CSC334/424: Homework 4

Due Sunday May 20, 2018 by midnight

1) **(30 points, group submission)** Submit a rough draft of your final project (Executive Summary) and Journal Article including the following, which will help you prepare your final document for the following week. **Look to How to Read a Journal Article within the Syllabus folder to assist in writing your journal article.**
    a. **Executive Summary (2 Pages)**
        i. Non-technical report (imagine writing a report to your boss, who is a business manager without any technical/statistical expertise)
        ii. A paragraph or two about the application and dataset including the research question
        iii. If needed, any necessary background information needed to understand the application or the results (be brief or short in this section)
        iv. Brief description of the methods
        v. Results in laymen terms and how it relates to application
        vi. Limitations of the research and future work
        vii. Final Conclusions about the research
    b. **Abstract (250 words)**
        i. Introductory sentence explaining the research question, which intrigues the audience to continue reading
        ii. A sentence about the methods used
        iii. A sentence or two about the results
        iv. A sentence about the conclusion and how the results tie into the research question and the application of the dataset.
    c. **Introduction**
        i. Explain the dataset and provide background on the application
    d. **Literature Review**
        i. Can be combined with the introduction
        ii. Provide any previous literature on the topic, research questions, or methods used to answer the topic related to your research question
    e. **Methods**
        i. The list of the techniques that you are using to analyze your data. Include descriptions of why each topic is appropriate.
    f. **Discussion and Results**
        i. For each technique, a preliminary analysis and preliminary results remember to answer your research question.
        ii. Plots or visualizations that back up or reinforce your analysis (in journals, they are labeled as figures)
        iii. Include a section that describes the common threads that link your analyses. Is there anything that is reinforced by two approaches? Are there any common threads that you are seeing? If your analyses are so distinct that they don't cross-over like this, what do each add to your understanding of the data?
        iv. Limitations of the research / how would you have completed the research differently
        v. Future work that could be conducted on the dataset or the application
    g. **Conclusion**
        i. Final conclusions about the research

2) **(10 points, individual submission)** Select one of the techniques covered in lectures 7 and 8 (i.e. LDA, CA, Cluster Analysis … any type) and apply it to some aspect of your data. Or research a new technique that we have not covered and apply it). Each team member should investigate a different aspect of the data.

3) **20 points):** A common application of Discriminant Analysis is the classification of bonds into various bond rating classes. These ratings are intended to reflect the risk of the bond and influence the cost of borrowing for companies that issue bonds. Various financial ratios culled from annual reports are often used to help determine a company's bond rating.

The Excel spreadsheet BondRating.xls (XLS) contains two sheets named Training data and Validation data. These are data from a sample of 95 companies selected from COMPUSTAT financial data tapes. The company bonds have been classified by Moody's Bond Ratings (1980) into seven classes of risk ranging from AAA, the safest, to C, the most risky. The data include ten financial variables for each company. These are:

LOPMAR: Logarithm of the operating margin,
LFIXMAR: Logarithm of the pretax fixed charge coverage,
LTDCAP: Long-term debt to capitalization,
LGERRAT: Logarithm of total long-term debt to total equity,
LLEVER: Logarithm of the leverage,
LCASHLTD: Logarithm of the cash flow to long-term debt,
LACIDRAT: Logarithm of the acid test ratio,
LCURRAT: Logarithm of the current assets to current liabilities,
LRECTURN: Logarithm of the receivable turnover,
LASSLTD: Logarithm of the net tangible assets to long-term debt.

The data are divided into 81 observations in the Training data sheet and 14 observations in the Validation data sheet. The bond ratings have been coded into numbers in the column with the title CODERTG, with AAA coded as 1, AA as 2, etc. Develop a Linear Discriminant Analysis model to classify the bonds in the Validation data sheet.

a) What is the performance of the classifier on the training data? Notice that there is order in the class variables (i.e., AAA is better than AA, which is better than A,…).
b) What is the performance of the classifier on the validation data?
c) Would certain misclassification errors be worse than others? If so, how would you suggest measuring this?