IS 507 Assignment 7

Due by Sunday, November 20, 2022 11:59PM

Problem 1: The Excel spreadsheet Alzheimer.csv contains one sheet named Alzheimer, which is data attempting to explain whether a patient has Alzheimer's Disease. These are data from a sample of 336 employees and consists of 9 variables for each patient. These are:

1) Dementia-Outcome variable-patient diagnosis
2) Gender-Female=0 and Male=1
3) Age-Age of patient (in years)
4) Education-Years of Education
5) SES-Socioeconomic Status 1=Low and 5=High
6) MMSE-Mini mental state examination score
7) CDR-Clinical Dementia Rating
8) eTIV-estimated total intracranial volume
9) nWBV-Normalize whole brain volume
10) ASF-Atlas Scaling Factor

Develop a **Linear Discriminant Analysis** model to classify the Dementia event from the other variables.

a) What is the performance of the classifier using cross-validation?
b) What is the performance of the classifier using training and testing?
c) Would certain misclassification errors be worse than others? If so, how would you suggest measuring this?

Problem 2: Select one of the techniques (i.e. CA, Cluster Analysis, LDA) and apply it to some aspect of your final project dataset. Or research a new technique that we have not covered and apply it). Each team member should investigate a different aspect of the dataset.

Problem 3: Using Google Scholar, locate **a journal article**, which uses cluster analysis in your field of interest. **Write a summary** of the journal article and how it utilizes the cluster analysis in **two to three paragraphs**. **Cite the paper in APA format.**

Extra Credit (5 points)

An academic paper from a conference or Journal will be posted to the Homework 4 content section of D2L. Review the paper and evaluate their usage of FA and LDA. In particular address the following: **(See article on Comparison of Latent Dirichlet Modeling and Factor Analysis for Topic Extraction A Lesson of History)**

- What is the application of this paper?
- What is the research question the authors wish to answer in this paper?
- What is Latent Dirichlet Modeling and what can we learn from it?
- How does this paper utilize Factor Analysis?
- What are the results and conclusions from this paper?
- What other areas or fields do you think would benefit from LDA?
- What other thoughts do you have on textual modeling?