Math 5366 Final Project

The objective of the final project this semester is to learn more about some technical skill that you are interested in or some other topic related to data mining. You will give a 10 to 30 minute presentation about this topic during March or April and write a report about it, which is due by May 3. Preferably, you will learn a new skill and apply it yourself to a problem of your choosing, but this may not be possible for all topics.

Rules:

- 1. You can work in groups of 1–3 people.
- 2. Each group must do a different project (check the sign-up sheet before choosing a topic). It's ok if there is some overlap between two groups' topics, but there should be some significant differences also.
- 3. Everyone should sign up for a topic by February 15.

Taking risks on this project is encouraged. It's better to do a decent job with a difficult topic than a very good job with a safe topic. The purpose of this project is to stretch yourself, learn something new, and teach other students in the class. Here are some possible topics:

- 1. Scraping data from web pages (with Python, R, or some other language).
- 2. 0xdata and H20 ("Fast Scalable Machine Learning for Smarter Applications")
- 3. RJDBC package in R (Enables SQL queries to be made with R)
- 4. Shiny by RStudio
- 5. CUDA (High performance computing with parallel processing. You can investigate RCUDA or PyCUDA also).
- 6. Using machine learning to train self-driving vehicles and other robots (real or simulated).
- 7. Sweave (Writing documents by interfacing R with latex).
- 8. plyr package in R
- 9. caret package in R
- 10. The twitteR package for analyzing Twitter data with R
- 11. A company actively involved in data mining (Revolution Analytics, for example).
- 12. Miscellaneous topics in Python, SAS, SQL, or R.
- 13. A topic of your choice (It may be a good idea to look at job postings, pick a skill you don't recognize, and learn more about that skill.)