CSC 324/424 Winter 2016 Take-home Final Exam Due Tue/Wed March 15/16, 2016 @ 10:00pm in D2L Dropbox

The purpose of this final exam is to test your ability to apply the techniques of this course without specific prompting to run a particular technique with a particular dataset. This is a course where it is especially true that, “You get out of it what you put in,” because you can do well on the assignments and other tests without digging in and working with the techniques we have discussed. However, digging in and getting practice is the only way to fully understand these. I encourage you to enjoy this assignment and look at some interesting data, but I also understand many people will have other obligations and I will not judge you for doing something straightforward.

Submission: turn in two documents. First (1) written responses to the questions below, labeled by question number and (2) your code and/or output from your analysis.

Questions:

1. Identify a multivariate dataset you will use for this assignment. If you haven’t already chosen and need ideas, take a look at the last lab.
   1. Provide a link to the dataset (if not possible, add the data as one of the posted files or explain why you cannot, e.g. corporate IP).
   2. Explain the variables of the data including their meaning and type (nominal, numeric, …)
   3. How many samples are there?
   4. Are there missing values?
2. Provide a question you have about this data that can be answered using a technique we’ve studied.
3. Explain what analysis technique from class you will use to answer your question, including why and how. I’m happy to see visualization in the results, but this is limited to the core techniques we learned plus SVM if you need it.
4. Test that the proposed technique can be used with the chosen data.
   1. Run the appropriate tests we discussed in class to determine if the technique is appropriate. Examples include testing for adequate sampling for PCA and Box’s M Test for LDA.
   2. Report the results of your test. If the data fail the tests, you must switch data or explain why you are going forward anyway.
5. Apply the chosen technique to the chosen data.
6. Validate the model constructed by the chosen technique by testing that it is significant and has a reasonable model fit. Use the criteria appropriate for the model you’ve chosen.
7. Finally, report on the results and answer your proposed question if possible.