November 1, 2015

Final Project

CSCI E-81 Fall 2015

Goal

The goal of the final project is to take what you have learned and apply it to something that interests you (that is hopefully cool). The field can be anything but the approaches should lie mostly within the machine learning and data mining area though many projects may also require some web scraping, programming, visualization, etc. We hope that your fellow classmates draw inspirations and ideas from your project as part of the overall learning experience. Our plan is to post the projects and make them available to your classmates.

Kinds of Projects

- Publishable research where you solve a problem in your area
- Extensions of research where the published solution is limited and you press forward with ideas on modifying the algorithm, trying new approaches, etc.
- Projects that solve a problem or create an opportunity in your field/work (with a very strong preference to those that do not require proprietary data)
- Projects that improve the world or our understanding of it

Individual or Team projects

After discussions with several students and within the teaching staff, we have decided to reverse course and accept either individual or team projects. Many of the teams seem to be working quite well so you can continue to work together. However, finding a project topic that you are passionate about may be more personal so we are allowing individual projects as well.

Language & Environments

- Python, Java, R, Matlab, etc. are all fine
- As long as the teaching staff can run it, there is no issue.
- For less standard environments, talk to us
- Be sure to give us instructions on how to set up and run your code

Get help from your friendly neighborhood teaching staff

Our teaching staff is looking forward to helping you in any way that we can including:

- choosing a project
- guiding the research
- interpreting the results
- working toward a successful outcome

We have office hours, sections, and are generally always available on piazza/email.

Grading criteria

The final project will be similar to a large homework assignment in scope. We expect you to be able to leverage your machine learning and data mining. You will be evaluated on how you approached the problem and the types of solutions that you tried to move past obstacles.

You will also be evaluated on the presentations, written report quality, and understandability of the code. The presentations will be 6 minutes for individuals and 10 minutes for teams which is short so you will have to practice and time your presentation to share the meaningful portion of your project.

If you are working with a team, we will also ask that you submit the partner evaluation spreadsheet used in the past paired project assignments. We want to ensure that those contributing the most are appropriately rewarded. Submit just one project submission per team and one partner assessment per member.

Bonus points? Those who go above and beyond will certainly be rewarded with bonus points. Since the path of research is never certain, it will be difficult to define these up front but we can certainly discuss project scope on an individual/team basis.

Report Format: roughly a journal article format

- Title
- Author(s)
- Abstract: 1 paragraph summary
- Introduction
- Data sets: what they are, where they're from, citations, etc.
- Methods & Results (may be split into different sections)
- Discussion
- Conclusion
- References (likely but depends on the type of report)

Target length is probably 7-8 pages of text excluding references with as many figures as is appropriate to tell the story. It is not necessary to include code in your report unless you feel it helps tell the story.

What to submit: A single compressed file with a directory structure

Create a directory with your project or team name and inside that directory put:

- Written report in top folder
- Code directory (do include "old" code directory of intermediate solutions). Include small libraries if necessary.
- Instructions on how to run the code if applicable in top folder. Include instructions for downloading large libraries
- Data set directory (unless impossible for proprietary or other reasons)
- Slides for presentation (if any) in top folder

Zip the directory and submit it to the Canvas system

Deadlines

- November 9 — Project idea (1-2 paragraphs) defining goals, scope, and team
  One submission per team is fine
- December 11 & 18 — In-class presentations
- December 18 — Written report due--no extensions unfortunately

Verbal presentations will occur during part of the class on December 11 and the entire class on December 18. We realize that the projects presented on December 11 will not be fully complete before the deadline. We will happily accept volunteers to present Dec 11, otherwise we plan to use a lottery system.