question is interesting or important.
    2. Identify a relevant data set. You should learn about how the data was collected and be able to explain key features of the data, for example: How many observations? What are the noise sources? What are the relevant predictors?
    3. Identify at least one relevant data analysis approach, choosing from the methods covered in the course (linear or nonlinear regression, classification, clustering, PCA, etc.). Explain why this analysis approach is appropriate for addressing your question.
    4. Identify and explain one or more hypotheses or initial expectation that you will test using the data.
    5. Model selection: You should compare and contrast at least TWO different models. Your comparison should make use of cross-validation, bootstrap sampling, regularization, and/or other relevant techniques. For example, you might compare K-Nearest Neighbors classification for a range of k values (k=1,2,…,50), and select the k value that provides the lowest test set (cross-validation) error.
    6. Model estimation: Implement your data analysis and present the results using a combination of data visualizations (box plots, scatter plots), statistical analyses and models.
    7. Present your conclusions and outlook for next steps/future directions.
- The final product will be a written report, 5-10 pages in length. In addition, you will create a poster explaining your project to be presented in a symposium session on the last day of class. We will provide more information about the final paper and poster in a few weeks.

Key dates:

Date	Item
11/15	Draft project proposal due. Post your group members, title and abstract using this form: https://docs.google.com/forms/d/e/1FAIpQLSfzJ39N4jGeAiHNYoRP-PCxsOLxnMy0ZeJ_F8hUO3q2kMTAsg/viewform Your proposal should list all the students in your group and provide a title and brief (5-10 sentences) abstract summarizing your topic and the data sources and analysis approaches you plan to use.
11/15-11/20	Public comment period. Every student will be randomly assigned to post a constructive comment/question about 2 other projects. You are welcome to read and post comments on more than 2.
11/20-12/4	Work on final projects. If you would like to consult with a TA, IA or professor about your project, visit them in office hours or schedule a meeting.
12/4	Final project report due
12/6	Poster session symposium

Additional guidelines:

- Remember that your project must use at least one of the data analysis methods covered in our class.
- Final proposal:
    - Submit your revised proposal on the course website discussion board by 11/15 at midnight.
- Final paper:
    - Must be submitted by email as a PDF or Word document to your TA, no later than **Monday Dec. 4** at midnight.
    - Total length should be 5-10 pages. Please use 11 point font and 1" margins. Either double- or single-spacing are fine.
    - Your paper should include at least 4 full pages of text.
    - The paper should include the following sections (in this order):
        - **Abstract**: A brief, 1-paragraph (3-7 sentences) summary of your project's topic, methods, and results.
        - **Introduction**: What is the scientific question? Why is it interesting and/or important? What was your hypothesis?
        - **Materials and Methods**: Describe the data set you used, include key characteristics such as the number of observations, number and nature of predictors, etc. Describe the methods you used, explaining why you chose those particular methods (i.e. why is the method appropriate for this type of data). The methods description should be concise, but it should include enough detail so that the reader could fully reproduce your analysis. Thus, if there are particular choices you made in the course of your analysis (e.g. variables to include or exclude, preprocessing steps), these should be described.
        - **Results**: For each of the analyses you performed, describe what you found. You may include figures and/or tables (see below), with captions (see below). For any result, you must include an analysis of its statistical significance. This could include reporting a p-value (if you performed a regression analysis), a cross-validated estimate of mean squared error, and/or a bootstrap-based confidence interval.
        - **Discussion and Conclusions**: What did you learn from your analysis? Were the results expected or surprising, and why? What would you recommend as next steps for future researchers interested in this topic?
    - Figures and/or tables may be included in your paper. Axes must be clearly labeled and the figures should clearly and concisely convey the results of your analysis. Make sure font sizes are large enough to be easily legible. Each figure should be captioned with a descriptive title, and a clear explanation of the figure's content. For example:
        - *Figure 1: Correlation between air temperature and taco sales. The figure shows a scatter plot of air temperature vs. taco sales (black dots), as well as a linear regression line fit by least squares (dotted line).*
- Final project poster
    - On the last day of class (Thursday Dec. 6) we will have a poster session to view and comment on each other's projects.

- Each group should prepare a poster explaining your project. Your poster may include the same sections as your paper (Abstract, Intro, Methods, Results, Discussion/Conclusions)
- At least one member of each team should be present at the poster throughout the poster session.
- The course teaching staff, as well as guest reviewers (CogSci graduate students), will be invited to view your posters and comment on them.